Online Appendix for "The (Structural) Gravity of Epidemics"*
-Not for Publication-

Alejandro Cuñat† Robert Zymek‡

May 2020

A Model

A.1 Pre-infection Economy

A.1.1 Bilateral Flows

Define

$$V_t(n,n') \equiv \ln\left[\frac{u_{n'}}{\delta_{nn'}} z_{n't}(n)\right] + \beta E_t\left[V_{t+1}(n')\right] \tag{1}$$

as the lifetime utility of an individual who begins day t in n and chooses to spend it in n'. Further define

$$G_{nn't}(V) \equiv \Pr\left[V_t(n,n') \leq V\right] \tag{2}$$

as the cumulative distribution function of $V_t(n,n')$. Then,

$$G_{nn't}(V) = \Pr\left[z_{n't} \leq \frac{\delta_{nn'} e^V}{u_{n'} e^{\beta E_t[V_{t+1}(n')]}}\right] = e^{-\left(\frac{\tau_{nn't}}{v_{n't}}\right)^{-\theta} e^{-\theta V}}, \tag{3}$$

where $v_{n't} \equiv u_{n'} \exp\left\{\beta E_t\left[V_{t+1}(n')\right]\right\}$.

Further define

$$m_{nn^*t} \equiv \Pr\left[V_t(n,n^*) \geq \max\left\{V_t(n,n')\,; n' \neq n^*\right\}\right] \tag{4}$$

*This work contains statistical data from the 2011 UK Census which is Crown Copyright.

†Department of Economics, University of Vienna and CESifo, Oskar Morgenstern Platz 1, Vienna 1090, Austria; alejandro.cunat@univie.ac.at.

‡School of Economics, University of Edinburgh and CESifo, 31 Buccleuch Place, Edinburgh, EH8 9JT, United Kingdom; robert.zymek@ed.ac.uk.

as the probability that an individual who begins day t in n will find it optimal to move to n^*. Then,

$$m_{nn^*t} = \int_{-\infty}^{\infty} \prod_{n' \neq n^*} G_{nn't}(V)\, dG_{nn^*t}(V) = \frac{(\tau_{nn^*t}/v_{n^*t})^{-\theta}}{\sum_{n'=1}^{N} (\tau_{nn't}/v_{n't})^{-\theta}}. \tag{5}$$

Finally, define

$$G_{nt}(V) \equiv \Pr[V_t(n) \leq V] \tag{6}$$

as the cumulative distribution function of $V_t(n)$. Then,

$$G_{nt}(V) = \prod_{n'=1}^{N} G_{nn'}(V) = e^{-e^{-\theta\left[V - \frac{1}{\theta}\ln\sum_{n'=1}^{N}\left(\frac{\tau_{nn't}}{v_{n't}}\right)^{-\theta}\right]}}. \tag{7}$$

$V_t(n)$ follows a Generalised Extreme Value (GEV or "Gumbel") distribution:

$$V_t(n) \sim GEV\left(\frac{1}{\theta}\ln\sum_{n'}(\tau_{nn'}/v_{n't})^{-\theta}, \frac{1}{\theta}, 0\right). \tag{8}$$

For any variable x such that $x \sim GEV(\mu, \sigma, 0)$, standard results imply that $E[x] = \mu + \sigma\gamma$, where γ is the Euler-Mascheroni constant.[1] Hence,

$$E_t[V_{t+1}(n)] = \frac{1}{\theta}\ln\sum_{n'=1}^{N}\left(\frac{\tau_{nn't+1}}{v_{n't+1}}\right)^{-\theta} + \frac{\gamma}{\theta}, \tag{9}$$

and

$$v_{nt} = e^{\frac{\beta\gamma}{\theta}} u_n \left[\sum_{n'=1}^{N}\left(\frac{\tau_{nn't+1}}{v_{n't+1}}\right)^{-\theta}\right]^{\frac{\beta}{\theta}}. \tag{10}$$

A.1.2 Welfare

Suppose pre-infection frictions to mobility are constant over time: $\tau_{nn't} = \tau_{nn'}\ \forall\ t$. Define average welfare as

$$\frac{V_t}{L} = \sum_{n=1}^{N}\sum_{n'=1}^{N}\frac{L_{nt}}{L} m_{nn'} E[V_t(n,n') | V_t(n,n') \geq \max\{V_t(n,n''); n'' \neq n'\}], \tag{11}$$

where $E[V_t(n,n') | V_t(n,n') \geq \max\{V_t(n,n''); n'' \neq n'\}]$ is the expected value of V_t for an individual that finds it optimal to move from n to n' at time t. It is easy to

[1] See McFadden (1978) for an exploration of the properties of the GEV distribution.

show that

$$\Pr\left[V_t\left(n,n^*\right)\leq V\middle|\, V_t\left(n,n^*\right)\geq\max\left\{V_t\left(n,n'\right);n'\neq n^*\right\}\right]=e^{-e^{-\theta\left[V-\frac{1}{\theta}\ln\sum_{n'=1}^{N}\left(\frac{\tau_{nn'}}{v_{n'}}\right)^{-\theta}\right]}}. \tag{12}$$

Thus, again using the properties of the GEV distribution,

$$E\left[V_t\left(n,n^*\right)\middle|\, V_t\left(n,n^*\right)\geq\max\left\{V_t\left(n,n'\right);n'\neq n^*\right\}\right]=\frac{1}{\theta}\ln\sum_{n'=1}^{N}\left(\frac{\tau_{nn'}}{v_{n'}}\right)^{-\theta}+\frac{\gamma}{\theta}. \tag{13}$$

