# Online Appendix

## Table of Contents

# A ADDITIONAL FIGURES

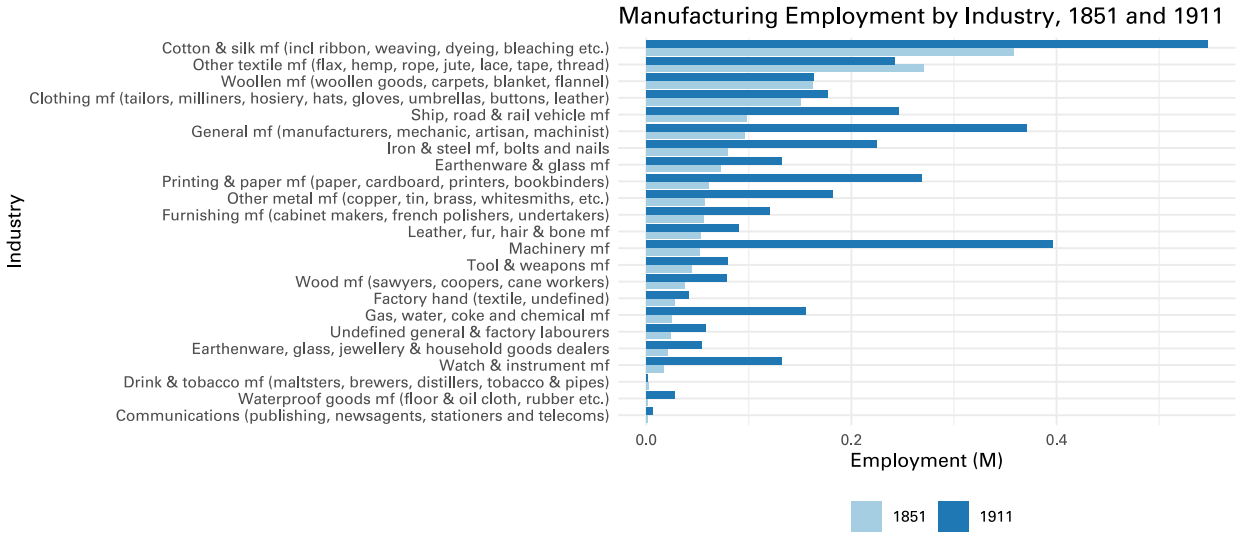


Figure A.1: Employment by manufacturing industry

This figure plots employment in different manufacturing industries in the 1851 and 1911 censuses, calculated using census microdata from Schürer and Higgs (2014) and industry categories from Bennett et al. (2017).

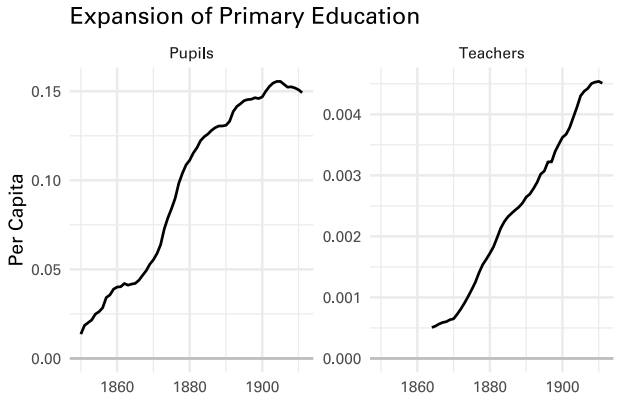


Figure A.2: Primary school teachers and pupils per capita

This figure plots the number of primary school teachers and primary school pupils, divided by total population. Source: Palgrave Macmillan Ltd (2013)


Figure A.3: Map of registration districts highlighting the location of Anglesey

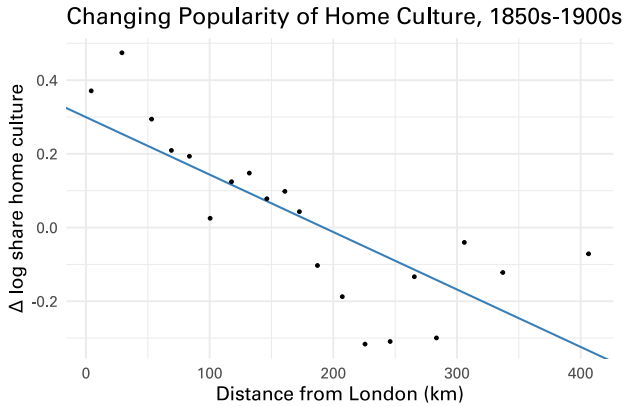


Figure A.4: Change in popularity of the home culture plotted against distance to London

This figure shows the change in the log share allocated names most associated with the home culture comparing those born between 1851–1860 and 1901–1910, plotted against the distance from the district centroid to the City of London.

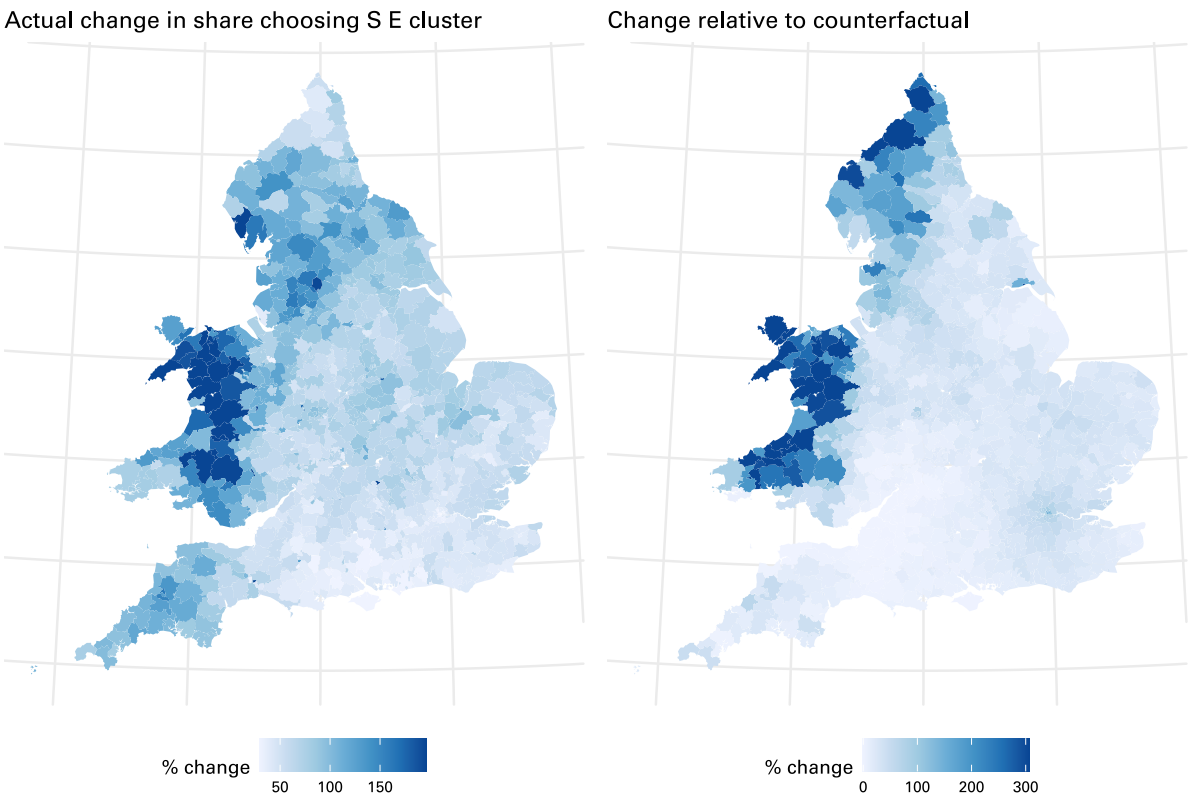


Figure A.5: Change in log share choosing the Southeast England cluster

Left panel is observed change between those born 1841–1860 and those born 1861–1895, right panel is change between counterfactual using 1851 estimates of migration costs, population, and destination utilities, and observed values for those born 1861–1895

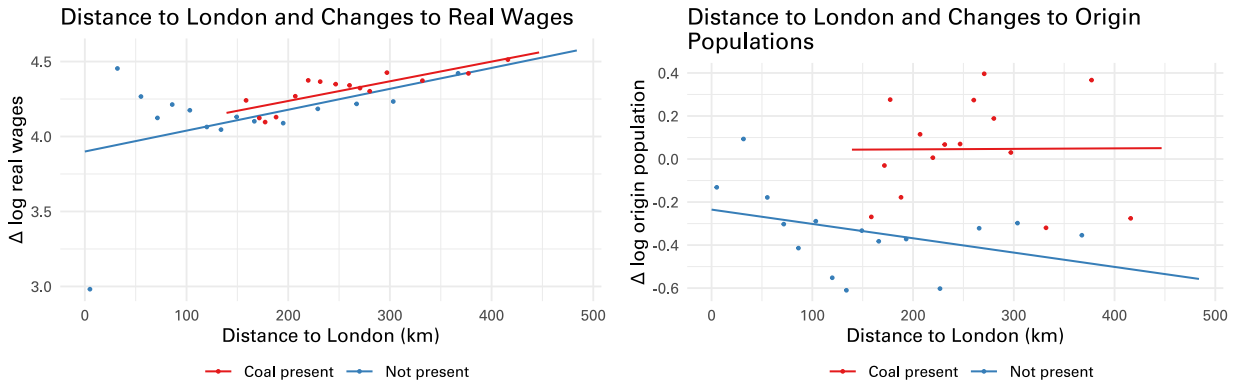


Figure A.7: Change in real wage and origin populations between 1851 and 1911, binned averages plotted against distance from London, subset by whether the district contains coal

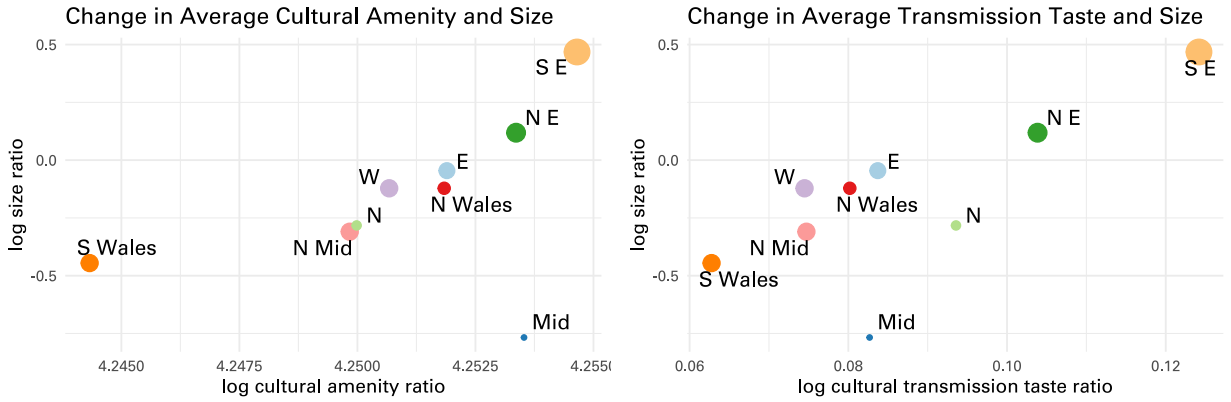


Figure A.6: Changes in average cultural amenities and transmission tastes and predicted changes in culture sizes

In both figures, the y axis is the log of the size of the cluster in the observed data, divided by the size of the cluster in a counterfactual with 1851 real wages and starting populations. To net out the mechanical effects of population growth, we use the predicted cultural choices ($\sigma$) from this counterfactual and multiply by observed starting populations. The x axis in the left panel is $\log \text{cultural amenity ratio}^k = \log\left(\frac{\sum_{d=1}^{N}\left(\xi_d^k v_{d,1911}\right)^\theta}{\sum_{d=1}^{N}\left(\xi_d^k v_{d,1851}\right)^\theta}\right)$. The x axis in the right panel is $\log \text{cultural transmission taste ratio}^k = \log\left(\frac{\sum_{o=1}^{N} n_{o,1911}\psi_o^k}{\sum_{o=1}^{N} n_{o,1851}\psi_o^k}\right)$. The size of the dots corresponds to the size of the cluster in the observed data.

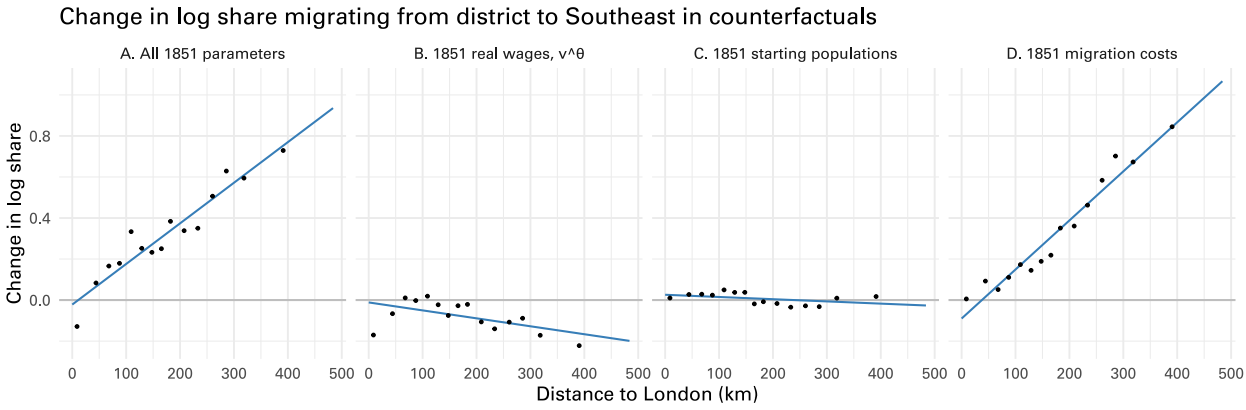


Figure A.8: Counterfactual changes in migration to districts allocated to the Southeast England cluster

The x axis is distance to London, the y axis is observed log share migrating from district in question to districts allocated to the Southeast, minus log share under counterfactual scenario.

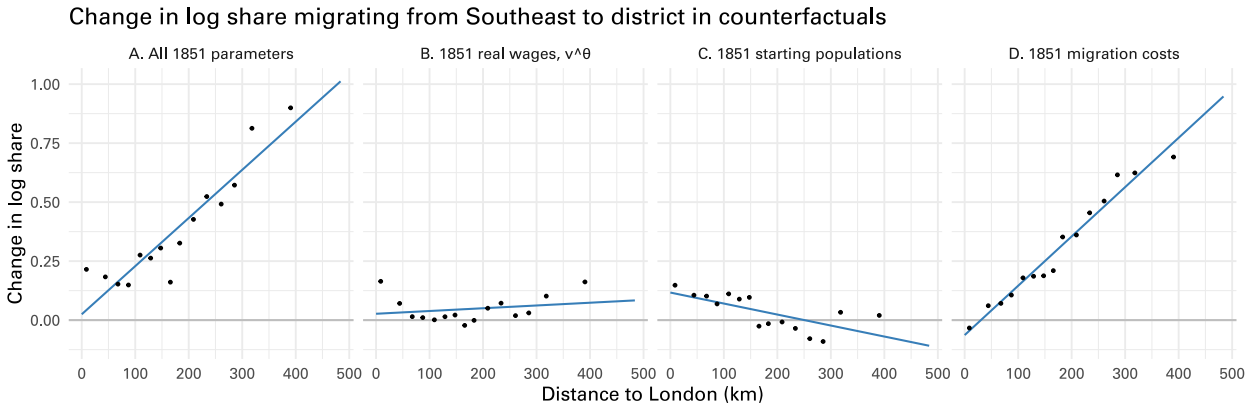


Figure A.9: Counterfactual changes in migration from the Southeast England cluster to districts

The x axis is distance to London, the y axis is observed log share migrating to district in question from districts allocated to the Southeast, minus log share under counterfactual scenario.