Hence, average social welfare is

$$\frac{V_t}{L}=\sum_{n=1}^{N}\frac{L_{nt-1}}{L}\left(-\frac{1}{\theta}\ln m_{nn}+\ln v_n+\frac{\gamma}{\theta}\right). \tag{14}$$

Substituting from $v_n=\left[\left(e^{-\gamma}m_{nn}\right)^{-\frac{1}{\theta}}\right]^{\frac{\beta}{1-\beta}}u_n^{\frac{1}{1-\beta}}$ yields

$$\frac{V_t}{L}=\frac{1}{1-\beta}\sum_{n}\frac{L_{nt-1}}{L}\left(-\frac{1}{\theta}\ln m_{nn}+\ln u_n+\frac{\gamma}{\theta}\right). \tag{15}$$

A.2 An Epidemic Outbreak

A.2.1 General Behaviour of the Susceptible, Infected and Recovered

Susceptible, infected and recovered people in location n at the start of period t have the following Bellman equations, respectively:

$$V_t\left(n,S\right)=\max_{n'\in N}\left\{\ln\left[\frac{u_{n'}}{\delta_{nn't}}z_{n't}\left(n\right)\right]+\left(1-\pi_s\frac{S_{nt}}{L_{nt}}\right)\beta E_t\left[V_{t+1}\left(n',S\right)\right]+\right.$$

$$\left.+\pi_s\frac{S_{nt}}{L_{nt}}\beta E_t\left[V_{t+1}\left(n',I\right)\right]\right\}, \tag{16}$$

$$V_t\left(n,I\right)=\max_{n'\in N}\left\{\ln\left[\frac{u_{n'}}{\delta_{nn't}}z_{n't}\left(n\right)\right]+\left(1-\pi_r-\pi_d\right)\beta E_t\left[V_{t+1}\left(n',I\right)\right]+\right.$$

$$\left.+\pi_r\beta E_t\left[V_{t+1}\left(n',R\right)\right]\right\}, \tag{17}$$

$$V_t\left(n,R\right)=\max_{n'\in N}\left\{\ln\left[\frac{u_{n'}}{\delta_{nn't}}z_{n't}\left(n\right)\right]+\beta E_t\left[V_{t+1}\left(n',R\right)\right]\right\}, \tag{18}$$

Using the same reasoning as in the pre-infection economy, we obtain movement probabilities

$$m_{nn't}\left(J\right)=\frac{\left[\tau_{nn't}/v_{n't}\left(J\right)\right]^{-\theta}}{\sum_{n'}\left[\tau_{nn't}/v_{n't}\left(J\right)\right]^{-\theta}}, \tag{19}$$

for $J = S, I, R$, with

$$v_{n't}(S) \equiv u_{n'} e^{\beta\{(1-f_{n't})E_t[V_{t+1}(n',S)]+f_{n't}E_t[V_{t+1}(n',I)]\}}, \tag{20}$$

$$v_{n't}(I) \equiv u_{n'} e^{\beta\{(1-\pi_r-\pi_d)E_t[V_{t+1}(n',I)]+\pi_r E_t[V_{t+1}(n',R)]\}}, \tag{21}$$

$$v_{n't}(R) \equiv u_{n'} e^{\beta E_t[V_{t+1}(n',R)]}, \tag{22}$$

$$E_t[V_{t+1}(J)] = \frac{1}{\theta}\ln\sum_{n'=1}^{N}\left[\frac{\tau_{nn't+1}}{v_{n't+1}(J)}\right]^{-\theta} + \frac{\gamma}{\theta}. \tag{23}$$

This implies

$$v_{nt}(S) = e^{\frac{\beta\gamma}{\theta}} u_n \left\{\sum_{n'=1}^{N}\left[\frac{\tau_{nn't+1}}{v_{n't+1}(S)}\right]^{-\theta}\right\}^{\left(1-\pi_s\frac{S_{nt}}{L_{nt}}\right)\frac{\beta}{\theta}} \left\{\sum_{n'=1}^{N}\left[\frac{\tau_{nn't+1}}{v_{n't+1}(I)}\right]^{-\theta}\right\}^{\pi_s\frac{S_{nt}}{L_{nt}}\frac{\beta}{\theta}}, \tag{24}$$

$$v_{nt}(I) = e^{\frac{(1-\pi_d)\beta\gamma}{\theta}} u_n \left\{\sum_{n'=1}^{N}\left[\frac{\tau_{nn't+1}}{v_{n't+1}(I)}\right]^{-\theta}\right\}^{\frac{(1-\pi_r-\pi_d)\beta}{\theta}} \left\{\sum_{n'=1}^{N}\left[\frac{\tau_{nn't+1}}{v_{n't+1}(R)}\right]^{-\theta}\right\}^{\frac{\pi_r\beta}{\theta}}, \tag{25}$$

$$v_{nt}(R) = e^{\frac{\beta\gamma}{\theta}} u_n \left\{\sum_{n'=1}^{N}\left[\frac{\tau_{nn't+1}}{v_{n't+1}(R)}\right]^{-\theta}\right\}^{\frac{\beta}{\theta}}. \tag{26}$$

Clearly, $v_{nt}(R) = v_{nt}$ for all n, t. This implies $m_{nn't}(R) = m_{nn't}$ for all n, n', t.

A.2.2 Behaviour of the Susceptible, Infected and Recovered if $\pi_d \approx 0$

If $\pi_d = 0$, we can guess that $v_{nt}(S) = v_{nt}(I) = v_{nt}$ for all n, t. This is easily verified. In turn, it implies $m_{nn't}(S) = m_{nn't}(I) = m_{nn't}$ for all n, n', t. Moreover, we can linearly approximate $v_{nt}(I)$ around $\pi_d = 0$ to obtain

$$\ln v_{nt}(I) - \ln v_{nt} \simeq (1-\pi_r-\pi_d)\beta\sum_{n'=1}^{N} m_{nn'}\left[\ln v_{n't}(I) - \ln v_{n't}\right] + \pi_d(\ln u_n - \ln v_{nt}) =$$

$$= (1-\pi_r-\pi_d)\beta\sum_{n'=1}^{N} m_{nn'}\left[\ln v_{n't}(I) - \ln v_{n't}\right] - \pi_d\beta\left(\ln v_{nt} - \frac{1}{\theta}\ln m_{nnt} + \frac{\gamma}{\theta}\right), \tag{27}$$

and it is now easy to see that $v_{nt}(S) \to v_{nt}(I) \to v_{nt}$ if $\pi_d \to 0$.

A.3 Mobility Barriers and Disease Spread: Special Cases

A.3.1 Single Location

Suppose $\pi_d = 0$. Assume there is only one location. Then:

$$S_{t+1} = S_t - \tilde{I}_t, \tag{28}$$

$$I_{t+1} = (1 - \pi_r) I_t + \tilde{I}_t, \quad \tilde{I}_t = \pi_s \frac{I_t S_t}{L}, \tag{29}$$

$$R_{t+1} = R_t + \pi_r I_t, \tag{30}$$

$$L = S_t + I_t + R_t. \tag{31}$$

Equation (31) implies that we can reduce this system to two difference equations:

$$I_{t+1} - I_t = \left[\pi_s \left(1 - \frac{I_t + R_t}{L} \right) - \pi_r \right] I_t, \tag{32}$$

$$R_{t+1} - R_t = \pi_r I_t, \tag{33}$$

with $R_1 = R_0 = I_0 = 0$, $I_1 = \tilde{I}_0$. Moreover, it is possible to show that $I_\infty = 0$ and $R_\infty = \chi L$, where $\chi \in [0, 1)$ for any $\tilde{I}_0 < L$.[2]

A.3.2 Many Locations with Perfect Mobility

Suppose $\pi_d = 0$ and that there are no mobility barriers: $\tau_{nn't} = 1$ for $\forall\, n, n', t$. In this case, movement probabilities are proportional to populations: $m_{nn'} = v^\theta_{n'} / \sum_{n'} v^\theta_{n'} = L_{n'}/L$.

It follows that, as in A.3.1, the aggregate spread of the epidemic is described by (32) and (33), with $R_t \equiv \sum_n R_{nt}$, $I_t \equiv \sum_n I_{nt}$, $R_1 = R_0 = I_0 = 0$, $I_1 = \sum_n \tilde{I}_{n0}$. Perfect mobility leads to an instant proportional spread of the epidemic across local populations. As a result, the aggregate dynamics are the same in a many-location economy without mobility barriers as in the single-location economy.

Figure A.1 shows the evolution of the shares of the susceptible, infected and recovered in the total population, as well as the number of deaths per day, assuming a Covid-19-style epidemic with $\pi_d = 0$ and no mobility barriers. Under the parameterisation chosen, 85% of the population ultimately contract the disease, just as in the "do nothing" baseline scenario discussed in Section 4.1 of the paper. However, in contrast with that scenario, the epidemic runs its course much faster: infections peak on day 29 (rather than day 67), and the long-run share of recovered is mostly attained by day 90 (rather than day 120). This illustrates the difference mobility barriers can make in determining the speed at which a disease spreads across the population.

[2] For a proof, see Allen (1994).