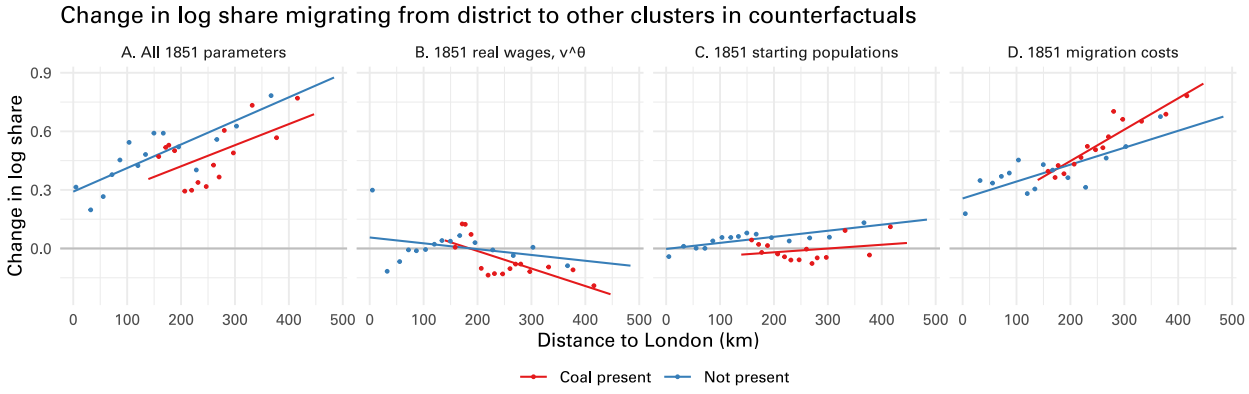


Figure A.10: Counterfactual changes in migration from the district out of the home cluster

The x axis is distance to London, the y axis is observed log share migrating from the district to districts not in the home cluster, minus log share under counterfactual scenario.

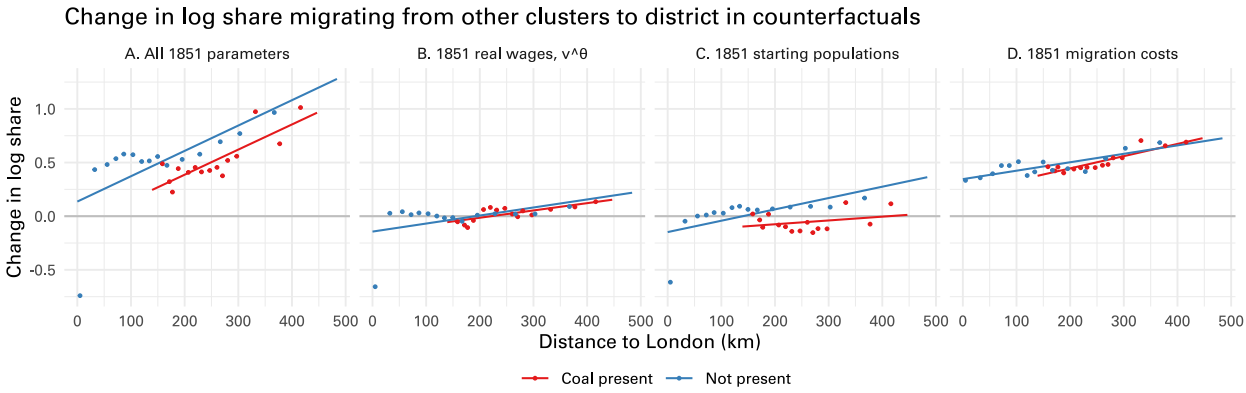


Figure A.11: Counterfactual changes in migration from non-home clusters into district

The x axis is distance to London, the y axis is observed log share migrating to the district from districts not in the home cluster, minus log share under counterfactual scenario.

## B ADDITIONAL TABLES

| | ln migrants | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| different cluster | −1.861 | −0.226 | −0.083 |
| | (0.074) | (0.050) | (0.033) |
| ln distance | | −1.532 | −1.475 |
| | | (0.083) | (0.075) |
| county-pair FE | | | x |
| N | 273400 | 273400 | 273400 |
| pseudo $R^2$ | 0.352 | 0.701 | 0.767 |

This table shows the results of regressions of log migration flows between registration districts in 1851 against an indicator for whether the origin and destination district are in different clusters, and log distance between the subdistrict centroids. The unit of analysis is the origin-destination district pair, restricted to pairs where the origin does not equal the destination. Models are estimated by poisson pseudo-maximum likelihood. All models have fixed effects for the origin and destination district. Model (3) adds fixed effects for each pair of origin and destination counties. Standard errors clustered by origin and destination district in parentheses.

Table B.1: People in 1851 tended to migrate within-cluster

| | probability of marriage ×1000 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| different cluster | −0.376 | −0.102 | −1.141 | −0.924 |
| | (0.028) | (0.010) | (0.065) | (0.056) |
| same district | | 0.931 | | 1.036 |
| | | (0.062) | | (0.071) |
| parish FE | x | x | | |
| cluster-by-parish FE | | | x | x |
| DV mean | 1.12 | 1.12 | 1.12 | 1.12 |
| N | 321382 | 321382 | 321382 | 321382 |
| $R^2$ | 0.531 | 0.539 | 0.535 | 0.545 |

This table shows the results of regressions of the probability of a given man-woman couple resident in the same parish being married (multiplied by 1000) against an indicator for whether both were born in the same cluster. The unit of analysis is the parish-man cluster-woman cluster-same district indicator, and observations are weighted by the number of potential couples in the unit—this is equivalent to a regression at the individual parish-man-woman level in which the dependent variable is a binary indicator of whether they are married. Models (1) and (2) have parish of residence fixed effects, (3) and (4) add fixed effects for the residence parish interacted with the man and woman's birth cluster. (2) and (4) also control for an indicator that both man and woman were born in the same registration district. Standard errors clustered by parish in parentheses.

Table B.2: People in 1851 tended to marry within-cluster

Table B.3: First names with highest name scores by cluster

| | | | Frequency 1841–1860 | | |
|---|---|---|---|---|---|
| Cluster | Metaphone | Example Name | Cluster | Total | Name Score |
| E | ELFN | Elvina | 51 | 142 | 0.918 |
| E | HKR | Hagar | 25 | 89 | 0.886 |
| E | MHL | Mahala | 108 | 449 | 0.863 |
| E | TMRS | Damaris | 17 | 72 | 0.860 |
| E | EFRT | Everett | 14 | 67 | 0.840 |
| E | ETKR | Edgar | 241 | 1190 | 0.835 |
| E | LRN | Lorna | 14 | 70 | 0.832 |
| E | OSBRN | Osborne | 12 | 61 | 0.830 |
| E | BRNBS | Barnabas | 12 | 62 | 0.827 |
| E | BLNT | Belinda | 27 | 143 | 0.822 |
| Mid | IX | Isiah | 48 | 86 | 0.842 |

| | | | | | |
|---|---|---|---|---|---|
| Mid | KMFRT | Comfort | 37 | 70 | 0.823 |
| Mid | RSHN | Rosehannah | 53 | 109 | 0.798 |
| Mid | IS | Isaiah | 218 | 499 | 0.767 |
| Mid | SLP | Zilpah | 35 | 81 | 0.762 |
| Mid | XTRK | Shadrack | 53 | 125 | 0.755 |
| Mid | SRHN | Sarahann | 207 | 530 | 0.731 |
| Mid | WLFB | Willoughby | 31 | 79 | 0.727 |
| Mid | NMN | Newman | 29 | 76 | 0.726 |
| Mid | TRSL | Drusilla | 98 | 263 | 0.716 |
| N | TKSN | Dixon | 28 | 63 | 0.909 |
| N | RXRTSN | Richardson | 25 | 58 | 0.907 |
| N | JKSN | Jackson | 41 | 107 | 0.887 |
| N | 0MPSN | Thompson | 61 | 162 | 0.887 |
| N | RBSN | Robson | 21 | 58 | 0.880 |
| N | WTSN | Watson | 67 | 185 | 0.879 |
| N | RBNSN | Robinson | 95 | 267 | 0.877 |
| N | MRMTK | Marmaduke | 27 | 78 | 0.873 |
| N | BRYN | Bryan | 29 | 91 | 0.859 |
| N | WLKNSN | Wilkinson | 25 | 84 | 0.848 |
| N E | K0BRT | Cuthbert | 96 | 226 | 0.961 |
| N E | RBSN | Robson | 21 | 58 | 0.950 |
| N E | TR0 | Dorothy | 696 | 2449 | 0.930 |
| N E | LNSLT | Lancelot | 50 | 180 | 0.928 |
| N E | BRBR | Barbara | 406 | 1470 | 0.927 |
| N E | ANBL | Annabella | 41 | 153 | 0.924 |
| N E | RLF | Ralph | 578 | 2548 | 0.907 |
| N E | FSTR | Foster | 20 | 91 | 0.904 |
| N E | 0MSN | Thomasina | 66 | 315 | 0.899 |
| N E | ISBL | Isabella | 3190 | 15396 | 0.897 |
| N Mid | STKLF | Sutcliffe | 67 | 70 | 0.986 |
| N Mid | HWR0 | Howarth | 51 | 55 | 0.977 |
| N Mid | KRNWT | Greenwood | 122 | 132 | 0.973 |
| N Mid | BT | Betty | 1816 | 2118 | 0.949 |
| N Mid | SKR | Squire | 326 | 409 | 0.923 |

| | | | | | |
|---|---|---|---|---|---|
| N Mid | HRTL | Hartley | 160 | 201 | 0.923 |
| N Mid | RFT | Wright | 263 | 333 | 0.921 |
| N Mid | NNS | Nancy | 2043 | 2948 | 0.874 |
| N Mid | HMLT | Hamlet | 57 | 84 | 0.869 |
| N Mid | LSTR | Lester | 80 | 119 | 0.861 |
| N Wales | KRF0 | Griffith | 480 | 803 | 0.973 |
| N Wales | OWN | Owen | 889 | 1869 | 0.956 |
| N Wales | KWN | Gwen | 207 | 460 | 0.952 |
| N Wales | HF | Hugh | 968 | 2835 | 0.926 |
| N Wales | IXML | Ishmael | 22 | 72 | 0.914 |
| N Wales | HMFR | Humphrey | 108 | 376 | 0.907 |
| N Wales | EFN | Evan | 665 | 3028 | 0.872 |
| N Wales | WTKN | Watkin | 26 | 124 | 0.865 |
| N Wales | LWLN | Llewellyn | 63 | 312 | 0.860 |
| N Wales | KMR | Gomer | 11 | 59 | 0.847 |
| S E | TRTN | Trayton | 56 | 69 | 0.919 |
| S E | FLTLF | Philadelphia | 93 | 126 | 0.882 |
| S E | LPLT | Leopold | 38 | 56 | 0.848 |
| S E | WR0 | Worthy | 39 | 58 | 0.844 |
| S E | ERL | Earl | 30 | 51 | 0.791 |
| S E | AKSTS | Augustus | 445 | 776 | 0.780 |
| S E | HSTR | Hester | 581 | 1023 | 0.776 |
| S E | ATLFS | Adolphus | 121 | 214 | 0.774 |
| S E | BRTRM | Bertram | 33 | 59 | 0.770 |
| S E | TKLS | Douglas | 134 | 242 | 0.767 |
| S Wales | KWNLN | Gwenllian | 234 | 247 | 0.998 |
| S Wales | JNKN | Jenkin | 201 | 216 | 0.997 |
| S Wales | HWL | Howell | 140 | 170 | 0.991 |
| S Wales | MRKN | Morgan | 529 | 649 | 0.990 |
| S Wales | EFN | Evan | 2017 | 3028 | 0.979 |
| S Wales | KMR | Gomer | 38 | 59 | 0.977 |
| S Wales | WTKN | Watkin | 71 | 124 | 0.969 |
| S Wales | ELFR | Elvira | 50 | 97 | 0.960 |
| S Wales | LWLN | Llewellyn | 150 | 312 | 0.955 |

| | | | | | |
|---|---|---|---|---|---|
| S Wales | KWN | Gwen | 214 | 460 | 0.953 |
| W | LFT | Loveday | 62 | 96 | 0.966 |
| W | WLMT | Wilmot | 42 | 101 | 0.921 |
| W | BS | Bessie | 957 | 2706 | 0.898 |
| W | NXLS | Nicholas | 342 | 1073 | 0.883 |
| W | 0MSN | Thomasina | 100 | 315 | 0.883 |
| W | TRFN | Tryphena | 68 | 246 | 0.861 |
| W | LSNT | Lucinda | 28 | 105 | 0.856 |
| W | XRT | Charity | 91 | 349 | 0.850 |
| W | NN | Nina | 120 | 519 | 0.829 |
| W | KRS | Grace | 1077 | 4728 | 0.826 |

| | Speaks Welsh | | Speaks English | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| North Wales score | 0.355 | 0.060 | −0.178 | −0.031 |
| | (0.087) | (0.017) | (0.057) | (0.010) |
| South Wales score | 0.498 | 0.266 | −0.019 | −0.020 |
| | (0.048) | (0.035) | (0.022) | (0.005) |
| Southeast England score | −0.543 | −0.246 | 0.104 | 0.040 |
| | (0.021) | (0.033) | (0.015) | (0.005) |
| Birthplace FE | | x | | x |
| N | 97932 | 97932 | 97932 | 97932 |
| $R^2$ | 0.146 | 0.507 | 0.023 | 0.171 |

This table presents evidence of the relationship between region-specific name scores and speaking Welsh or English in the 1911 census. The language question was only asked in Wales and the Isle of Mann. We exclude those recorded as speaking neither language. Data is aggregated to the name-birthplace district level, observations are weighted by the number of individuals in each cell, to give the same coefficients as an individual-level regression. In (1) and (2) the dependent variable is an indicator that the individual speaks Welsh, in (3) and (4) and indicator that they speak English. The independent variables are the name scores for the individual's name and the North Wales, South Wales and Southeast England cultural regions. Models (2) and (4) add fixed effects for the district of birth, (1) and (3) include and intercept. Standard errors clustered by district of birth in parentheses.

Table B.4: Relationship between name scores and language

| | Δ ln S E | Δ ln home culture | | Δ ln S E | Δ ln home culture | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Distance to London (100km) | 0.186 | −0.168 | −0.169 | 0.202 | −0.156 | −0.157 |
| | (0.008) | (0.016) | (0.018) | (0.009) | (0.019) | (0.027) |
| Contains coal | | | −0.607 | | | −0.379 |
| | | | (0.140) | | | (0.101) |
| Distance to London × coal | | | 0.194 | | | 0.118 |
| | | | (0.057) | | | (0.043) |
| Weighted | x | x | x | | | |
| N | 827 | 798 | 798 | 827 | 798 | 798 |
| $R^2$ | 0.545 | 0.195 | 0.239 | 0.408 | 0.103 | 0.114 |

This table presents evidence of the relationship between distance to London, coal, and the strength of the Southeast English and home cultures, using data on naming for the 1851–1860 and 1901–1910 cohorts. Regressions are at the district-level. The dependent variable in (1) and (4) is the change in the log share allocated names most associated with the Southeast English cluster, between those born 1851–1860 and 1901–1910, in (2), (3), (5) and (6) it is the change in the log share allocated names most associated with the home culture. Coal presence is an indicator that the district contains coal deposits, distance to London is distance from the district centroid to the City of London, in hundreds of kilometers. Models (1)–(3) are weighted by the number allocated names in the 1851–1860 generation, (4)–(6) are unweighted. (1) and (3) correspond to the left and right hand panels of Figure 5. All models include an intercept. Robust standard errors in parentheses.

Table B.5: Relationship between coal, distance to London, and changing popularity of cultures

| | Δ ln share home culture | |
|---|---|---|
| | (1) | (2) |
| Δ ln out-migrants / pop | −0.281 | |
| | (0.061) | |
| Δ ln in-migrants / pop | | −0.066 |
| | | (0.026) |
| N | 783 | 782 |
| $R^2$ | 0.062 | 0.027 |

This table presents the results of district-level regressions of the change in the log share assigned names most associated with the culture to which we allocate the district based on surnames before 1800, against migration and industrialization. The independent variable in (1) is the change in the log number of people over 16 living outside the cluster divided by the number born in the district, 1851–1901. In (2), it is the change in the log number of people over 16 living in the district born outside the cluster, divided by the number born in the district, 1851–1901. Observations are weighted by the number allocated namescores in the 1851–1860 cohort. Robust standard errors in parentheses.

Table B.6: Relationship between migration and the popularity of the home culture

| | Δ ln share S E culture | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Δ ln Mf workers | 0.062 | 0.038 | 0.013 | −0.011 |
| | (0.024) | (0.028) | (0.019) | (0.013) |
| Weights | 1850s names | 1900s names | 1851 pop | none |
| N | 809 | 809 | 809 | 809 |
| $R^2$ | 0.048 | 0.024 | 0.001 | 0.001 |

This table presents the results of district-level regressions in which the independent variable is the change in log manufacturing employment between 1851 and 1901 and the dependent variable the change in the log share given names associated with the Southeast cluster between the 1851–1860 and 1901–1910 generations. In (1), as in Table 1, observations are weighed by the number allocated names in the 1851–1860 cohort, (2) weights by the number allocated names in the 1901–1910 cohort, (3) by 1851 population, and (4) is unweighted. Robust standard errors in parentheses.

Table B.7: Non-Robustness of Relationship Between Manufacturing Growth and Popularity of Southeast Culture

| | Observed cultural change ($\Delta \ln \sigma$) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Predicted cultural change ($\Delta \ln \sigma$) | | | | | | |
| from real wages ($v_d^\theta$) | 0.739 | 0.305 | | | | |
| | (0.021) | (0.054) | | | | |
| from migration costs ($\delta_{od}^\theta$) | | | 0.123 | 0.260 | | |
| | | | (0.036) | (0.027) | | |
| from starting population ($n_o$) | | | | | 0.388 | 0.445 |
| | | | | | (0.026) | (0.102) |
| Cluster FE | | x | | x | | x |
| N | 7362 | 7362 | 7362 | 7362 | 7362 | 7362 |
| $R^2$ | 0.234 | 0.521 | 0.004 | 0.530 | 0.043 | 0.523 |

This table shows OLS estimates at the district-by-cluster level. The independent variable is the change in the log share choosing each culture between the counterfactual estimated using 1851 destination real wages $v_d^\theta$ in (1) and (2), migration costs in (3) and (4), and starting populations in (5) and (6), and the observed value for those born 1861–1895. The dependent variable is the change between the observed value for those born 1841–1860 and those born 1861–1895. Even-numbered models add cluster fixed effects. Standard errors clustered by district in parentheses.

Table B.8: Relationship between the change in log cultural choice shares, $\sigma$, relative to different 1851 counterfactuals

| | Change in home culture ($\Delta \ln \sigma$) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\Delta \ln$ real wage ($v_d^\theta$) | −0.032 | −0.056 | −0.216 | −0.161 | −0.063 | 0.003 |
| | (0.027) | (0.026) | (0.050) | (0.035) | (0.023) | (0.009) |
| Distance to London (100km) | | −0.392 | −0.867 | −0.567 | −0.354 | −0.112 |
| | | (0.102) | (0.123) | (0.073) | (0.065) | (0.052) |
| $\Delta \ln v_d^\theta \times$ distance | | 0.078 | 0.183 | 0.120 | 0.081 | 0.014 |
| | | (0.025) | (0.028) | (0.016) | (0.015) | (0.012) |
| Counterfactual | Realized | Realized | All | Wages | Pop. | Migration costs |
| N | 812 | 812 | 812 | 812 | 812 | 812 |
| $R^2$ | 0.002 | 0.031 | 0.083 | 0.092 | 0.027 | 0.081 |

This table shows OLS estimates at the district-level of the relationship between wage growth and the popularity of the home culture across counterfactuals. The dependent variable is the log share choosing the home culture minus the log share predicted choosing the home culture in a specific counterfactual scenario. Models (1) and (2) examine the observed change between the 1841–1860 generation and the 1861–1895 generation. (3) examines the output from the counterfactual setting real wages, starting populations and migration costs to 1851 levels, (4) just real wages, (5) just populations, and (6) just migration costs. The independent variable is the change in real wages between 1851 and 1911, interacted with distance to London. All models include an intercept. Robust standard errors in parentheses.

Table B.9: Relationship between wage growth, distance to London, and counterfactual changes in the popularity of the home culture

| | Change in log share migrating outside home culture | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\Delta \ln$ real wage ($v_d^\theta$) | −0.125 | −0.296 | 0.035 | 0.135 |
| | (0.049) | (0.021) | (0.024) | (0.018) |
| Distance to London (100km) | 0.559 | 0.433 | 0.011 | −0.127 |
| | (0.230) | (0.065) | (0.116) | (0.064) |
| $\Delta \ln v_d^\theta \times$ distance | −0.100 | −0.098 | −0.000 | 0.048 |
| | (0.052) | (0.015) | (0.026) | (0.015) |
| Counterfactual | All | Wages | Pop. | Migration costs |
| N | 828 | 828 | 828 | 828 |
| $R^2$ | 0.145 | 0.677 | 0.019 | 0.317 |

This table shows OLS estimates at the district-level of the relationship between wage growth and migration from the home culture across counterfactuals. The dependent variable is the log share migrating to locations outside the cluster to which we allocate the district based on historical surnames in the observed data, minus the predicted log share in the relevant counterfactual. (1) examines the output from the counterfactual setting real wages, starting populations and migration costs to 1851 levels, (2) just real wages, (3) just populations, and (64 just migration costs. The independent variable is the change in real wages between 1851 and 1911, interacted with distance to London. All models include an intercept. Robust standard errors in parentheses.

Table B.10: Relationship between wage growth, distance to London, and counterfactual changes in migration from the cluster

Table B.11: Data construction overview

| | (1) | (2) | (3) |
|---|---|---|---|
| **Task** | Allocating districts to clusters | Calculating name scores | Calculating origin-destination-cluster migration flows |
| **Data** | Surnames of household heads born before 1800, allocated to district of birth | First names of people born 1841–1860, allocated to cluster of birth | People born 1861–1911, allocated to clusters by first name, to origin districts by birth, to destinations by residence |
| **Source** | 1851 Census | 1911 Census | 1911 Census |

Table B.12: Estimation overview

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **Parameters** | $\delta_{od}^{\theta}$, bundled $\left(\left(v_d\xi_d^k\right)^{\theta}\left(m_d^k\right)^{\alpha}\right)$ | $\alpha, \left(\xi_d^k\right)^{\theta}, v_d^{\theta}$ | $\Omega_o^k$ | $\varphi, \psi_o^k$ | $v_{d,1851}^{\theta}, \delta_{od,1851}^{\theta}$ |
| **Estimation** | Poisson regression | Two-stage least squares | Calculated directly from (1) and (2) | Poisson regression with control function | Poisson regression |
| **Level** | Origin-destination-culture | Destination-culture | Origin-culture | Origin-culture | Origin-destination |
| **Data** | 1911 census | 1911 census | 1911 census | 1911 census | 1851 census |

## C EQUILIBRIUM UNIQUENESS

Allen, Arkolakis and Li (2020) provide conditions for equilibrium existence and uniqueness in spatial economic models. Their analysis focuses on an economic model with $N$ agents and $H$ types of interactions in which the equilibrium can be reduced to a set of $N \times H$ equations of the form

$$x_{ih} = \sum_{j=1}^{N} f_{ijh}(x_{j1}, ..., x_{jH}) \tag{7}$$

where $x_{ih}$ is the equilibrium outcome for each agent in each interaction and $f_{ijh}$ is a differentiable function that governs the interactions between agents. One can construct an $H \times H$ matrix of uniform bounds of the elasticities of these $f_{ijh}$ functions,

$$\mathbf{A}_{hh'} = \sup_{i,j} \left( \left| \frac{\partial \ln f_{ijh}}{\partial \ln x_{jh'}} \right| \right)$$

Theorem 1 of Allen, Arkolakis and Li (2020) states that if the spectral radius of this matrix—the largest absolute value of its eigenvalues—is less than one, the equilibrium exists, is unique, and can be computed by iterating equation (7).

The main theorem requires that the $f_{ijh}$ function only takes as arguments equilibrium outcomes pertaining to agent or location $j$ and returns strictly positive values. Remark 1 gives an additional result for cases where $f_{ijh}$ takes as arguments equilibrium outcomes pertaining to other agents or locations as well, and returns zeros. The results of Theorem 1 hold in such cases if one replaces the uniform bound on the elasticity with the uniform bound on the sum of elasticities.

To apply this theorem, we need to rewrite the equilibrium conditions of our model in a form consistent with equation (7).

Note that we can write (1), the number with culture $k$ migrating from $o$ to $d$, as

$$m_{od}^k = \frac{\left(v_d \delta_{od} \xi_d^k\right)^\theta \left(m_d^k\right)^\alpha}{\Omega_o^k} \frac{\left(\Omega_o^k\right)^\varphi \psi_o^k}{\sum_{l=1}^K \left(\Omega_o^l\right)^\varphi \psi_o^l} n_o = \left(v_d \delta_{od} \xi_d^k\right)^\theta \left(m_d^k\right)^\alpha n_o \frac{\left(\Omega_o^k\right)^{\varphi-1} \psi_o^k}{\sum_{l=1}^K \left(\Omega_o^l\right)^\varphi \psi_o^l}$$

we can therefore write the total number of migrants of culture $k$ in $d$ as

$$m_d^k = \sum_{o=1}^{N} \left(v_d \delta_{od} \xi_d^k\right)^\theta \left(m_d^k\right)^\alpha n_o \frac{\left(\Omega_o^k\right)^{\varphi-1} \psi_o^k}{\sum_{l=1}^K \left(\Omega_o^l\right)^\varphi \psi_o^l}$$

Defining $q_d^k = \left(m_d^k\right)^{1-\alpha}$, we can rewrite this equation as

$$q_d^k = \sum_{o=1}^{N} \left(v_d \delta_{od} \xi_d^k\right)^{\theta} n_o \frac{\left(\Omega_o^k\right)^{\varphi-1} \psi_o^k}{\sum_{l=1}^{K} \left(\Omega_o^l\right)^{\varphi} \psi_o^l}$$

Similarly we can rewrite

$$\Omega_o^k = \sum_{j=1}^{N} \left(v_j \delta_{oj} \xi_j^k\right)^{\theta} \left(q_j^k\right)^{\frac{\alpha}{1-\alpha}}$$

We can thus express our equilibrium in terms of a set of $N \times K \times 2$ equations of the form $\Omega_o^k = \sum_{j=1}^{N} \sum_{l=1}^{K} g_{oj}^{k,l}(\mathbf{q}, \mathbf{\Omega})$ and $q_o^k = \sum_{j=1}^{N} \sum_{l=1}^{K} h_{oj}^{k,l}(\mathbf{q}, \mathbf{\Omega})$, where

$$g_{oj}^{k,l} = \left(v_j \delta_{oj} \xi_j^l\right)^{\theta} \left(q_j^l\right)^{\frac{\alpha}{1-\alpha}} \mathbf{1}_{\{l=k\}},\ h_{oj}^{k,l} = \left(v_o \delta_{jo} \xi_o^l\right)^{\theta} n_j \frac{\left(\Omega_j^l\right)^{\varphi-1} \psi_j^l}{\sum_{m=1}^{K} \left(\Omega_j^m\right)^{\varphi} \psi_j^l} \mathbf{1}_{\{l=k\}}.$$

In relation to Allen, Arkolakis and Li (2020)'s notation, here we treat $N \times K$ location-by-culture pairs as the set of locations and $\Omega_o^k$ and $q_o^k$ as the two types of equilibrium outcomes. Note that the $g_{oj}^{k,l}$ function depends only on $q_l^k$, but can return zero values if $k \neq l$, and so is not consistent with equation (1) of Allen, Arkolakis and Li (2020). The $h_{oj}^{k,l}$ function depends on both $\Omega_j^l$ and $\Omega_j^m$ for other values $m \neq l$. We must apply Remark 1 of Allen, Arkolakis and Li (2020) which gives a condition related not to the uniform bound on the elasticity but to the uniform bound on the sum of the elasticities.

The relevant matrix of uniform bounds on the elasticities is then

$$\mathbf{A} = \begin{pmatrix} \sup_{o,k} \sum_{j=1}^{N} \sum_{l=1}^{K} \left| \frac{\partial \ln \sum_{m=1}^{K} \sum_{d=1}^{N} h_{od}^{k,m}}{\partial \ln q_j^l} \right| & \sup_{o,k} \sum_{j=1}^{N} \sum_{l=1}^{K} \left| \frac{\partial \ln \sum_{m=1}^{K} \sum_{d=1}^{N} h_{od}^{k,m}}{\partial \ln \Omega_j^l} \right| \\ \sup_{o,k} \sum_{j=1}^{N} \sum_{l=1}^{K} \left| \frac{\partial \ln \sum_{m=1}^{K} \sum_{d=1}^{N} g_{od}^{k,m}}{\partial \ln q_j^l} \right| & \sup_{o,k} \sum_{j=1}^{N} \sum_{l=1}^{K} \left| \frac{\partial \ln \sum_{m=1}^{K} \sum_{d=1}^{N} g_{od}^{k,m}}{\partial \ln \Omega_j^l} \right| \end{pmatrix}$$

For the top right-hand corner, note that $\sum_{m=1}^{K} \sum_{d=1}^{N} h_{od}^{k,m} = \sum_{d=1}^{N} h_{od}^{k,k} = q_o^k$

$$\frac{\partial \ln q_o^k}{\partial \ln \Omega_j^l} = \frac{\partial \ln q_o^k}{\partial \ln h_{oj}^{k,k}} \frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^l} = \frac{h_{oj}^{k,k}}{q_o^k} \frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^l}$$

Note,

$$\frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^k} = \varphi - 1 - \sigma_j^k \varphi,\ \frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^l} = -\sigma_j^l \varphi$$

where $\sigma_j^l$ is the share choosing culture $l$ in location $j$. The sum of these absolute values is then

$$\sum_{j=1}^{N} \sum_{l=1}^{K} \left| \frac{\partial \ln q_o^k}{\partial \ln \Omega_j^l} \right| = \sum_{j=1}^{N} \frac{h_{oj}^{k,k}}{q_o^k} \left( \sum_{l=1}^{K} \left| \frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^l} \right| \right)$$

Writing out the term in parentheses

$$\sum_{l=1}^{K}\left|\frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^l}\right| = \left|\varphi - 1 - \sigma_j^k\varphi\right| + \sum_{l\in\{1,\ldots,K\setminus k\}} \sigma_j^l\varphi$$

Given that $\sum_{l\in\{1,\ldots,K\setminus k\}} \sigma_j^l = 1-\sigma_j^k$, we can rewrite this equation as

$$\sum_{l=1}^{K}\left|\frac{\partial \ln h_{oj}^{k,k}}{\partial \ln \Omega_j^l}\right| = \left|\varphi\left(1-\sigma_j^k\right) - 1\right| + \varphi\left(1-\sigma_j^k\right)$$

If $\varphi\left(1-\sigma_j^k\right) \leq 1$, this expression equals 1. If $\varphi\left(1-\sigma_j^k\right) > 1$, it equals $2\varphi\left(1-\sigma_j^k\right)-1$, which is maximized at $\sigma_j^k = 0$. This expression is therefore bounded by $\max\left(2\varphi - 1, 1\right)$ So we have

$$\sum_{j=1}^{N}\sum_{l=1}^{K}\left|\frac{\partial \ln q_o^k}{\partial \ln \Omega_j^l}\right| \leq \sum_{j=1}^{N}\frac{h_{oj}^{k,k}}{q_o^k}\max\left(2\varphi-1,1\right) = \max\left(2\varphi-1,1\right)$$

For bottom-left entry in $\mathbf{A}$,

$$\sum_{m=1}^{K}\sum_{d=1}^{N} g_{od}^{k,m} = \sum_{d=1}^{N} g_{od}^{k,k} = \Omega_o^k$$

$$\frac{\partial \ln \Omega_o^k}{\partial \ln q_j^k} = \frac{\partial \ln \Omega_o^k}{\partial \ln g_{od}^{k,k}}\frac{\partial \ln g_{od}^{k,k}}{\partial \ln q_j^k} = \frac{g_{od}^{k,k}}{\Omega_o^k}\frac{\alpha}{1-\alpha}$$

Note that $\frac{\partial \ln \Omega_o^k}{\partial \ln q_j^l} = 0$ for $l \neq k$. Then

$$\sum_{j=1}^{N}\sum_{l=1}^{K}\left|\frac{\partial \ln \Omega_o^k}{\partial \ln q_j^l}\right| = \sum_{j=1}^{N}\left|\frac{\partial \ln \Omega_o^k}{\partial \ln q_j^k}\right| = \sum_{j=1}^{N}\frac{g_{od}^{k,k}}{\Omega_o^k}\frac{\alpha}{1-\alpha} = \frac{\alpha}{1-\alpha}$$

The two diagonals are zero because $\frac{\partial h_{od}^{k,m}}{\partial q_j^l} = 0$ and $\frac{\partial g_{od}^{k,m}}{\partial \Omega_j^l} = 0$ for all $o, d, k, m, j, l$.
We can therefore write

$$\mathbf{A} = \begin{pmatrix} 0 & \max(1, 2\varphi - 1) \\ \left|\frac{\alpha}{1-\alpha}\right| & 0 \end{pmatrix}$$

The spectral radius of this matrix is $\rho(\mathbf{A}) = \sqrt{\left|\frac{\alpha}{1-\alpha}\right|\max(1, 2\varphi-1)}$. The condition that $\rho(\mathbf{A}) < 1$ will be satisfied if $\frac{\alpha}{1-\alpha}\max(1, 2\varphi - 1) < 1$

## D ADDITIONAL DETAIL ON INSTRUMENTAL VARIABLES

### D.1 *Migration Flows in Table 2*

We want to regress the log share given names associated with a given culture in a location against the log share migrating from that location to districts historically associated with that culture. The endogeneity concern is that cultural choices and migration may both be influenced by unobserved preferences for the culture in question. We instrument for migration flows using coal deposits.