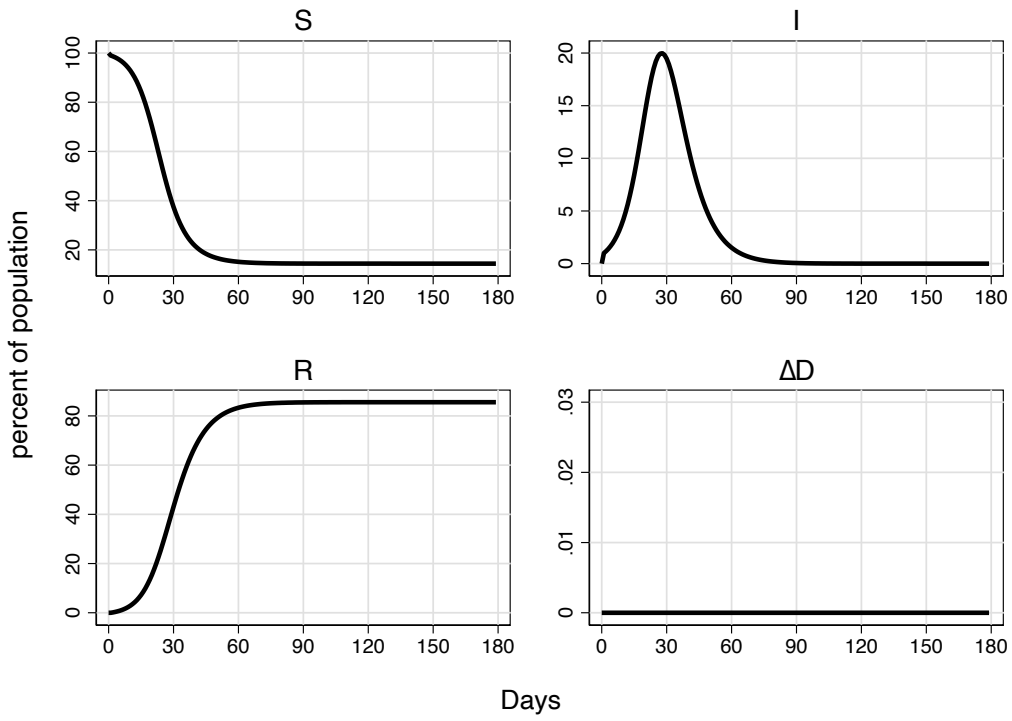

Figure A.1: Susceptible, infected, recovered and deaths per day under perfect mobility

The top-left panel plots the susceptible share of the population against time in days. The top-right panel plots the infected share of the population against time in days. The bottom-left panel plots the recovered share of the population against time in days. The bottom-right panel plots deaths per day against time in days. Graph drawn using equations (31)-(33), imposing $\pi_d = 0$, $\pi_r = 1/6.5$, $\pi_s = 2.2\pi_r$, and $\tilde{I}_0 = \sum_n \tilde{I}_{n0} = .01L$.

A.3.3 Many Locations without Mobility

Suppose $\pi_d = 0$ and there are prohibitive frictions to mobility, $\tau_{nn't} \to \infty$ for $\forall\ n \neq n', t$. This implies no flows of people between locations: $m_{nn} = m_{n'n'} = 1 - m_{nn'} = 1$ for all $n' \neq n$.

The aggregate spread of epidemic is now described by (the sum over N of)

$$I_{nt+1} - I_{nt} = \left[\pi_s \left(1 - \frac{I_{nt} + R_{nt}}{L_n}\right) - \pi_r\right] I_{nt}, \tag{34}$$

$$R_{nt+1} - R_{nt} = \pi_r I_{nt}, \tag{35}$$

with $R_{n1} = R_{n0} = I_{n0} = 0$, $I_{n1} = \tilde{I}_{n0}$. Moreover, by the same logic as in section A.3.1, $I_{n\infty} = 0$ and $R_{n\infty} = \chi L_n$, where $\chi \in [0, 1)$ for any $\tilde{I}_{n0} < L_n$.

It is easy to see that if $\tilde{I}_{n0} > 0$ for all n, $R_\infty = \sum_n R_{n\infty}$ will be the same as in the one-location/perfect mobility economy for given π_r and π_s. However, if $\tilde{I}_{n0} = 0$ for some n, then $R_\infty = \sum_n R_{n\infty}$ will be strictly smaller.

In addition, we argue in Section 2.3.2 of the paper that there is a generic concave relationship between $I_{nt} + R_{nt}$ and $\tilde{I}_{n0}$ for any t. To see why, note from (34) that growth rate of infections decreases in the share of "live" infections and in the share of

the recovered in the population. The share of people who have contracted the disease thus grows faster from a smaller base of initial infections.

B Data and Calibration

B.1 Data

B.1.1 Population, Migration and Commuting Flows from 2011 UK Census

We rely on information on population, migration and commuting from the latest UK Census, conducted in 2011.[3] The data is reported at the level of so-called "census output areas" that can be aggregated to the local-authority level using concordances provided by the UK's Office for National Statistics (ONS). The boundaries of local authorities circumscribe areas administered by a single local government, typically a local council. After aggregation, we obtain data for 378 local authorities covering all of Great Britain.[4]

Local authorities vary dramatically in terms of geographic area. The smallest ("Kensington and Chelsea") is just 1,200 hectares; the largest ("Highlands") is 2.5m hectares. However, they are more comparable in terms of population size. The median local-authority district had a population of 130,000 in 2018, with 90% of local-authority populations in the range of 60,000 to 360,000. Figure B.1 provides a map of British local authorities and their populations in 2018.

The 2011 Census reports information on all regular residents of an area in 2011 who lived at a different address one year prior. It also reports the location of an individual's usual place of residence and place of work in 2011. After aggregation this allows us to compute, for any two local authorities A and B, what share of residents of A moved to B permanently in 2010-11 and what share of residents of A commuted for work to B in 2011. Table B.1 reports the results of a "naive" and a structural gravity regression for the bilateral flows of migrants and commuters from the 2011 Census.

For the "naive" gravity regression, we regress the number of people going from A to B on a constant term, the origin and destination populations (L_{nt-1} and L_{nt} in the model) and two measures of geographic distance: the distance between the geographic mid-points of the origin and destination local authority, and a contiguity indicator taking value 1 if these local authorities share a common border. For migrants, the origin population is defined as the resident population in 2010 (equal to resident population in 2011 minus 2010-11 inward migrants plus 2010-11 outward migrants). The

[3] See Office for National Statistics (2015).

[4] Note that the ONS concordances merge the distinct local authorities of "Cornwall" and "Isles of Scilly" into the single unit "Cornwall, Isles of Scilly", and the distinct local authorities "City of London" and "Westminster" into "City of London, Westminster".

destination population is defined as the resident population in 2011. For commuters, the origin population is defined as the resident population in 2011. The destination population is defined as the local workforce in 2011 (equal to resident population in 2011 plus in-commuters minus out-commuters). For the structural gravity regression, we regress the number of people going from A to B on an A-as-origin and B-as-destination dummy, and the same distance indicators as in the "naive" regressions. All regressions are estimated using Poisson Pseudo-Maximum Likelihood (PPML).

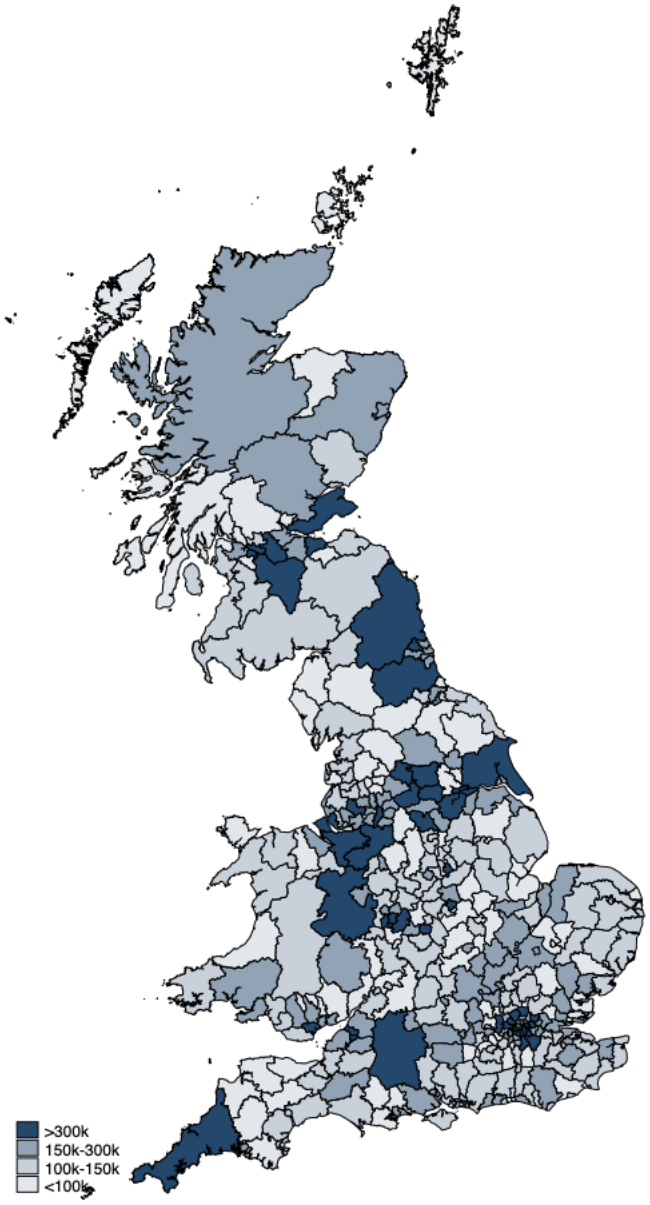

Figure B.1: 2018 map of local authorities in Great Britain, with populations

The map shows local-authority areas in Great Britain, defined consistently with 2011 UK Census concordances. Colouring reflects ONS 2018 mid-year population estimates. See Office for National Statistics (2019).

Dep. variable:	Migrants		Commuters	
# persons from origin going to destination	(1)	(2)	(3)	(4)
Ln distance (km)	-2.024***	-2.013***	-1.444***	-1.449***
	(0.014)	(0.012)	(0.014)	(0.011)
= 1 if contiguous	0.541***	0.517***	1.488***	1.610***
	(0.052)	(0.046)	(0.063)	(0.049)
Ln origin population	0.481***		0.323***	
	(0.025)		(0.069)	
Ln destination population	0.522***		0.704***	
	(0.025)		(0.075)	
Observations	142,884	142,884	142,884	142,884
Places	378	378	378	378
Adjusted R^2	.998	.996	.973	.992
Fixed effects:				
- Constant term	Yes	No	Yes	No
- Place-origin	No	Yes	No	Yes
- Place-destination	No	Yes	No	Yes

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

Table B.1: Gravity regressions on bilateral flows between GB local authorities from 2011 Census

Regressions estimated with Poisson Pseudo-Maximum Likelihood (PPML). Robust standard errors in parentheses. The dependent variable is the daily bilateral flow of people from the origin local authority to the destination local authority. This flow is calculated on the basis of UK 2011 Census data (see text for details). "Ln distance (km)" is the natural logarithm of the kilometre distance between the geographic mid-points of the origin and destination local authorities; "= 1 if contiguous" represents a dummy which takes value 1 if two local authorities share a common border, 0 otherwise; "Ln origin/destination population" is the natural logarithm of the origin/destination population. For migrants, the origin population is defined as the resident population in 2010 (equal to resident population in 2011 minus 2010-11 inward migrants plus 2010-11 outward migrants). The destination population is defined as the resident population in 2011. For commuters, the origin population is defined as the resident population in 2011. The destination population is defined as the local workforce in 2011 (equal to resident population in 2011 plus in-commuters minus out-commuters).