First, we aggregate migration data at the origin-destination level, and regress log migration flows against origin and destination fixed effects, log distance, and an indicator for the origin being the same as the destination:

$$\ln m_{od} = \gamma_d + \beta_1 \ln \text{distance}_{od} + \beta_2 \mathbf{1}_{\{o=d\}} + \gamma_o + \varepsilon_{od}$$

This gravity regression gives a set of migration cost coefficients, $\beta_1, \beta_2$, and destination fixed effects $\gamma_d$. We generate predicted destination fixed effects by regressing the estimated destination fixed effects against an indicator for coal presence:

$$\gamma_d = \delta_0 + \delta_1 \mathbf{1}_{\{d \text{ contains coal}\}} + \varepsilon_d$$

Given these various estimates, we can predict migration flows between each location:

$$\text{coal predicted } m_{od} = \frac{\exp\left(\hat{\delta}_0 + \hat{\delta}_1 \mathbf{1}_{\{d \text{ contains coal}\}} + \hat{\beta}_1 \ln \text{distance}_{od} + \hat{\beta}_2 \mathbf{1}_{\{o=d\}}\right)}{\sum_{j=1}^{N} \exp\left(\hat{\delta}_0 + \hat{\delta}_1 \mathbf{1}_{\{j \text{ contains coal}\}} + \hat{\beta}_1 \ln \text{distance}_{oj} + \hat{\beta}_2 \mathbf{1}_{\{o=j\}}\right)} n_o$$

where $n_o$ is the number originating in $o$.

We then add up the number predicted to migrate to districts allocated to each culture, divide by the total number of migrants, and take the natural logarithm:

$$\ln \text{coal predicted } s_d^k = \ln\left(\frac{1}{n_o} \sum_d \text{coal predicted } m_{od} \mathbf{1}_{\{\text{historic culture}(d)=k\}}\right)$$

where $\mathbf{1}_{\{\text{historic culture}(d)=k\}}$ takes a value of 1 if the culture to which district $d$ is allocated based on surnames in 1800 is $k$, and 0 if otherwise.

This gives the log share predicted to migrate to districts allocated to a given culture only due to coal deposits and proximity. To further isolate the component from coal deposits, we follow Borusyak and Hull (2020) and recenter the instrument. We randomly permute the vector of coal deposits 1,000 times. On each permutation, we calculate predicted migration

flows and log migration shares. We then subtract the average of these permutate log migration shares from the instrument.

### D.2 $\varphi$ *Equation*

We construct the instrument as follows. First, we aggregate migration data at the origin-destination level, ignoring clusters, and regress log migration flows against origin and destination fixed effects, log distance, and an indicator for the origin being the same as the destination:

$$\ln m_{od} = \gamma_d + \beta_1 \ln \text{distance}_{od} + \beta_2 \mathbf{1}_{\{o=d\}} + \gamma_o + \varepsilon_{od}.$$

This gravity regression gives an alternative set of migration cost coefficients, $\beta_1, \beta_2$, and destination fixed effects $\gamma_d$. We generate predicted destination fixed effects by regressing the estimated destination fixed effects against an indicator for coal presence:

$$\gamma_d = \delta_0 + \delta_1 \mathbf{1}_{\{d \text{ contains coal}\}} + \varepsilon_d.$$

These estimates allow us to predict the migration pull to different locations, using just the presence of coal and geographic distance. We use these predicted migration incentives to construct on instrument for $\Omega_o^k$:

$$\text{coal-predicted } \Omega_o^k = \sum_{d=1}^{N} \exp\left(\hat{\delta}_0 + \hat{\delta}_1 \mathbf{1}_{\{d \text{ contains coal}\}} + \hat{\beta}_1 \ln \text{distance}_{od} + \hat{\beta}_2 \mathbf{1}_{\{o=d\}}\right) \mathbf{1}_{\{\text{historic culture}(d)=k\}}$$

where $\hat{v}_d^\theta$ is the exponential of the coal-predicted destination utility, and the indicator function on the right takes a value of 1 if the district $d$ is allocated to cluster $k$ based on surnames before 1800. The logic is that coal and distance provide pull factors unrelated to cultural sorting, while locations allocated to a specific culture due to historic surnames will have higher cultural amenities $\xi_d^k$ for that culture.

Again following Borusyak and Hull (2020) we recenter the instrument by permuting the vector of coal deposits 1,000 times, calculating $\Omega_o^k$ on each permutation using the permuted coal deposits, and subtracting the average $\ln \Omega_o^k$ over all permutations from our instrument. Recentering helps address the concern that districts close in space to a given historic culture may also have more similar cultural transmission tastes.

We then estimate $\varphi$ with the control function as follows:

$$\ln \Omega_o^k = \beta \text{ recentered} \ln \text{coal-predicted } \Omega_o^k + \gamma_o + \gamma_{k(o)}^k + e_o^k$$

$$\ln \sigma_o^k = \varphi \ln \Omega_o^k + \delta \left( \ln \Omega_o^k - \ln \hat{\Omega}_o^k \right) + \gamma_o + \gamma_{k(o)}^k + \ln \psi_o^k$$

where $\gamma_{k(o)}^k$ is a fixed effect for the surname culture cluster $o$ is located in interacted with the culture in question $k$, and $\left( \ln \Omega_o^k - \ln \hat{\Omega}_o^k \right)$ is the first-stage residual. The $\gamma_{k(o)}^k$ fixed effects help further address the concern that districts close to a given surname cluster might have stronger cultural transmission tastes for that cluster. We use our estimate of $\varphi$ and observed $\sigma_d^k$ and $\Omega_o^k$ to back out $\psi_o^k$.

### D.3 $\alpha$ *Equation*

The coal-related influence on cultural choices that we use to instrument for $\Omega_o^k$ can also be used to instrument for destination populations and estimate $\alpha$. We can predict migration flows from origins to destinations using coal, distance, and starting populations:

$$\text{coal-predicted } m_{od} = \frac{\exp\left( \hat{\delta}_0 + \hat{\delta}_1 \mathbf{1}_{\{d \text{ contains coal}\}} + \hat{\beta}_1 \ln \text{distance}_{od} + \hat{\beta}_2 \mathbf{1}_{\{o=d\}} \right)}{\sum_{j=1}^{N} \exp\left( \hat{\delta}_0 + \hat{\delta}_1 \mathbf{1}_{\{j \text{ contains coal}\}} + \hat{\beta}_1 \ln \text{distance}_{oj} + \hat{\beta}_2 \mathbf{1}_{\{o=j\}} \right)} n_o.$$

We then use these predicted migration flows—which do not take into account cultural differences—to predict how shocks to choices in the origins translate into destination populations:

$$\text{coal-predicted } m_d^k = \sum_{o=1}^{N} \text{coal-predicted } \Omega_o^k \cdot \text{coal-predicted } m_{od}.$$

This measure gives the weighted sum of shocks to cultural choices at the origin, weighted by the predicted number of migrants from that origin.

We also recenter this instrument following Borusyak and Hull (2020). As for the instrument for $\Omega_o^k$, we permute the vector of coal deposits 1,000 times, calculate the instrument on each permutation, and substract the average of $\ln$ coal-predicted $m_d^k$ from our instrument. When recentering the instrument, we only permute the coal deposits used to calculate coal-predicted $\Omega_o^k$, not those used to calculate migration flows. Thus the average of the permuted instruments takes into account how coal infuences migration to specific origins, and so subtracting that average nets out such an effect. We then estimate the following equations by two-stage least squares:

$$\ln m_d^k = \beta \text{ recentered} \ln \text{coal-predicted } m_d^k + \gamma_d + \gamma_{k(d)}^k + e_d^k$$

$$\gamma_d^k = \alpha \ln \hat{m}_d^k + \gamma_d + \gamma_{k(d)}^k + \theta \ln \xi_d^k.$$

| | ln religious distance | | ln surname similarity | | ln average distance | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Recentered coal-predicted $\ln m_d^k$ | −0.015 | | 0.734 | | −0.200 | |
| | (0.056) | | (0.584) | | (0.180) | |
| Recentered coal-predicted $\ln \Omega_o^k$ | | 0.023 | | 0.273 | | −0.115 |
| | | (0.044) | | (0.418) | | (0.107) |
| N | 7425 | 7425 | 7264 | 7264 | 7452 | 7452 |
| $R^2$ | 0.842 | 0.848 | 0.906 | 0.907 | 0.845 | 0.859 |

This table validates the exclusion restriction assumption for our instruments used to estimate $\alpha$ and $\varphi$ by examining the relationship between the instruments and religious distance and surname similarity, measures of cultural proximity that correlate with $\xi$ and $\psi$, the residuals in the second stage regressions, as well as average geographic distance to districts allocated to the culture in question. Data is at the district-cluster level. As in Table 4, observations are weighted by the relevant origin or destination population, and all models include district and cluster-by-home-cluster fixed effects. Standard errors clustered by district in parentheses.

Table D.1: Orthogonality of Instruments to Cultural Proximity

### D.4 *Validating Instruments*

The exclusion restrictions necessary for our instrumental variables estimations are that the instrument for destination populations is uncorrelated with cultural amenities ($\xi_d^k$), and that the instrument for origin expected utilities is uncorrelated with cultural transmission tastes ($\psi_o^k$). While we cannot directly test that assumption, we can examine the relationship between our instruments and measures of cultural similarity that should be correlated with the taste and amenity terms. We examine the religious distance and surname similarity measures used in Section 6.6, and the average distance to districts allocated to the cluster in question, which should pick up other forms of cultural similarity that are geographically clustered.

Table D.1 regresses these measures of cultural similarity against our instruments, as in the first stages of our instrumental variables estimation routines. In all models, the coefficient on the instrument is close to zero and not statistically significant, which raises our confidence in the exclusion restriction.

## E BOOTSTRAPPING ELASTICITIES

Table E.1 reports bootstrapped confidence intervals for the homophily ($\alpha$) and culture ($\varphi$) elasticities. In our model, there are two sources of uncertainty. First, there uncertainty from the sampling of locations into our dataset. Second, origin-destination-culture migration flows are affected by transitory shocks that we treat as an error term in the estimation of distance

elasticities and destination bundles of attributes. One can think of migration flows as a draw from a multinomial distribution. Randomness in the realization of this draw will affect our estimates.

To account for sampling uncertainty, one would normally use the nonparametric bootstrap, blocking by location. That is, one would resample locations or weights for locations to obtain a set of new datasets, and then run the entire estimation routine on each new dataset to obtain the sampling distribution of the quantity of interest. The key idea is that each resampled dataset approximates a draw from the same data generating process as the primary dataset. In our context that assumption would not be satisfied. Our model and counterfactual exercises indicate that changing the set of locations or populations alters individuals' cultural choices and migration decisions. If a bootstrap replication sampled London twice, we would expect more people to migrate there and the popularity of the southeast to increase. Observed cultural choice and migration shares that do not take into account this response would be inconsistent with the model.

Our preferred approach combines the clustered sampling aspect of the nonparametric bootstrap with the structure of our model. On each bootstrap replication we do the following:

1. Sample weights for origin locations from a Dirichlet distribution, as in a fractional random weight bootstrap
2. Multiply origin populations by these sampled weights
3. Solve for equilibrium migration flows and cultural choices given these new origin populations
4. Sample migration flows from a multinomial distribution, using the predicted share of members of culture $k$ going from $o$ to $d$ as the choice probabilities
5. Estimate gravity model on these sampled migration flows to estimate destination fixed effects and migration cost parameters
6. Calculate instruments using these sampled migration flows
7. Use calculated instruments to estimate elasticities.

Doing so 1,000 times gives a distribution of estimated elasticities over resamples. We construct confidence intervals using the 2.5th and 97.5th percentiles of these distributions in Table E.1.

| | $\ln m_d^k$ | $\gamma_d^k$ | | $\ln \Omega_o^k$ | $\ln \sigma_o^k$ | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Recentered coal-predicted $\ln m_d^k$ | 0.641 | | | | | |
| | [0.095; 2.564] | | | | | |
| $\ln m_d^k$ [$\alpha$] | | 0.604 | 0.521 | | | |
| | | [0.571; 0.634] | [0.385; 0.873] | | | |
| Recentered coal-predicted $\ln \Omega_o^k$ | | | | 0.214 | | |
| | | | | [0.070; 0.718] | | |
| $\ln \Omega_o^k$ [$\varphi$] | | | | | 2.185 | 1.803 |
| | | | | | [2.038; 2.390] | [0.844; 2.355] |
| First-stage residuals | | | | | | 0.388 |
| | | | | | | [−0.040; 3.510] |
| Model | First stage | OLS | TSLS | First stage | PPML | Control function |
| N | 7418 | 7418 | 7418 | 7452 | 7452 | 7452 |
| $R^2$ | 0.989 | 0.999 | 0.999 | 0.998 | | |

This table shows estimates of $\alpha$ and $\varphi$ with confidence intervals from a semiparametric bootstrap. All models are estimated at the district-by-culture level. Models (1) and (4) show the first stage, (2) and (5) the non-instrumented second stage and (3) and (6) the instrumented second stage. (1), (2), and (4) are estimated by OLS, (3) by TSLS, and (5) and (6) by Poisson Pseudo-Maximum Likelihood, with (6) implementing the control function approach by including the residuals from (4). In models (2) and (3) the dependent variable is the destination-by-culture fixed effect from an origin-by-destination-by-culture gravity model, in (5) and (6) it is the log share of people born in the district assigned to the culture. All models include fixed effects for the district and the culture interacted with the historic culture of the district. Standard errors are calculated through a semiparametric bootstrap clustered by location. On each bootstrap iteration we sample weights for origin locations, solve for equilibrium migration and cultural choices, draw origin-destination-culture migration flows from a multinomial distribution given these choices and run the entire estimation routine on the resulting dataset. Confidence intervals are the 2.5th and 97.5th percentiles of the bootstrap distribution.

Table E.1: Estimates of model elasticities with bootstrapped confidence intervals

## F ESTIMATING 1851 ECONOMIC FUNDAMENTALS

This appendix details how we estimate real wages and migration costs using 1851 data. In the main estimation, we use origin-destination-culture migration flows to separately estimate the real wage $v_d^\theta$ and cultural amenities $\xi_d^k$. For 1851, we do not have data on names and so cannot allocate individuals to cultures. We estimate the following gravity model using origin-destination migration flows from 1851:

$$\ln m_{od,1851} = \gamma_{d,1851} + \beta_{1,1851} \ln \text{distance}_{od} + \beta_{2,1851} \mathbf{1}_{\{o=d\}} + \gamma_{o,1851} + \varepsilon_{od,1851} \tag{8}$$

The parameters $\beta_{1,1851}$ and $\beta_{2,1851}$ pin down migration costs as a function of geographic distance and an indicator that the origin is the same as the destination, much as in our main estimation in Section 6.1.