Across the columns of Table B.1 is is evident that gravity-style regressions can account for a large share of the observed variation in bilateral flows of people between local authorities: all columns report very high values of R^2. In addition, distance and contiguity exert a significant influence on migration and commuting patterns. For migrants, the distance elasticity − i.e. the percentage change in bilateral flows in response to a 1% difference in distance − is equal to −2.0. For commuters, it is somewhat lower at −1.4. By contrast, contiguity only raises migrant flows by a factor ($e^{.517}$ =) 1.7, but commuting flows by a factor ($e^{1.610}$ =) 5. Unsurprisingly, the distance elasticity and contiguity effect reported in Section 3.1.2 of the paper represent a linear combination of the estimates from the underlying migration and commuting data.

B.1.2 Daily Bilateral Flows of People Between Local Authorities

We combine the Census migration and commuting data to calculate a daily flow of people between any two local authorities. For migrants, we divide the 2010-11 figures by 365 to obtain the daily flow. For commuters, we first "balance" flows to reflect that commuting represents gross flows that do not cause a net change in local populations. For example, if 60 people report commuting from A to B in 2011, and 40 people report commuting from B to A, we put the potential number of people from each place who could spend the day in the other at (60+40)/2=50. We then adjust for the fact that commuters will travel between A and B only for half of the average workday. In our example, this implies that on the average day $(5/7-34/365)\times 50/2$ people go for work from A to B, and from B to A, where we assume that the average work week is 5 days and the average annual number of holidays is 34.

Adding average daily migrant and commuter flows thus constructed, we obtain our final measure of the daily bilateral flow of people − including the shares of resident populations that tend to stay within their respective local authorities on the average day. Table B.2 provides summary of the bilateral movement propensities and shares of "stayers".

$m_{nn'}$	# obs.	mean	st. dev.	min.	10th pctl.	median	90th pctl.	max.
$n \neq n'$	142,506	.00015	.00111	0	0	.00001	.0007	.00389
$n = n'$	378	.94424	.03704	.37728	.92091	.94638	.97052	.99714

Table B.2: Daily bilateral movement propensities

Daily bilateral movement propensities between 378 local authorities, based on 2011 UK Census. See text for details of data construction.

B.2 Calibration

B.2.1 Bilateral Mobility Barriers and Relative Place Values

As shown in Fally (2015), the estimated fixed effects in a PPML specification of our structural gravity regression on daily bilateral flows are consistent with the definition of the inward and outward MRTs in Section 2.1.2 of the paper, as well as the equilibrium constraints that these need to satisfy. In particular, for any t,

$$\sum_{n=1}^{N} m_{nn'} L_{nt-1} = \sum_{n=1}^{N} \hat{m}_{nn'} L_{nt-1} = L_{n't}, \text{ and } \sum_{n'=1}^{N} m_{nn'} L_{nt-1} = \sum_{n'=1}^{N} \hat{m}_{nn'} L_{nt-1} = L_{nt-1}, \tag{36}$$

$$P_{nt}^{-\theta} = \frac{L_{nt}}{L_{Nt}} e^{-\hat{\Pi}_{nt}} = L_{nt} v_{nt}^{-\theta}, \tag{37}$$

where N is the benchmark local authority (in our case, "Hartlepool").

From this, our structural gravity regression supplies us with two sets of parameter restrictions of the form

$$\frac{v_n^{-\theta}}{v_N^{-\theta}} = e^{-\hat{\Pi}_n}, \tag{38}$$

$$\tau_{nn'}^{-\theta} = (dist_{nn'})^{\hat{\phi}_1} \left(e^{contig_{nn'}}\right)^{\hat{\phi}_2}, \tag{39}$$

where we now drop the t-subscript. These parameter restrictions imply daily bilateral movement propensities:

$$\hat{m}_{nn'} = \frac{(dist_{nn'})^{\hat{\phi}_1} \left(e^{contig_{nn'}}\right)^{\hat{\phi}_2} e^{-\hat{\Pi}_{n'}}}{\sum_{n'=1}^{N} (dist_{nn'})^{\hat{\phi}_1} \left(e^{contig_{nn'}}\right)^{\hat{\phi}_2} e^{-\hat{\Pi}_{n'}}}. \tag{40}$$

Using (10), (38) and (40), we can also back out the relative flow utilities associated with different places for a given values of β and θ:

$$\frac{u_n^{-\theta}}{u_N^{-\theta}} = e^{\gamma\beta} \left(\frac{v_n^{-\theta}}{v_N^{-\theta}}\right)^{1-\beta} \left(\frac{\hat{m}_{NN}}{\hat{m}_{nn}}\right)^{\beta}. \tag{41}$$

Descriptive statistics for the relative place values, mobility barriers and relative flow utilities derived in line with (10), (39) and (41) are shown in Table B.3.

variable	# obs.	mean	st. dev.	min.	10th pctl.	median	90th pctl.	max.
$-\theta \ln (v_n/v_N)$	378	-.18	.28	-1.23	-.55	-.15	.13	.72
$-\theta \ln \tau_{nn'}$	142,884	-10.82	1.63	-14.38	-12.64	-11.06	-8.85	0
$-\theta \ln (u_n/u_N)$	378	.04	.07	-.03	-.02	.01	.11	.47

Table B.3: Relative place values, mobility barriers and relative flow utilities

Calibrated relative place values (v_n/v_N), mobility barriers ($\tau_{nn'}$) and relative flow utilities (u_n/u_N) based on structural gravity regression. To compute relative flow utilities (u_n/u_N), we impose a daily discount factor of $\beta = .96^{1/365}$. See text for details.

B.2.2 Initial Infections

To simulate an epidemic in our model, we need to "seed" initial infections, $\tilde{I}_{n0}$, in at least some locations n. We do this at the local-authority level in a manner consistent with the pattern of Covid-19 cases reported by the UK, Scottish and Welsh Governments on 10 March 2020.