In a model without cultural differentiation, the destination fixed effect $\gamma_{d,1851}$ would correspond to $\theta \ln v_{d,1851}$, the real wage. However, in our model agents have a preference for being in a location with other members of the same culture, so some of the migration pull to a particular location is due to the number of others migrating there. To estimate the real wage, we need to rescale these destination fixed effects. We estimate the equivalent of (8) using 1911 data, and then regress the log real wages, $\theta \ln v_d$, estimated for 1911 in Section 6.2 against $\gamma_{d,1911}$, the destination fixed effects from this 1911 gravity model. In that regression, the coefficient on the destination fixed effects is 0.43 and the $R^2$ is 0.88. We then rescale the destination fixed effects from the 1851 model by 0.43 to account for this sorting effect: $v_{d,1851}^\theta = \exp(\gamma_{d,1851})^{0.43}$

## G MODEL IMPLIED ELASTICITIES

Table G.1 examines whether model-implied elasticities moderate the relationship between changes in real wages and cultural choices. The model-implied elasticities in question, described in more detail in Section 8 are the predicted change in the popularity of a given culture in a given district in a simulation that increases the district's real wage by 1%. We take the log of this predicted change and divide by ln(1.01) so it can be interpreted as an elasticity. These elasticities should tell us whether and how much a culture grows in a district in response to economic change in the district. We test this claim by regressing the change in log cultural choices for each culture and district against the change in log real wages in that district, which we interact with the model-implied elasticity. The coefficient of interest is on the interaction between changes in real wages and the elasticity. A positive coefficient would indicate that increasing real wages leads to increases in the prevalence of the cultures with positive culture-wage elasticities. In models (1)–(3) we find this to be the case. In (5) we find that this effect also holds when subsetting to the home culture, the culture to which we allocate the district based on historical surnames. In (4) we find that wage growth in districts further from London preserved the home culture, in (6) we find that this relationship is driven by greater home-culture wage elasticities further from London.

| | Change in cultural choice $(\Delta \ln \sigma) \times 100$ | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Model-implied elasticity | −2.122 | −1.740 | −1.533 | | −8.044 | −8.022 |
| | (0.813) | (0.743) | (0.635) | | (2.581) | (2.654) |
| Distance to London (km) | | | | −0.392 | | −0.236 |
| | | | | (0.102) | | (0.098) |
| $\Delta$ ln real wage $(\Delta\, \theta \ln v_d)$ | 2.091 | 1.968 | | −5.574 | −3.830 | −2.572 |
| | (0.692) | (0.698) | | (2.575) | (2.343) | (2.065) |
| $\Delta\, \theta \ln v_d \times$ elasticity | 0.427 | 0.330 | 0.309 | | 1.867 | 1.912 |
| | (0.189) | (0.176) | (0.151) | | (0.639) | (0.656) |
| $\Delta\, \theta \ln v_d \times$ distance to London | | | | 0.078 | | 0.039 |
| | | | | (0.025) | | (0.023) |
| Clusters | All | All | All | Home | Home | Home |
| Cluster FE | | x | x | | | |
| District FE | | | x | | | |
| N | 7362 | 7362 | 7362 | 812 | 812 | 812 |
| $R^2$ | 0.006 | 0.522 | 0.586 | 0.031 | 0.029 | 0.058 |

This table shows OLS estimates at the district-by-cluster level. The dependent variable is the change in the log share choosing the culture in the district, between the 1841–1860 cohort and the 1861–1895 cohort, multiplied by 100 to aid legibility. Models (1)–(3) are estimated using all clusters and interact changes in real wages with the model-implied elasticity of $\sigma$ to $v_d^\theta$, estimated by simulating a 1% increase in $v_d^\theta$ separately for each district. The coefficient of interest is this interaction, which should be positive. Models (2) and (3) include cluster fixed effects, (3) includes district fixed effects which are collinear with the change in real wages. Models (4)–(6) subset to the culture cluster the district is allocated based on historical surnames. (4) shows that further from London, wage growth was more positively associated with the preservation of the home culture, (5) shows the interaction between wage growth and the model-implied elasticity, (6) includes both and shows that the interaction effect of being far from London is attributable to the model-implied elasticity. In (1) and (4)–(6) an intercept is not shown. Standard errors clustered by district in (1)–(3) or robust standard errors in (4)–(6) in parentheses.

Table G.1: Relationship between industrialization and cultural change, moderated by model-implied elasticities

| | General equilibrium elasticity | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Partial equilibrium elasticity | 5.650 | 344.083 | 5.523 | 344.234 |
| | (0.763) | (5.424) | (0.755) | (5.307) |
| District FE | | x | | x |
| Cluster FE | | | x | x |
| N | 7345 | 7345 | 7345 | 7345 |
| $R^2$ | 0.011 | 0.924 | 0.018 | 0.925 |

This table shows OLS estimates at the district-by-cluster level. The dependent variable is the model-implied general equilibrium elasticity of $\sigma_o^k$ to $v_o^\theta$, the independent variable is the partial equilibrium elasticity, calculated as the gravity-model predicted share of people of culture $k$ in $o$ not migrating from location $o$. Models (2) and (4) include district fixed effects, (3) and (4) cluster fixed effects. Standard errors clustered by district in parentheses.

Table G.2: Relationship between partial and general equilibrium elasticities

## H ADDITIONAL ANALYSES

### H.1 *Effects of Cultural Homogeneity on Migration and Economic Outcomes*

This section examines whether rising cultural homogeneity affected economic outcomes. Because culture-specific preferences and homophily influenced where people migrated, changing the popularity of different cultures should alter the spatial distribution of population, and, if doing so directs more people towards locations with higher real wages, increase average income ($v_d$). Table H.1 shows the results of different counterfactual simulations, primarily changing the $\psi_o^k$ matrix of place-specific preferences for choosing particular cultures. Total assimilation into the Southeast culture, achieved by setting all non-Southeast $\psi_o^k$ values to zero, increases the average $v_d^\theta$ by 0.76%.

If we think of $v_d$ as representing the real wage in location $d$, under constant returns to scale, then given an estimate of $\theta$, the elasticity of migration to wages, we can infer the effect on average wages. Tombe and Zhu (2019), using data from China, estimate elasticities of migration to wages between 1.2 and 1.6, and use 1.5 in their analysis, which would suggest that the estimate in the first row of Table H.1 corresponds to around a 0.5% increase in average wages. Morten and Oliveira (2023) estimate a migration elasticity of 4.5 in Brazil, corresponding to a 0.2% increase in average wages in response to full homogenization into the Southeast culture. Caliendo et al. (2021) estimate a much lower elasticity of 0.5—corresponding to a 1.5% increase in average wages—albeit in a dynamic model of migration in Europe. The positive effect of homogenization into the Southeast culture on economic output is due to the correlation between cultural amenities ($\xi_d^k$) specific to the Southeast culture and real wages: increasing the prevalence of the Southeast culture directs more people to high real wage locations. If we set all culture-specific destination migration preferences to 1, we find that cultural homogeneity in fact decreased average $v_d^\theta$. Figure H.1 shows that both counterfactuals redistribute population across regions, away from rural parts of the North and Wales. More moderate changes to the popularity of the Southeast cluster have more moderate effects, while rescaling $\psi_o^k$ so that the average value for all cultures is equal decreases average $v_d^\theta$.

These small effects of homogenization on economic output make sense given the paper's other findings. Distance-related migration costs that help preserve peripheral cultures by diminishing out-migration also keep workers in low-productivity peripheral regions, even in the absence of cultural sorting and preferences. The model already assumes that individuals migrate towards higher-wage locations, which limits the extent to which eliminating other factors can shift migration towards such locations.

Table H.1: Counterfactual estimates from homogenizing culture

| Counterfactual | %Δ Ave. $v_d^\theta$ | %Δ share migrant |
|---|---|---|
| Removing non-S E $\psi_o^k$ | 0.757 | 0.702 |
| Removing non-S E $\psi_o^k$ and fixing $\xi_d^k = 1$ | -0.152 | 0.249 |
| Doubling S E $\psi_o^k$ | 0.755 | 0.696 |
| Halving S E $\psi_o^k$ | -0.317 | -0.015 |
| Fixing $\psi_o^k$ to have same average for all cultures | -0.116 | 0.144 |

This table shows percentage changes in the average $v_d^\theta$ and share migrating under different counterfactual scenarios. The first replaces all cultural choice preferences $\psi_o^k$ with zero for cultures other than the Southeast. The second adds to this specification by also fixing $\xi_d^k$ to one for all destinations. The third replaces $\psi_o^k$ with double its value for the Southeast culture, the fourth with half its value. The fifth rescales $\psi_o^k$ so that the average value for all cultures is equal.

### H.2 *Measuring Transportation Costs*

In the main version of the model, transportation costs are simply a function of geographic distance and an indicator that the origin is the same as the destination. In this section we explore directly calculating transportation costs using shapefiles for railways, stations, shipping routes, and ports.

We use data on railways in 1851 and 1881 from Satchell et al. (2023) and Marti-Henneberg et al. (2023), stations in 1851 and 1881 from Marti-Henneberg et al. (2018) and Henneberg et al. (2018), and ports and shipping routes from Alvarez-Palau et al. (2019).

Our method closely follows Donaldson and Hornbeck (2016) in constructing a matrix of transportation costs between Registration District Centroids. With this data for 1851 and 1881, we create a network database in which nodes are locations—district centroids, stations, and ports—and edges are railways, shipping routes, and land routes connecting them. In addition to railways and shipping we allow any two nodes within 100km of each other to be connected by land. We weight these edges using the transportation cost parameters from Donaldson and Hornbeck (2016) and then use Dijkstra's algorithm to calculate the least cost path between any two district centroids. One should note that Donaldson and Hornbeck are studying trade costs in the US during the same period, and so it is possible that the relative cost of different modes of transport for shipping goods differs from the relative cost for transporting people.

Figure H.2 plots these calculated transport costs against distance in 1851 and 1881. As expected, transportation costs strongly correlate with geographic distance. The expansion of the railway network served to reduce transportation costs at intermediate-level distances.

Table H.2 shows the relationship between these transportation costs and migration flows. These models regress log migration flows against log transportation costs, with origin-year and destination-year fixed effects. In the context of the gravity model of migration, one

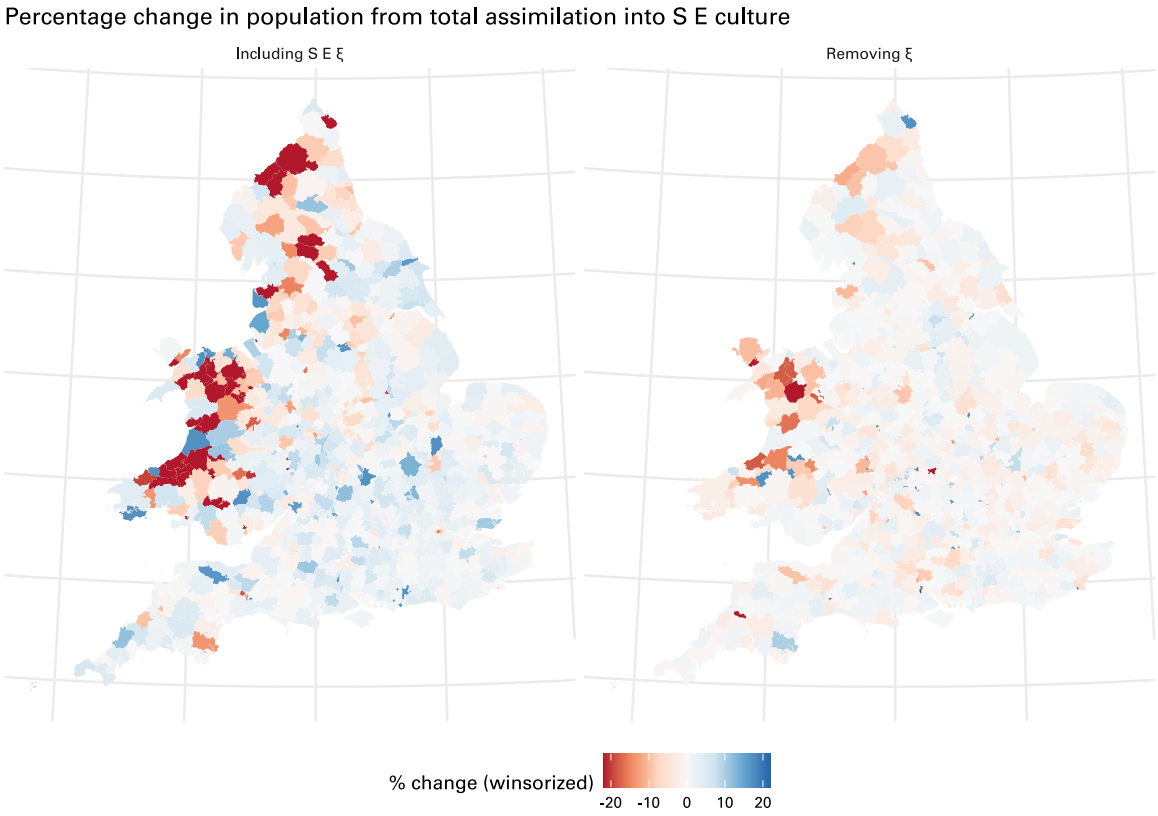


Figure H.1: Changes in population distribution from homogenizing culture

These maps show the percentage change in destination populations under counterfactual simulations that replace cultural preference terms $\psi_o^k$ for all cultures except the Southeast culture with zero. The figure on the right also replaces all culture-specific migration preferences $\left(\xi_d^k\right)^\theta$ with 1.

can think of the resulting estimates as indicating the relationship between transportation costs and the migration costs implied by migration flows. In model (1) we find a strong negative relationship between transportation costs and migration flows, consistent with the idea that transportation costs influenced migration costs. However, in (2) we find that controlling for geographic distance shrink the coefficient on transportation costs to zero. The implication is that (1) picks up the effects of distance and not a distinct effect of infrastructure. In (3) we add origin-destination fixed effects and similarly find a null effect of transportation costs on migration flows. Note that in all these specifications, we would expect the coefficient on transportation costs to be biased away from zero in the negative direction due to the endogeneity of infrastructure construction: demand for migration from an origin to a destination could drive investment in connections between the two locations.

Our conclusion from this exercise is that geographic distance provides a better measure of migration costs than measures of transportation costs. In addition to possible error in the

measurement of transportation costs, it is plausible that distance captures something related to ease of access or familiarity that is distinct from the physical cost of transportation.

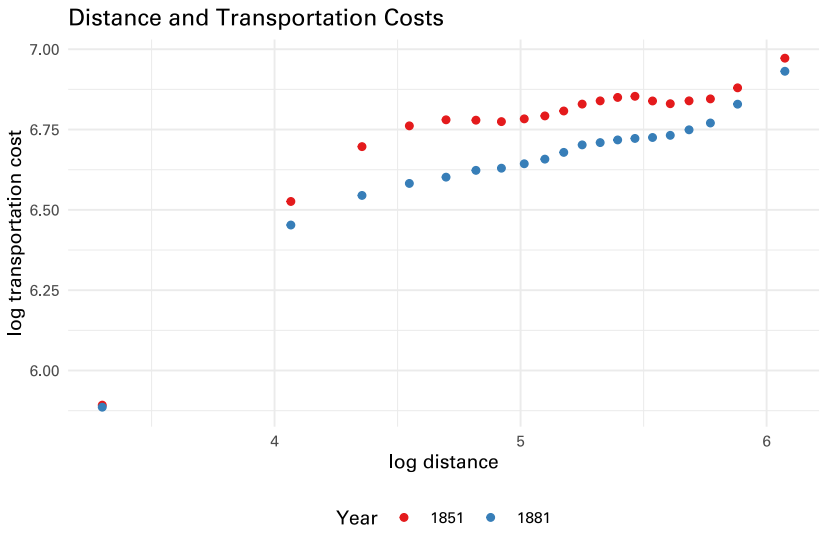


Figure H.2: Distance and Transport Costs, 1851 and 1881

Binned scatterplot of log transportation costs (y axis) in 1851 and 1881 against log distance.

### H.3 *Relationship Between Amenities and Wages*

Do cultural amenities capture just taste, or also local wage differentials? If members of a given culture are paid more than members of other cultures in a given location, that would induce members of that culture in particular to migrate there, and would be picked up by the cultural amenity parameter in our model. Such a wage difference could emerge if there is demand for specific skills in a location, which members of some cultures but not others tend to have, or if there is taste-based discrimination against members of certain cultures in the labor market.

Table H.3 examines this claim, using HISCAM scores (Lambert et al., 2013). These provide a measure of the socioeconomic status of each occupation. Our dependent variable is the average HISCAM score, at the culture-district level, for men born 1861–1895. A higher value implies that men from that culture living in that location tended to have higher status occupation. Our data does not allow us to examine whether men were paid more to do the same job. We regress average HISCAM scores against real wages ($v_d^\theta$), cultural amenities ($\xi_d^k$), and transmission tastes ($\psi_o^k$), weighting by the number of individuals in each cell. This specification gives equivalent coefficients to an individual-level regression. We find an ambiguous relationship between cultural amenities and transmission tastes, and HISCAM

| | log migrants | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| log transportation cost | −2.422 | 0.105 | −0.089 |
| | (0.098) | (0.081) | (0.060) |
| log distance | | −1.743 | |
| | | (0.071) | |
| Origin-year FE | x | x | x |
| Destination-year FE | x | x | x |
| Origin-destination FE | | | x |
| N | 1382742 | 1382742 | 795729 |

This table presents evidence of the relationship between transportation cost, distance, and migration flows in the 1851 and 1881 censuses. Models are estimated by Poisson Pseudo-Maximum Likelihood. Data is at the origin-destination-year level. The dependent variable is the log number migrating from the origin to the destination in that year's census. The independent variable of interest is the log transportation cost. All models include origin-year and destination-year fixed effects. Model (2) controls for the log distance between the origin and destination. (3) adds origin-destination fixed effects, which are collinear with distance. Standard errors clustered by origin and destination in parentheses.

Table H.2: Null relationship between transport costs and migration

scores: in two specifications the taste parameter is positively correlated with status, in two they are negatively correlated, though only the negative coefficients are statistically significant. These results provides evidence against the idea that cultural amenities capture positive local wage differentials.

A long tradition in urban economics dating back to Roback (1982) argues that amenities that affect workers' preferences but not firm productivity should lead to lower wages. The logic is that there is a compensating differential: if in equilibrium the marginal worker is indifferent across locations, a factor that makes a location more desirable will attract more workers and push down wages. While we do not explicitly model a culture-by-location process that would generate such a compensating differential, Bryan and Morten (2019) suggest one way to reconcile a negative relationship between cultural amenities and wages with our theoretical and empirical framework. In their model of migration, the individual-by-destination Frèchet shock (the equivalent of $\varepsilon_{id}$ in our model) is interpreted as a location-specific skill. A higher amenity value for members of a given culture in a given location serves to attract members of that culture with lower Frèchet shocks for that location, who have less-valuable location-specific skills and earn less.

| | Average economic status | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| log real wage ($v_d^\theta$) | 0.312 | | 0.283 | |
| | (0.364) | | (0.368) | |
| log cultural amenity ($\xi$) | −5.109 | 0.332 | | |
| | (1.285) | (0.303) | | |
| log cultural transmission taste ($\psi$) | | | 0.733 | −0.809 |
| | | | (0.439) | (0.122) |
| Cluster FE | x | x | x | x |
| District FE | | x | | x |
| N | 7385 | 7385 | 7375 | 7375 |
| $R^2$ | 0.080 | 0.979 | 0.071 | 0.981 |

This table shows OLS estimates at the district-cluster level. The dependent variable is the average economic status inferred using the HISCAM score corresponding to each occupation, for men born 1861–1895 resident in the district, allocated to clusters by their first names. In models (1) and (2) the independent variable is the destination (log) cultural amenity ($\xi$), in (3) and (4) the (log) cultural transmission taste ($\psi$). All models include cluster fixed effects. Models (1) and (3) also include the destination (log) real wage ($v_d^\theta$), (2) and (4) add district fixed effects which are collinear with real wages. Observations are weighted by the number of individuals for which occupational status data is available Standard errors clustered by district in parentheses.

Table H.3: Relationship between cultural amenities, real wages, transmission tastes, and economic status

### H.4 *Accounting for Emigration*

During this period, vast numbers of English and Welsh people emigrated, especially to the US, Canada, Australia, and New Zealand. These emigrants are not in our dataset, and to the best of our knowledge, there is no comprehensive data on emigrants. The question is how emigration affects our results. Here we show that our theory and evidence related to internal migration is unaffected by the existence of emigration. The potential for emigration does affect cultural choices, but in our framework it is captured by the location-by-culture cultural transmission tastes. Our counterfactual exercises must be interpreted as holding the ratio of domestic to international migration opportunities fixed. This assumption is appropriate given we are interested in examining how changing patterns of industrialization and internal migration affect identity choice, not in how changing international opportunities affect cultural choices.

Suppose that in addition to the locations we observe, there are a number of unobserved locations, to which people can also migrate. In the main body of the paper, locations 1 through $N$ are domestic. Let us add $F - N$ foreign locations, so that there are $F$ total locations, where 1 through $N$ are domestic and $N + 1$ through $F$ are foreign.

Denote the number of people born in $o$ assigned to culture $k$, including those who emigrate, with $n_o^k + e_o^k$, where $e_o^k$ refers to the number of emigrants from $o$ of culture $k$, and $n_o^k$, as in our analysis, is the number originating in $o$ of culture $k$ who remain in England and Wales. Then the number migrating to each location $d$ from domestic locations (Equation (1)) becomes

$$m_{od}^k = \frac{\left(v_d \delta_{od} \xi_d^k\right)^\theta \left(m_d^k\right)^\alpha}{\sum_{f=1}^F \left(v_f \delta_{of} \xi_f^k\right)^\theta \left(m_f^k\right)^\alpha} \left(n_o^k + e_o^k\right). \tag{9}$$

Note that the number from $o$ of culture $k$ who remain in England and Wales, $n_o^k$, is the sum over domestic locations:

$$n_o^k = \sum_{j=1}^N m_{oj}^k = \frac{\sum_{j=1}^N \left(v_j \delta_{oj} \xi_j^k\right)^\theta \left(m_j^k\right)^\alpha}{\sum_{f=1}^F \left(v_f \delta_{of} \xi_f^k\right)^\theta \left(m_f^k\right)^\alpha} \left(n_o^k + e_o^k\right),$$

which we rearrange as follows

$$\frac{n_o^k}{n_o^k + e_o^k} \frac{\sum_{f=1}^F \left(v_f \delta_{of} \xi_f^k\right)^\theta \left(m_f^k\right)^\alpha}{\sum_{j=1}^N \left(v_j \delta_{oj} \xi_j^k\right)^\theta \left(m_j^k\right)^\alpha} = 1.$$

Multiplying (9) by the left-hand side of this equation gives

$$m_{od}^k = \underbrace{\frac{\left(v_d\delta_{od}\xi_d^k\right)^\theta \left(m_d^k\right)^\alpha}{\sum_{j=1}^N \left(v_j\delta_{oj}\xi_j^k\right)^\theta \left(m_j^k\right)^\alpha} n_o^k}_{m_{od}^k \text{ in (1)}} \underbrace{\frac{\sum_{f=1}^F \left(v_f\delta_{of}\xi_f^k\right)^\theta \left(m_f^k\right)^\alpha}{\sum_{f=1}^F \left(v_f\delta_{of}\xi_f^k\right)^\theta \left(m_f^k\right)^\alpha} \frac{n_o^k + e_o^k}{n_o^k + e_o^k}}_{=1}.$$

This exercise shows that allowing for emigration and unobserved destination locations does not affect the part of our model that analyzes migration choices and so does not affect our calculations of the relevant $v_d, \delta_{od}$ and $\xi_d^k$ components. The logic is that we are already implicitly conditioning on people remaining in England and Wales.

Counterfactuals that set wages back to 1850s levels should affect emigration, for these analyses one must in effect assume that we are holding the ratio of domestic to foreign migration opportunities fixed.

Including foreign locations does alter the expressions for cultural choices. If foreign migration opportunities also matter for cultural choices, then in addition to the domestic $\Omega_o^k$, there is an additional foreign term:

$$\Omega_o^{k,t} := \sum_{f=1}^F \left(v_f\delta_{of}\xi_f^k\right)^\theta \left(m_f^k\right)^\alpha = \Omega_o^k + \Omega_o^{k,f}.$$

Here $\Omega_o^{k,t}$ is the culture-specific expected utility inclusive of emigration, and

$$\Omega_o^{k,f} := \sum_{f=N+1}^F \left(v_f\delta_{of}\xi_f^k\right)^\theta \left(m_f^k\right)^\alpha$$

is the expected utility just from foreign locations. Modifying the expression for cultural choices in (3) to include foreign migration opportunities gives

$$\sigma_o^k = \frac{\left(\Omega_o^{k,t}\right)^\varphi \psi_o^k}{\sum_{l=1}^K \left(\Omega_o^{l,t}\right)^\varphi \psi_l^k}.$$

Defining the ratio of total to domestic culture-specific expected utility as $\omega_o^k := \frac{\Omega_o^{k,t}}{\Omega_o^k}$, we have

$$\sigma_o^k = \frac{\left(\Omega_o^k\right)^\varphi \left(\omega_o^k\right)^\varphi \psi_o^k}{\sum_{l=1}^K \left(\Omega_o^l\right)^\varphi \left(\omega_o^l\right)^\varphi \psi_l^k}.$$

Taking logarithms as in (5) to create an estimating equation, we see that the error term in

this regression now includes the $\omega_o^k$ term:

$$\ln \sigma_o^k = \varphi \ln \Omega_o^k - \underbrace{\ln \left( \sum_{l=1}^{K} \left(\Omega_o^l\right)^{\varphi} \left(\omega_o^l\right)^{\varphi} \psi_l^k \right)}_{\text{place FE}} + \underbrace{\varphi \ln \omega_o^k + \ln \psi_o^k}_{\text{Error}}.$$

The implication is that our estimates of the cultural transmission taste $\psi_o^k$ also include the ratio of total to domestic migration opportunities, $\omega_o^k$. If it were easier to practice one culture in the United States than others, then locations with greater ease of migration to the United States would appear to have a stronger cultural transmission taste for that culture. Our counterfactuals should then be interpreted as providing the estimated effects of changes in the domestic economy, holding fixed the ratio of domestic to foreign opportunities.

### H.5 *Alternative Elasticities*

A key result in this paper is that the change in cultural choices going from a counterfactual estimated using 1851 economic fundamentals to the observed 1911 data matches specific observed cultural changes. We show that the culture of the Southeast of England expands, especially in places further from London, while peripheral home cultures decline, an effect moderated by the presence of coal. Here we investigate whether the ability of our model to reproduce these patterns is sensitive to the estimated homophily ($\alpha$) and culture ($\varphi$) elasticities.

For an arbitrary pair of elasticities, we back out the other parameters of the model, and solve for counterfactual cultural choices under 1851 real wages, populations, and migration costs. Figure H.3 plots the equivalent of Figure 10A at each of these pairs of elasticities, Figure H.4 plots the equivalent of Figure 11A. Note that in both figures the scale of the y axis varies across rows. We only include pairs of elasticities such that $\alpha\varphi < 1$, because for values above that bound the equilibrium is not unique and so simulation results are hard to interpret. The broad takeaway is that the qualitative predictions of an increase in the Southeast culture further from London, and a decline in the home culture further from London moderated by the presence of coal hold across non-zero values of these elasticities. However, the magnitude of the predicted changes, and the steepness of the slope linking distance to London and predicted changes, increases considerably as we increase the magnitude of the elasticities.

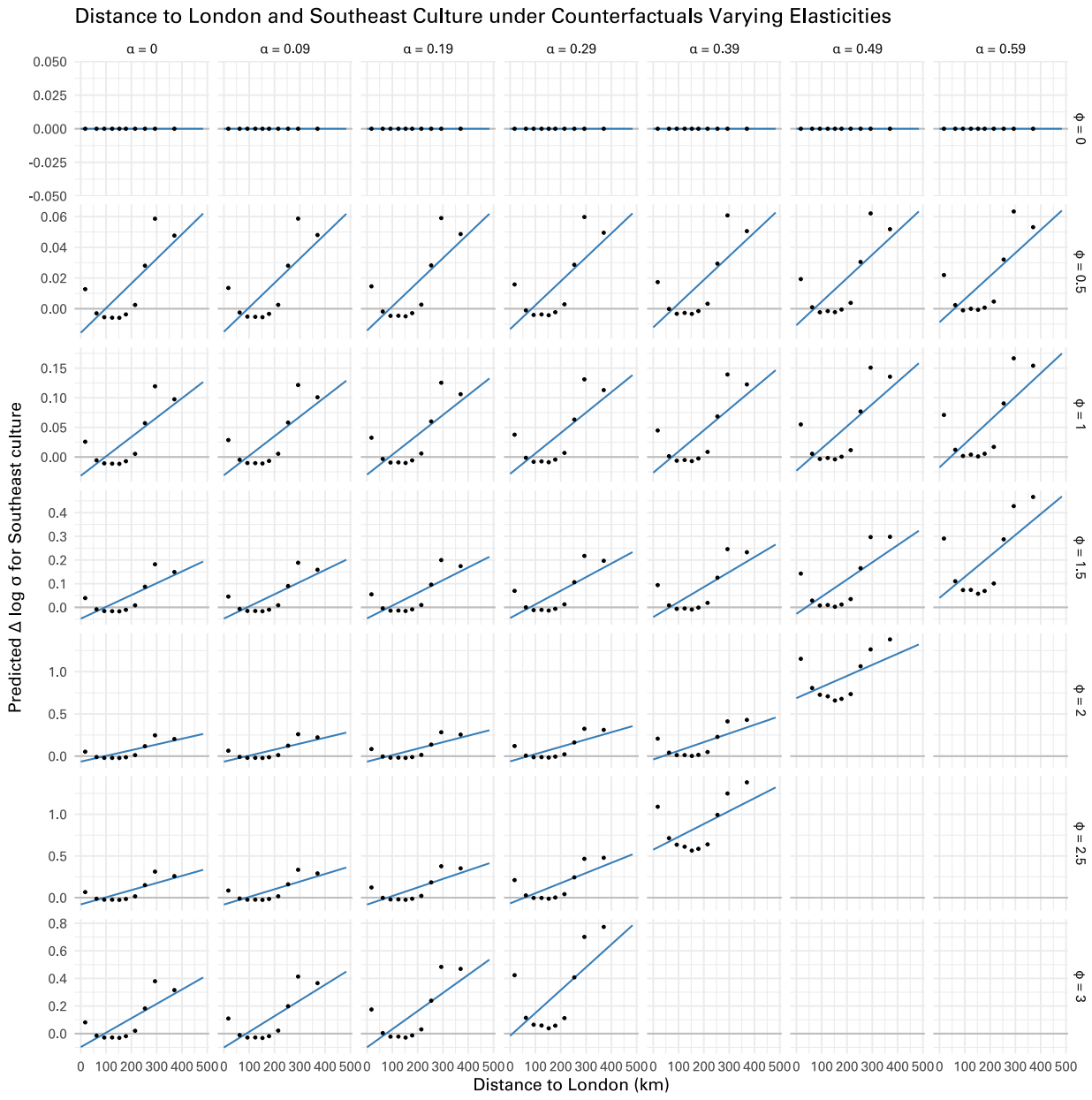


Figure H.3: Relationship between distance to London and the rise of the Southeast culture under different elasticities

Each grid cell shows the binned scatter relationship between distance to London and the log share choosing the Southeast culture in the observed data minus that in a counterfactual estimated using 1851 real wages, populations and migration costs. Each cell varies the homophily ($\alpha$) and culture elasticities ($\varphi$). Instead of estimating these elasticities using instrumental variables, we fix them at a given value and proceed to back out the other parameters of the data, then solve for counterfactuals using these assumed elasticities and backed out parameters. Columns vary $\alpha$, rows $\varphi$. Note that the scale of the y axis varies across rows. We do not calculate counterfactuals for values of $\alpha\varphi \geq 1$ as the equilibrium is non-unique and unstable.