Official statistics on Covid-19 cases in the UK have been released since 9 March 2020. This data can be found at https://coronavirus.data.gov.uk/ (retrieved 7 April 2020). On 10 March, 324 cases had been reported in England, 27 in Scotland, and 6 in Wales. However, data is provided at different levels of aggregation for the different countries, reflecting differences in the local administrative structure of the National Healthcare System (NHS).

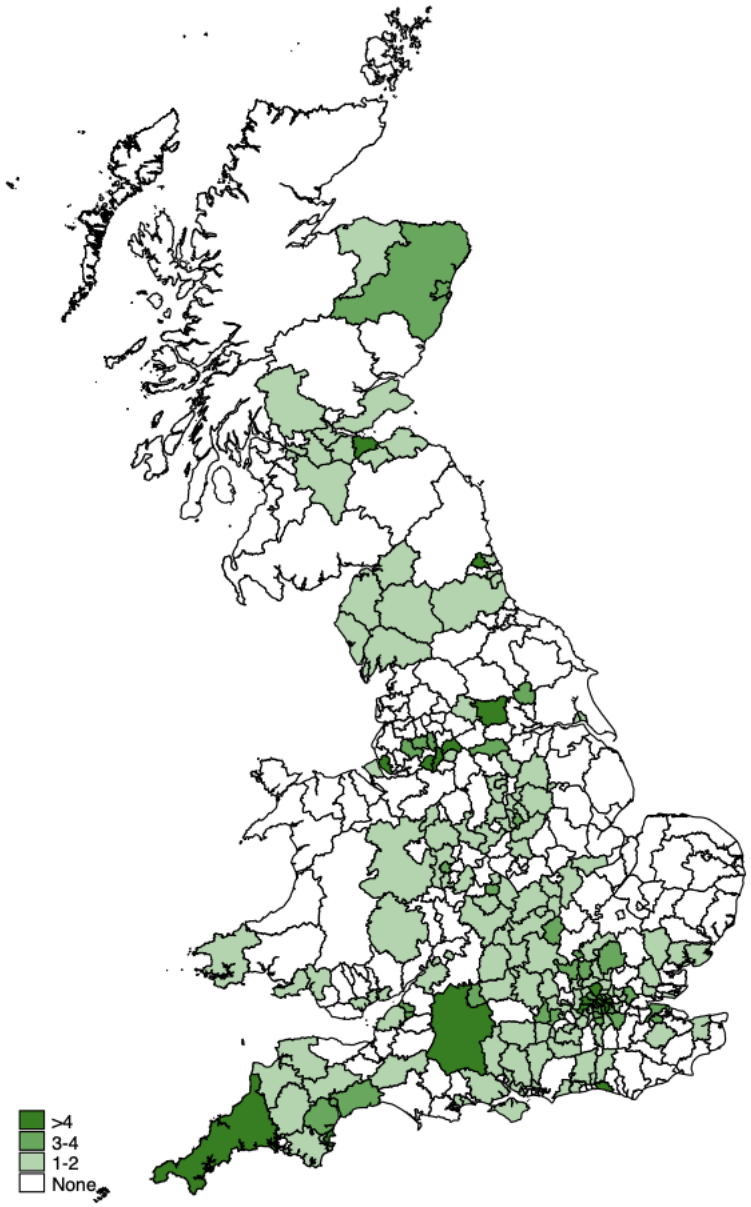

Figure B.2: Inferred Covid-19 cases at the local-authority level on 10 March 2020

The map shows local-authority areas in Great Britain, defined consistently with 2011 UK Census concordances. Colouring reflects the inferred number of Covid-19 cases on 10 March 2020, consistent with officially reported case numbers. See text for details.

The case data for England is broken down by local authority of residence, but with 15 out of 324 cases unattributed. The case data for Scotland is only provided at the country level. However, the Scottish government maintains its own Covid-19 website at https://www.gov.scot/publications/coronavirus-covid-19-daily-data-for-scotland/, where Scottish case numbers are broken down by NHS regional health board (above local-authority level). Finally, Welsh case numbers are only available for the whole of Wales

For England, we attribute the unattributed cases on 10 March to local authorities in proportion to attributed cases. For Scotland, we distribute cases reported at the level of regional health boards across local authorities in proportion to local populations. Finally, since there had been only 6 reported cases in Wales on 10 March,

these cases are easy to attribute to local authorities from national media reports.[5] The resulting inferred numbers of Covid-19 cases at the local-authority level across Great Britain is shown in Figure B.2.

Ferguson et al. (2020) cite evidence from China and repatriation flights suggesting that 40-50% of infections are not identified as cases. To reflect this, and the relative initial scarcity of Covid-19 testing in the UK, we assume that the number of cases reported at the local-authority level on 10 March reflected 30% of actual infections. Therefore, we divide the case numbers shown in Figure B.2 by .30 in order to arrive at the "actual" case numbers that constitute our seeding of initial infections.

C Simulations

C.1 Baseline: "Do Nothing"

There is a short time window during which the output from our "do nothing" baseline simulation can be evaluated against actual data. After initially adopting a fairly light-touch approach to the containment of Covid-19, the UK Government imposed a lockdown on March 23: from that time onwards, the general public was requested to leave their homes for no longer than 1 hour per day, and only for essential activities. However, during the preceding period, there was very limited containment of the Covid-19. This means we can compare the regional spread of the disease in our baseline scenario with the actual data for the period 10-23 March.

To do this, we generate the number of detected cases at the local-authority level for days 0-13 of our baseline simulation (equal to .30 times the infected plus recovered in a local authority, in line with our assumptions about detection rates in Section B.2.2). We then aggregate this number up to the 9 main administrative regions of the NHS: East of England, Midlands, London, North East and Yorkshire, North West, Scotland, South East, South West, Wales. We choose this level of aggregation because we have consistently reported official case numbers at this level for the period 10-23 March, and thus need not rely on proportionality assumptions, as in Section B.2.2.

Figure C.1 plots the log change of detected cases from our model in these regions during days 0-13 against the log change in actual cases reported by the British authorities between 10 and 23 March. As can be seen from the figure, the two are positively but far-from-perfectly correlated: the correlation is .48. London and Wales are particularly noteworthy outliers: both experienced significantly stronger case growth in practice than predicted by our simulation.

[5]For example, see https://www.bbc.co.uk/news/uk-wales-51827012 (retrieved on 16 April 2020).

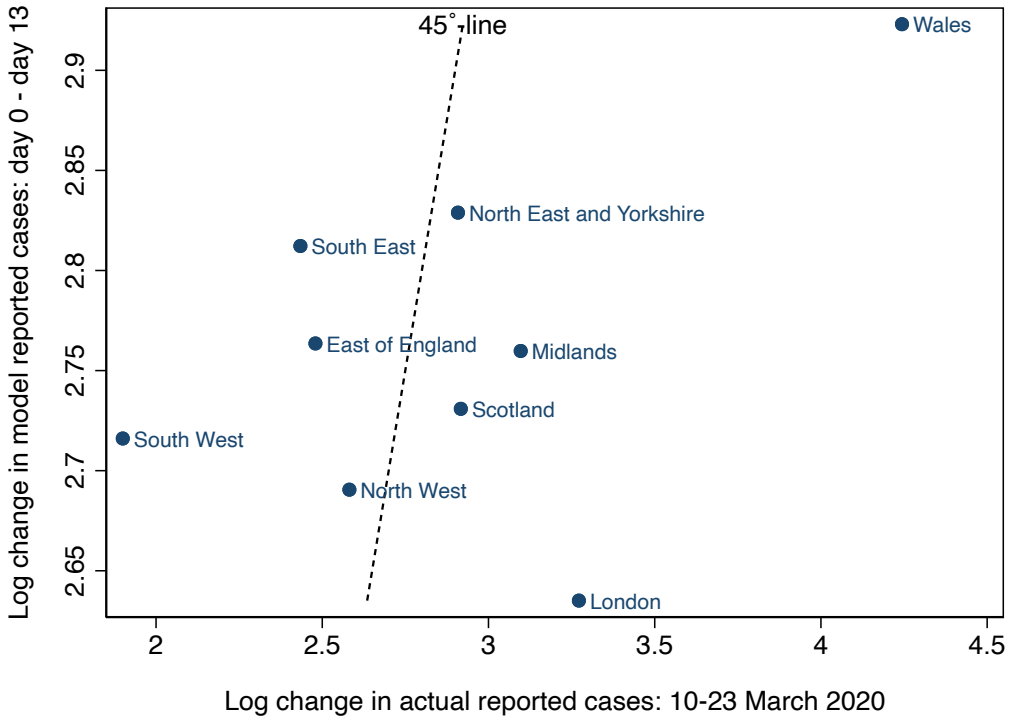

Figure C.1: Change in model-reported versus actual cases in NHS regions, 10-23 March 2020

The log change in model reported cases refers to the log change in 0.30 times the number of infected and recovered in a region according to the baseline "do nothing" simulation between days 0 and 13. The log change in actual reported cases refers to the log change in the case numbers reported by British authorities in this region between 10 March and 23 March 2020. The dashed line represents the 45-degree line. The properties of the baseline "do nothing" simulation are described in Sections 3 and 4.1 of the paper.

There are a number of reasons why our simple exercise may underpredict case growth. One may be the assumption of a constant 30% detection rate: it is likely that Covid-19 detections may have been increasing as a proportion of all infections in the UK during the early March period. Moreover, our model assumes no "exogenous" appearance of new cases after day 0. This assumption precludes, for example, the arrival of infected individuals from abroad − which may be a particular limitation for predicting case growth in London. By contrast, Ferguson et al. (2020) assume some exogenously determined new infections each period, which arrive in proportion to local populations.

References

[1] Allen, Linda J.S., 1994. "Some Discrete-Time SI, SIR and SIS Epidemic Models," *Mathematical Biosciences,* 124, pp. 83-105.

[2] Fally, Thibault, 2015. "Structural Gravity and Fixed Effects," *Journal of International Economics*, 97, 1, pp. 76-85.

[3] Ferguson, Neil M., Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C. Okell, Sabine van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick G.T. Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A. Donnelly, Steven Riley, and Azra C. Ghani, 2020. *Report 9: Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand.*

[4] McFadden, Daniel, 1978. "Modeling the Choice of Residential Location," *Transportation Research Record* 673, pp. 72–77.

[5] Office for National Statistics, 2015. *2011 Census: Flow Data* [data collection]. UK Data Service, SN: 7713.

[6] Office for National Statistics, 2019. *Population Estimates for UK, England and Wales, Scotland and Northern Ireland: Mid-2018*. Release number: MYE12.