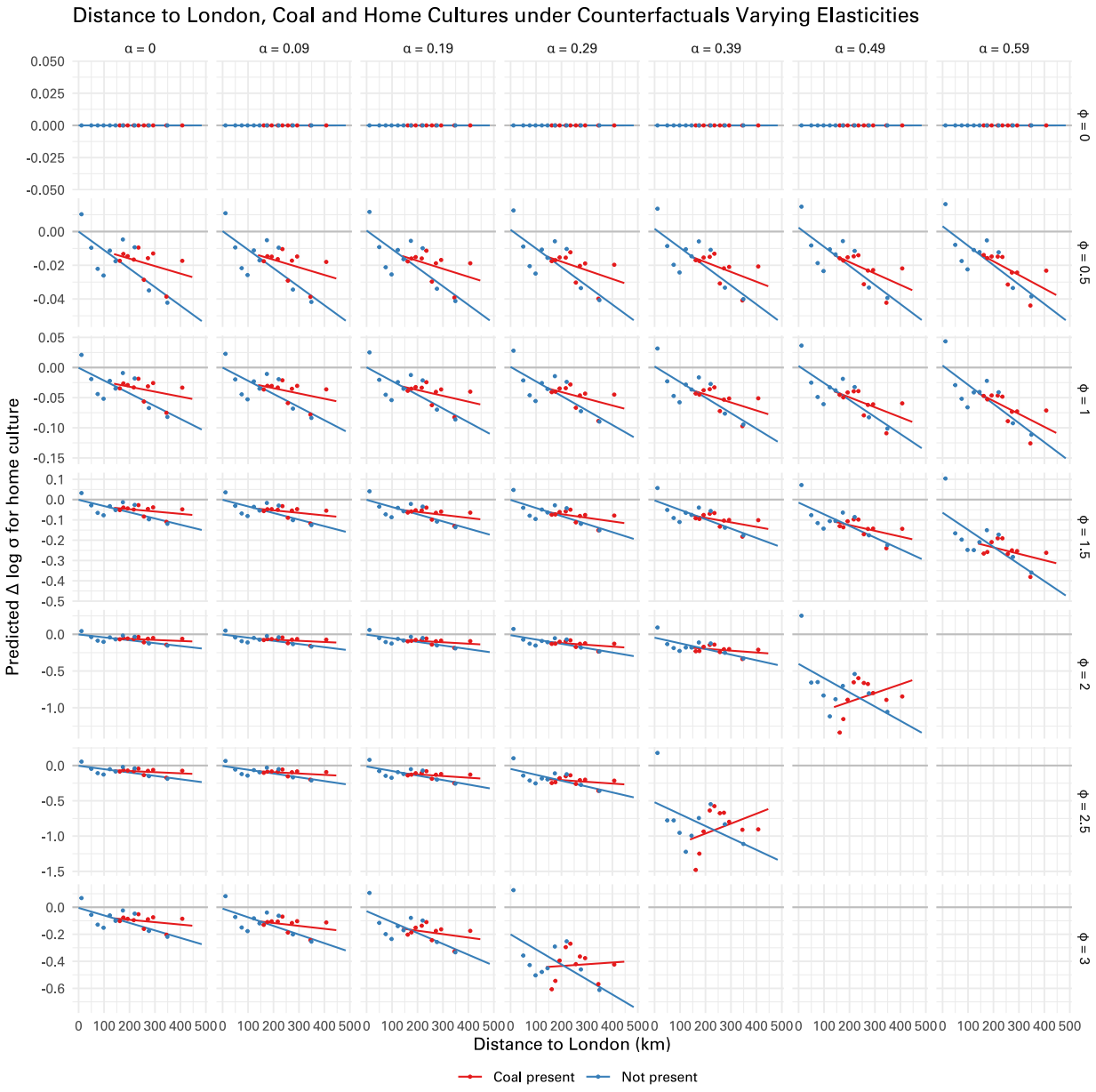


Figure H.4: Relationship between distance to London, coal, and the decline of the home culture under different elasticities

Each grid cell shows the binned scatter relationship between distance to London and the log share choosing the home culture in the observed data minus that in a counterfactual estimated using 1851 real wages, populations and migration costs, subset by whether the district contains coal. Each cell varies the homophily ($\alpha$) and culture elasticities ($\varphi$). Instead of estimating these elasticities using instrumental variables, we fix them at a given value and proceed to back out the other parameters of the data, then solve for counterfactuals using these assumed elasticities and backed out parameters. Columns vary $\alpha$, rows $\varphi$. Note that the scale of the y axis varies across rows. We do not calculate counterfactuals for values of $\alpha\varphi \geq 1$ as the equilibrium is non-unique and unstable.

### H.6 *Alternative Cultural Clusters*

In this section we examine how results from the paper change using alternative cultural clusters. We focus on two sets of clusters. First, we group counties into the regions used to coordinate civil defence during the Second World War. This grouping was made after the changes we study, but it was one of the first administrative geographies above the county level, and gives a similar number of regions to the cultural clusters we estimate. The point is not that this grouping combines culturally-similar regions, but rather provides a way of combining adjacent counties. The left panel of Figure H.5 shows these regions.

Second, we estimate clusters using data on medieval English from *A Linguistic Atlas of Late Medieval English* (Benskin et al., 2013). This source codes the appearance of linguistic features in geocoded medieval texts. We create a measure of the similarity between these sources based on the share of common linguistic features, and then run the Louvain graphical clustering algorithm on the similarity matrix. Relative to the spectral algorithm we use with surname data, the Louvain algorithm is well-suited to data that already encodes the distance between entries. It also automatically estimates the optimal number of clusters. We allocate districts to clusters based on the most common cluster among the 7 sources closest to the district. Because the data only applies to England, we allocate Wales to its own cluster. The right panel of Figure H.5 shows these clusters.

The pattern we observe in Figure 4, whereby by the start of the twentieth century, most of England was sucked into the Southeast-English culture, with exceptions in Wales and the Northeast, holds under these different clusters. Figures H.6 and H.7 both show the cluster with the highest name score among those born in each district 1851–1860 and 1901–1910, calculated using the two alternative clusters. Using civil defence regions, the hegemonic culture by 1910 is London, with linguistic cultures, the large Southern English cluster. These patterns suggest that the rise of the Southeast that we observe in the main text is primarily the rise of London.

We examine how both our reduced-form and structural findings change as we change the cultural clusters. For the reduced-form analyses in Table 2, we simply re-calculate name scores, migration flows from districts to clusters, and our coal-based instrument, using the different cluster allocations. For the structural analyses, we calculate name scores and cluster-by-migration flows using the new clusters, and calibrate the model using the elasticities we estimate in the main body of the paper. We then examine whether the cultural changes predicted by the model conform to actual cultural changes, as in Table 5. We re-run the main counterfactual simulations and examine the new predictions for the aggregate changes of the hegemonic culture (London in the case of the civil defence regions, and the large Southern cluster in the case of the linguistic clusters).

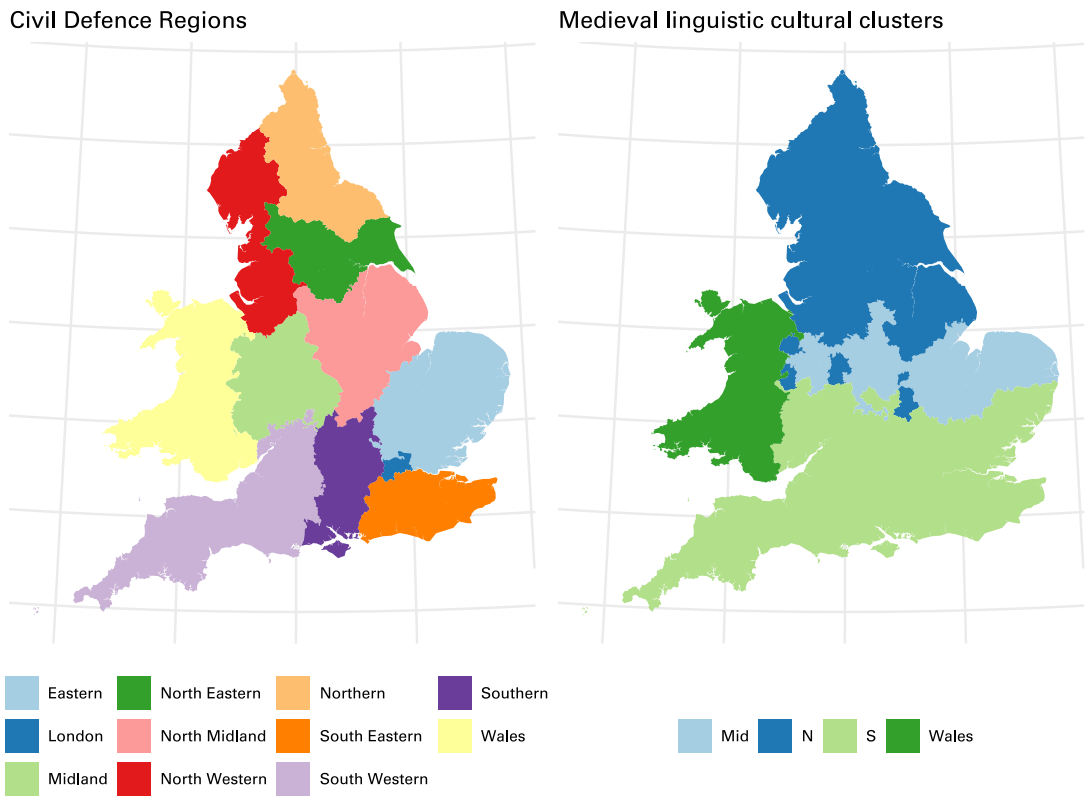


Figure H.5: Alternative cultural clusters

Broadly speaking, our results with these alternative clusters are similar to those using clusters based on historical surnames. Tables H.4 and H.5 show the reduced form analyses with these new clusters, and Tables H.6 and H.7 show the relationship between changes in predicted and actual cultural choices. The results for the civil defence regions are very close to our baseline estimates. In a few cases the results do not hold up with the linguistic clusters, but this is likely attributable to there being fewer clusters, which means there is less variation. For instance, the TSLS effect of migration to a cluster on name scores is smaller and very imprecisely estimated (Table H.5). That is to be expected as the instrument identifies off the migration pull to different clusters due to coal deposits, which affects fewer people if the clusters are larger.

In the structural analyses, the different-sized clusters lead to differences in quantitative magnitudes, and to the effects on the home culture, but other effects are qualitatively similar. Tables H.8 and H.9 show the same counterfactual outputs as Table 6, using the alternative clusters. The effects on the popularity of the Southeastern, Southern, and London cultures, are very similar across these specifications, which increases our confidence in our conclusion that

| | ln share migrating | | | ln share names | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| name score | 2.409 | 0.741 | | | |
| | (0.092) | (0.030) | | | |
| recentered coal-predicted ln share migrating | | | 4.421 | | |
| | | | (0.608) | | |
| ln share migrating | | | | 0.139 | 0.157 |
| | | | | (0.004) | (0.011) |
| Name x district FE | x | x | | | |
| District FE | | | x | x | x |
| Cluster FE | x | | x | x | x |
| Cluster x district FE | | x | | | |
| Model | OLS | OLS | First Stage | OLS | TSLS |
| First stage F-stat | | | | | 52.9 |
| N | 1033825 | 1033825 | 8990 | 8990 | 8990 |
| $R^2$ | 0.240 | 0.919 | 0.262 | 0.941 | 0.940 |

This table presents evidence of the relationship between cultural naming choices and migration, and replicates Table 2 using civil defence regions in the place of clusters based on surnames before 1800. Models (1) and (2) are estimated at the name-district-cultural cluster level: the dependent variable is the log share of people with a given name born in a given district migrating to a district in a given cultural cluster. The independent variable is the name score for that name for the destination cultural cluster. Both models include fixed effects for the name-district of birth combination, (1) includes fixed effects for the destination cluster, (2) interacts these with the district of birth. (1) and (2) are weighted by the number of people with each name born in each district. Models (3)–(5) are estimated at the district-cultural cluster level. In (4) and (5) the dependent variable is the log share given names most associated with the cultural cluster, and the independent variable is the log share of individuals migrating from the district to that cluster. In (5) this is instrumented for with the log share of migrants predicted by the location of coal deposits in a gravity model, recentered following Borusyak and Hull (2023). We permute the vector of coal deposits across district, calculate predicted log share of migrants under each permutation, and substract the mean of this from the instrument. (3) shows the first stage. (3)–(5) all include district and cluster fixed effects, and are weighted by the number of individuals with name scores born in each district. Standard errors clustered by district in parentheses.

Table H.4: Relationship between migration and naming patterns, with civil defence regions

| | ln share migrating | | | ln share names | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| name score | 2.398 | 0.587 | | | |
| | (0.119) | (0.030) | | | |
| recentered coal-predicted ln share migrating | | | 1.925 | | |
| | | | (0.544) | | |
| ln share migrating | | | | 0.127 | 0.039 |
| | | | | (0.004) | (0.032) |
| Name x district FE | x | x | | | |
| District FE | | | x | x | x |
| Cluster FE | x | | x | x | x |
| Cluster x district FE | | x | | | |
| Model | OLS | OLS | First Stage | OLS | TSLS |
| First stage F-stat | | | | | 12.5 |
| N | 565872 | 565872 | 3292 | 3292 | 3292 |
| $R^2$ | 0.488 | 0.965 | 0.509 | 0.889 | 0.845 |

This table presents evidence of the relationship between cultural naming choices and migration, and replicates Table 2 using linguistic clusters in the place of clusters based on surnames before 1800. Models (1) and (2) are estimated at the name-district-cultural cluster level: the dependent variable is the log share of people with a given name born in a given district migrating to a district in a given cultural cluster. The independent variable is the name score for that name for the destination cultural cluster. Both models include fixed effects for the name-district of birth combination, (1) includes fixed effects for the destination cluster, (2) interacts these with the district of birth. (1) and (2) are weighted by the number of people with each name born in each district. Models (3)–(5) are estimated at the district-cultural cluster level. In (4) and (5) the dependent variable is the log share given names most associated with the cultural cluster, and the independent variable is the log share of individuals migrating from the district to that cluster. In (5) this is instrumented for with the log share of migrants predicted by the location of coal deposits in a gravity model, recentered following Borusyak and Hull (2023). We permute the vector of coal deposits across district, calculate predicted log share of migrants under each permutation, and substract the mean of this from the instrument. (3) shows the first stage. (3)–(5) all include district and cluster fixed effects, and are weighted by the number of individuals with name scores born in each district. Standard errors clustered by district in parentheses.

Table H.5: Relationship between migration and naming patterns, with linguistic clusters

| | Observed cultural change ($\Delta \ln \sigma$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Predicted cultural change ($\Delta \ln \sigma$) | 0.290 | 0.278 | 0.166 | 0.065 |
| | (0.011) | (0.012) | (0.014) | (0.013) |
| District FE | | x | | x |
| Cluster FE | | | x | x |
| N | 8999 | 8999 | 8999 | 8999 |
| $R^2$ | 0.077 | 0.111 | 0.645 | 0.686 |

This table shows OLS estimates at the district-by-cluster level, replicating Table 5 using civil defence regions. The independent variable is the change in the log share choosing each culture between the counterfactual estimated using 1851 destination real wages $v_d^\theta$, starting populations, and migration costs and the observed value for those born 1861–1895. The dependent variable is the change between the observed value for those born 1841–1860 and those born 1861–1895. Model (2) adds district fixed effects, (3) adds cluster fixed effects, (4) adds both. Standard errors clustered by district in parentheses.

Table H.6: Relationship between the change in log cultural choice shares, $\sigma$, relative to the 1851 counterfactual and relative to the 1841–1860 cohort, with civil defence regions

| | Observed cultural change ($\Delta \ln \sigma$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Predicted cultural change ($\Delta \ln \sigma$) | 0.549 | 0.609 | 0.085 | −0.070 |
| | (0.025) | (0.027) | (0.013) | (0.012) |
| District FE | | x | | x |
| Cluster FE | | | x | x |
| N | 3306 | 3306 | 3306 | 3306 |
| $R^2$ | 0.199 | 0.247 | 0.802 | 0.867 |

This table shows OLS estimates at the district-by-cluster level, replicating Table 5 using linguistic clusters. The independent variable is the change in the log share choosing each culture between the counterfactual estimated using 1851 destination real wages $v_d^\theta$, starting populations, and migration costs and the observed value for those born 1861–1895. The dependent variable is the change between the observed value for those born 1841–1860 and those born 1861–1895. Model (2) adds district fixed effects, (3) adds cluster fixed effects, (4) adds both. Standard errors clustered by district in parentheses.

Table H.7: Relationship between the change in log cultural choice shares, $\sigma$, relative to the 1851 counterfactual and relative to the 1841–1860 cohort, with linguistic clusters

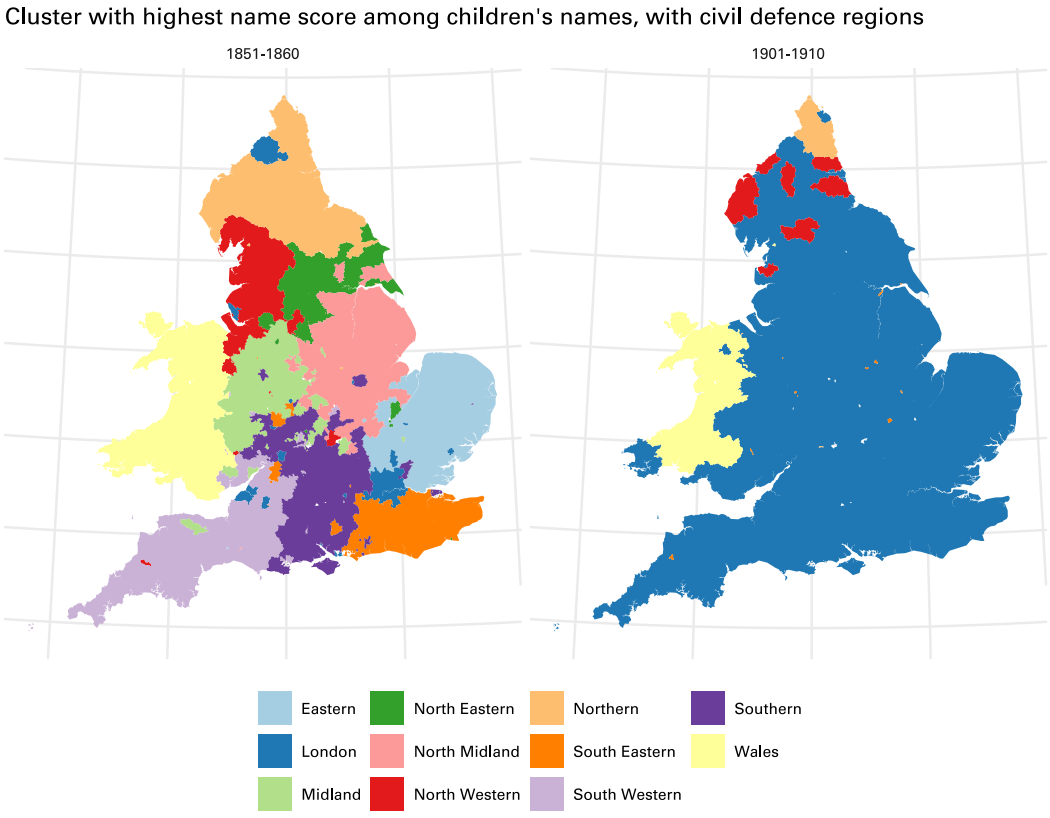


Figure H.6: Changing popularity of cultures, using civil defence regions as clusters

economic changes during the 1851–1911 period bolstered the hegemonic culture. The spatial patterns of counterfactual changes in adoption of these cultures are also similar (Figures H.8 and H.9). Our estimates for the effects on the popularity of home cultures differ in their aggregate magnitudes but follow similar spatial patterns (Figures H.10 and Figures H.11).

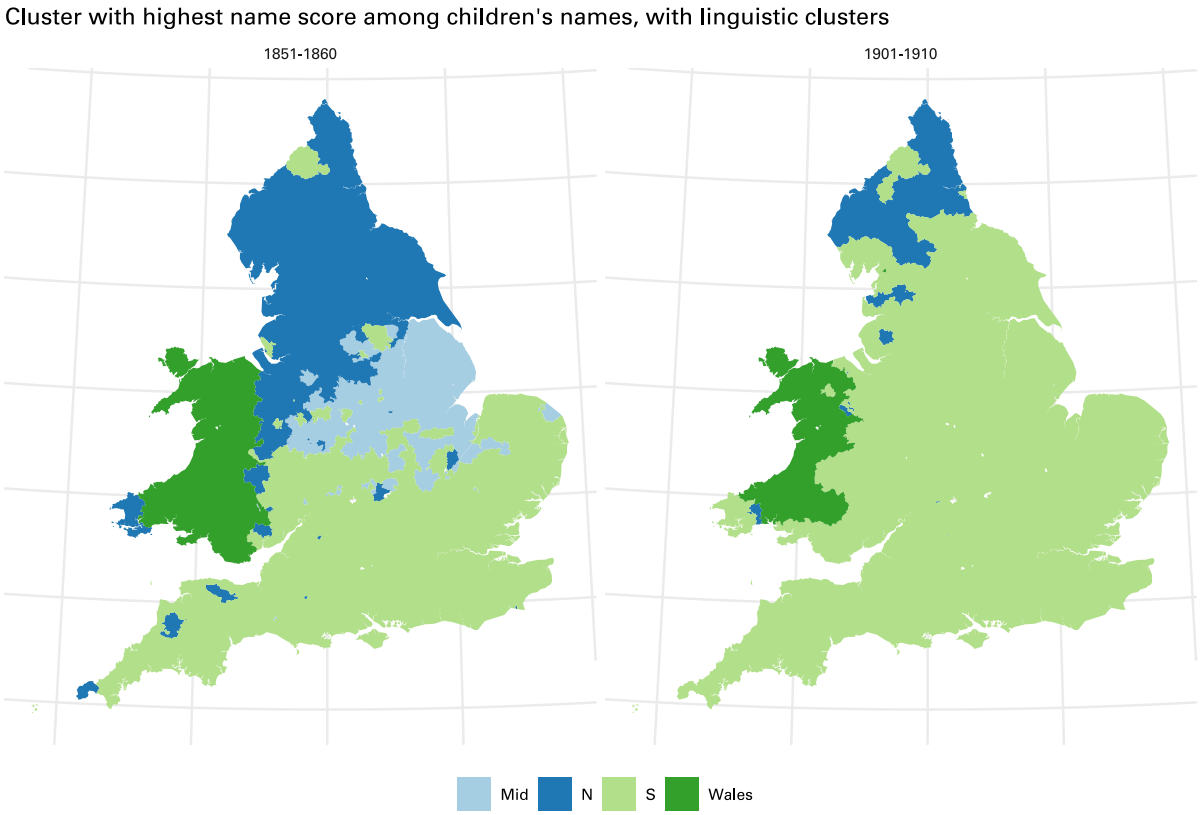


Figure H.7: Changing popularity of cultures, using linguistic clusters

Table H.8: Counterfactual estimates with civil defence regions

| | Home culture pop. | | London culture pop. | | Share migrant | |
|---|---|---|---|---|---|---|
| Counterfactual | %Δ | fixing pop. | %Δ | fixing pop. | %Δ | fixing pop. |
| All 1851 parameters | 5.926 | 7.309 | -43.122 | -40.602 | -35.698 | -31.062 |
| 1851 real wages, $v_d^\theta$ | -5.401 | -5.401 | -42.189 | -42.189 | -1.561 | -1.561 |
| 1851 starting populations | -13.544 | -5.174 | -29.714 | -22.306 | -0.727 | 0.580 |
| 1851 migration costs | 22.241 | 22.241 | -3.300 | -3.300 | -31.383 | -31.383 |

This table shows the same counterfactual outputs as Table 6, using civil defence regions in place of clusters based on surnames before 1800.

Table H.9: Counterfactual estimates with linguistic clusters

| | Home culture pop. | | S culture pop. | | Share migrant | |
|---|---|---|---|---|---|---|
| Counterfactual | %Δ | fixing pop. | %Δ | fixing pop. | %Δ | fixing pop. |
| All 1851 parameters | 3.668 | -2.359 | -20.859 | -18.306 | -35.218 | -30.776 |
| 1851 real wages, $v_d^\theta$ | -9.744 | -9.744 | -26.941 | -26.941 | -1.597 | -1.597 |
| 1851 starting populations | -1.544 | -4.509 | -14.534 | -7.274 | -0.690 | 0.605 |
| 1851 migration costs | 10.277 | 10.277 | -0.328 | -0.328 | -31.099 | -31.099 |

This table shows the same counterfactual outputs as Table 6, using linguistic clusters in place of clusters based on surnames before 1800.

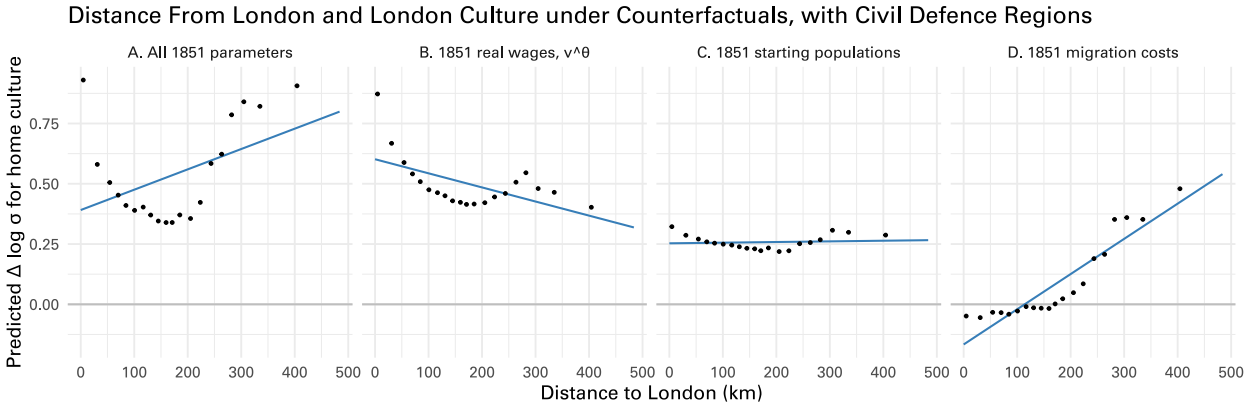


Figure H.8: Distance to London, coal, and the changing popularity of the London culture over the late 19th century, with Civil Defence Regions

This figure replicates Figure 10 using Civil Defence Regions as cultural clusters, and the London region in place of the Southeast region. The figure shows the predicted change in the log share allocated names most associated with the London culture, across different counterfactual scenarios. The y axis is the observed log share minus the counterfactual log share, the x axis distance from the City of London. Panel A uses 1851 migration costs, real wages, and starting populations to calculate the counterfactual, B only 1851 real wages, C 1851 starting populations, and D 1851 migration costs.

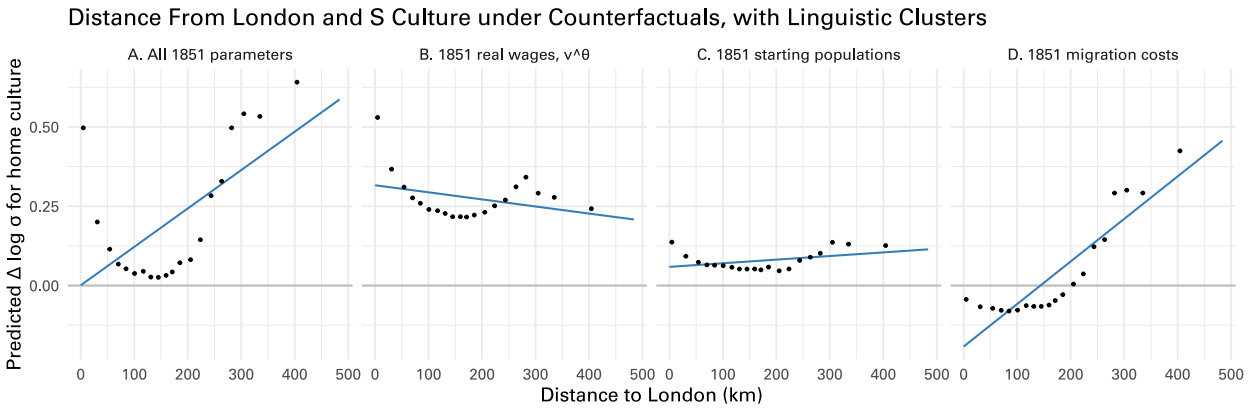


Figure H.9: Distance to London, coal, and the changing popularity of the South English culture over the late 19th century, with linguistic clusters

This figure replicates Figure 10 using linguistic clusters, and the South English cluster, containing London, in place of the Southeast English cluster. The figure shows the predicted change in the log share allocated names most associated with the Southern culture, across different counterfactual scenarios. The y axis is the observed log share minus the counterfactual log share, the x axis distance to the City of London. Panel A uses 1851 migration costs, real wages, and starting populations to calculate the counterfactual, B only 1851 real wages, C 1851 starting populations, and D 1851 migration costs.

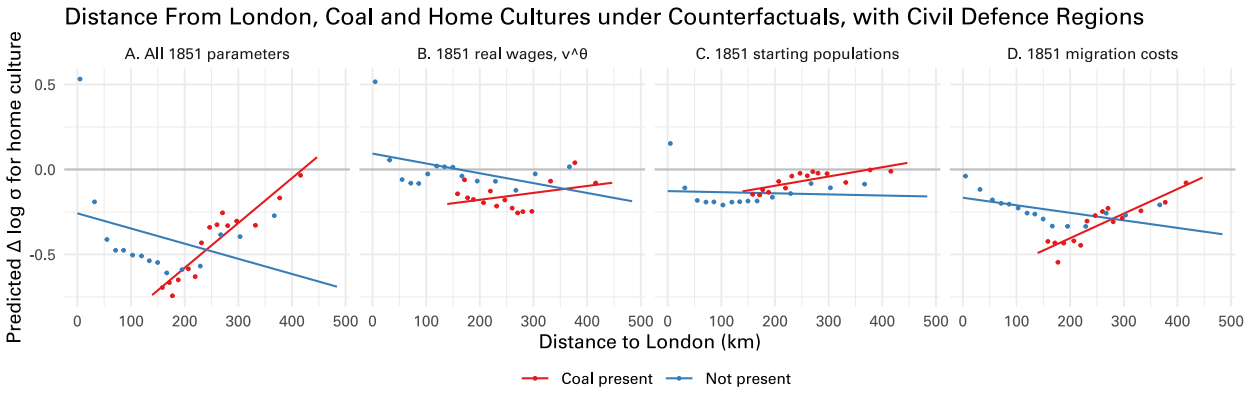


Figure H.10: Average distance, coal, and the changing popularity of home cultures over the late 19th century, with Civil Defence Regions

This figure replicates Figure 11 using Civil Defence Regions as cultural clusters. The figure shows the predicted change in the log share allocated names most associated with the home culture, across different counterfactual scenarios, subset by whether the district contains a coal deposit. The y axis is the observed log share minus the counterfactual log share, the x axis distance to the City of London. Panel A uses 1851 migration costs, real wages, and starting populations to calculate the counterfactual, B only 1851 real wages, C 1851 starting populations, and D 1851 migration costs.

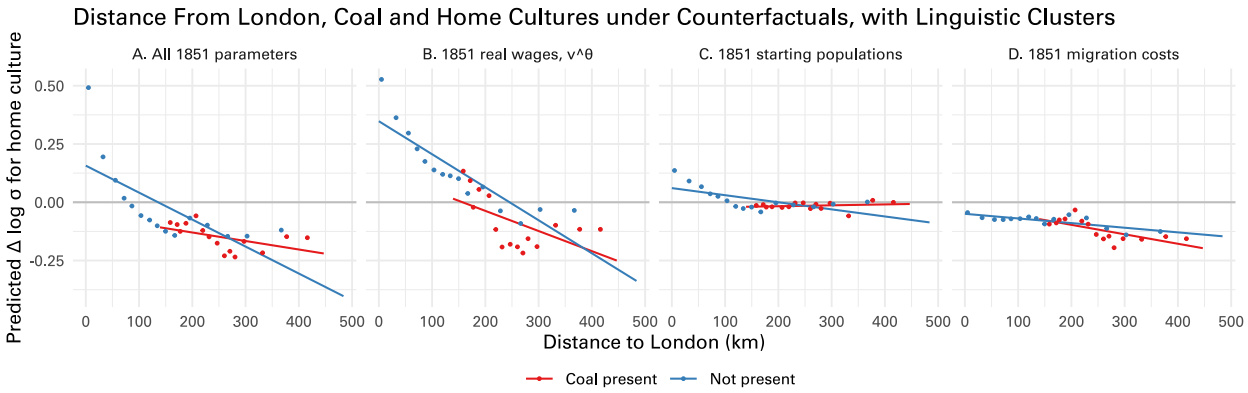


Figure H.11: Average distance, coal, and the changing popularity of home cultures over the late 19th century, with linguistic clusters

This figure replicates Figure 11 using linguistic clusters. The figure shows the predicted change in the log share allocated names most associated with the home culture, across different counterfactual scenarios, subset by whether the district contains a coal deposit. The y axis is the observed log share minus the counterfactual log share, the x axis distance to the City of London. Panel A uses 1851 migration costs, real wages, and starting populations to calculate the counterfactual, B only 1851 real wages, C 1851 starting populations, and D 1851 migration costs.