Online Appendix for "The Making of the Modern Metropolis: Evidence from London" (Not for Publication)

Stephan Heblich*

University of Toronto

Stephen J. Redding†

Princeton University, NBER and CEPR

Daniel M. Sturm‡

London School of Economics and CEPR

April 2020

Table of Contents

*Munk School of Global Affairs & Public Policy and Dept. of Economics, 1 Devonshire Place, Toronto, ON, M5S 3K7, Canada. Tel: 1 416 946 8935. Email: stephan.heblich@utoronto.ca.

†Dept. Economics and WWS, Julis Romo Rabinowitz Building, Princeton, NJ 08544. Tel: 1 609 258 4016. Email: reddings@princeton.edu.

‡Dept. Economics, Houghton Street, London, WC2A 2AE. UK. Tel: 44 20 7955 7522. Email: d.m.sturm@lse.ac.uk.

A Introduction

This technical appendix contains additional supplementary material for the paper. Section B proves Lemma 1 in Section 6 of the paper. Section C derives the expressions for the worker choice probabilities and expected utility in Section 5 of the paper. Section D establishes a number of isomorphisms, in which we show that our quantitative predictions for the impact of the removal of the railway network on historical workplace employment and commuting patterns hold in an entire class of urban models. In Section E, we consider an alternative specification for the idiosyncratic shock to worker commuting decisions, in terms of effective units of labor instead of preferences. Although the relationship between wages in the model and data is different in this alternative specification, we show that our baseline quantitative analysis for the impact of the construction of the railway network takes a similar form as in our baseline specification in the paper. In Section F, we develop a dynamic model, in which there is a gradual response to transport infrastructure improvements, because of adjustments costs for investments in durable building capital, and show that our baseline quantitative analysis continues to apply. Section G reports additional results for our solutions for the supply of floor space, productivity and amenities in Section VII. of the paper. Section H provides further details on our counterfactuals in Section VIII. of the paper. Section I reports additional empirical results discussed in the paper. Section J contains further information about the data sources and definitions.

B Proof of Lemma 1

*Proof.*We first determine the unique vector of relative changes in wages ($\hat{\mathbf{w}}_{\mathbf{t}}$) and then recover the unique vector of relative changes in employment ($\hat{\mathbf{L}}_{\mathbf{t}}$). From equation (20) in the paper, the combined land and commuter market clearing condition for an earlier year $\tau < t$ can be written as:

$$\hat{\mathbb{Q}}_t \mathbb{Q}_t = T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}}), \tag{B.1}$$

where $\hat{\mathbb{Q}}_t$ is the observed vector of relative changes in rateable values; $\mathbb{Q}_t$ is the observed vector of rateable values in our baseline year $t = 1921$; $\mathbf{Z}_{\mathbf{t}}$ is a known matrix of relative changes in variables between years τ and t and values for variables in our baseline year t, including R_{nt}, w_{nt}, $\hat{R}_{nt}$ and $\lambda^R_{nit|n}$; $\hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}$ is the matrix of changes in commuting costs; $\hat{\mathbf{w}}_{\mathbf{t}}$ is the vector of relative changes in wages to be determined; and $T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}})$ is an operator that is defined as:

$$\begin{aligned} T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}}) \quad &= (1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it}w_{it}\right]\hat{R}_{nt}R_{nt} \\ &+\left(\frac{\beta^H}{\beta^L}\right)\hat{w}_{nt}w_{nt}\left[\sum_{i\in\mathbb{N}} \frac{\lambda^R_{int|i}\hat{w}^{\epsilon}_{nt}\hat{\kappa}^{-\epsilon}_{int}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{i\ell t|i}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{i\ell t}}\hat{R}_{it}R_{it}\right]. \end{aligned} \tag{B.2}$$

We now establish the following properties of equation (B.1) and the operator $T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}})$.

Property (i): $\hat{\mathbb{Q}}_t \mathbb{Q}_t > 0$ such that $\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} > 0$ for all $n \in \mathbb{N}$

Property (ii): $\lim_{\hat{\mathbf{w}}_{\mathbf{t}}\to 0} T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}}) = 0$.

Property (iii): $T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}})$ is monotonic in the vector of relative changes in wages ($\hat{\mathbf{w}}_{\mathbf{t}}$), since:

$$\begin{aligned} \frac{dT(\cdot)}{d\hat{\mathbf{w}}_{\mathbf{t}}}d\hat{\mathbf{w}}_{\mathbf{t}} \quad &= (1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\frac{d\hat{w}_{it}}{\hat{w}_{it}}\hat{w}_{it}w_{it}\right]\hat{R}_{nt}R_{nt} \\ &+\epsilon(1-\alpha)\left[\sum_{i\in\mathbb{N}}\left(1-\frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\right)\frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\frac{d\hat{w}_{it}}{\hat{w}_{it}}\hat{w}_{it}w_{it}\right]\hat{R}_{nt}R_{nt} \\ &+\left(\frac{\beta^H}{\beta^L}\right)\frac{d\hat{w}_{nt}}{\hat{w}_{nt}}\hat{w}_{nt}w_{nt}\left[\sum_{i\in\mathbb{N}}\frac{\lambda^R_{int|i}\hat{w}^{\epsilon}_{nt}\hat{\kappa}^{-\epsilon}_{int}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{i\ell t|i}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{i\ell t}}\hat{R}_{it}R_{it}\right] \\ &+\epsilon\left(\frac{\beta^H}{\beta^L}\right)\hat{w}_{nt}w_{nt}\left[\sum_{i\in\mathbb{N}}\left(1-\frac{\lambda^R_{int|i}\hat{w}^{\epsilon}_{nt}\hat{\kappa}^{-\epsilon}_{int}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{i\ell t|i}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{i\ell t}}\right)\frac{\lambda^R_{int|i}\hat{w}^{\epsilon}_{nt}\hat{\kappa}^{-\epsilon}_{int}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{i\ell t|i}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{i\ell t}}\frac{d\hat{w}_{nt}}{\hat{w}_{nt}}\hat{R}_{it}R_{it}\right], \end{aligned} \tag{B.3}$$

where $\frac{dT(\cdot)}{d\hat{\mathbf{w}}_{\mathbf{t}}}d\hat{\mathbf{w}}_{\mathbf{t}} > 0$ for $d\hat{\mathbf{w}}_{\mathbf{t}} > 0$.

Property (iv): $T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}})$ is homogeneous of degree one in the vector of relative changes in wages ($\hat{\mathbf{w}}_{\mathbf{t}}$) such that $T(\boldsymbol{\xi}\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}}) = \boldsymbol{\xi}T(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}})$ for any positive scalar ξ.

From properties **(i)-(iv)**, starting from $\hat{w}_{nt} = 0$ for all locations n, and increasing $\hat{w}_{nt}$ for each location n, there exists a unique value for $\hat{w}_{nt}$ for which $\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = T_n(\hat{\mathbf{w}}_{\mathbf{t}}; \hat{\boldsymbol{\kappa}}_{\boldsymbol{t}}; \mathbf{Z}_{\mathbf{t}})$ and equation (B.1) is satisfied.

Using these unique solutions for $\hat{\mathbf{w}}_t$, the unique vector of relative changes in employment ($\hat{\mathbf{L}}_t$) can be recovered from the commuter market clearing condition in equation (18) in the paper, as reproduced below.

$$\hat{L}_{it}L_{it} = \sum_{n\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{B.4}$$

where we have solved for $\hat{w}^{\epsilon}_{it}$ and we observe ($\hat{R}_{nt}$, L_{nt}, R_{nt}, $\lambda^R_{nit|n}$). Finally, the unique relative change in commuting flows ($\hat{L}_{nit}$) can be recovered from the conditional commuting probabilities in equation (21) in the paper, as reproduced below:

$$\hat{L}_{nit}L_{nit} = \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{B.5}$$

where we have solved for $\hat{w}^{\epsilon}_{it}$ and we observe ($\hat{R}_{nt}$, L_{nt}, R_{nt}, $\lambda^R_{nit|n}$). ■

C Derivation of Choice Probabilities and Expected Utility

In this section of the online appendix, we report additional results for the derivation of worker commuting probabilities and expected utility in Section V. of the paper.

C1 Distribution of Utility

From the indirect utility function in equation (3) in the paper, we have the following monotonic relationship between idiosyncratic amenities ($b_{ni}(\omega)$) and utility ($U_{ni}(\omega)$):

$$b_{ni}(\omega) = \frac{U_{ni}(\omega)\,\kappa_{ni}P^{\alpha}_nQ^{1-\alpha}_n}{B_{ni}w_i}. \tag{C.1}$$

We assume that idiosyncratic amenities ($b_{ni}(\omega)$) are drawn from an independent extreme value (Fréchet) distribution for each residence-workplace pair and each worker:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{C.2}$$

where we normalize the Fréchet scale parameter in equation (C.2) to one, because it enters worker choice probabilities isomorphically to the common bilateral amenities parameter B_{ni}.

Together equations (C.1) and (C.2) imply that the distribution of utility for residence n and workplace i is:

$$G_{ni}(u) = e^{-\Psi_{ni}u^{-\epsilon}}, \qquad \Psi_{ni} \equiv (B_{ni}w_i)^{\epsilon}\left(\kappa_{ni}P^{\alpha}_nQ^{1-\alpha}_n\right)^{-\epsilon}. \tag{C.3}$$

From all possible pairs of residence and workplace, each worker chooses the bilateral commute that offers the maximum utility. Since the maximum of a sequence of Fréchet distributed random variables is itself Fréchet distributed, the distribution of utility across all possible pairs of residence and workplace is:

$$1 - G(u) = 1 - \prod_{k\in\mathbb{M}}\prod_{\ell\in\mathbb{M}} e^{-\Psi_{k\ell}u^{-\epsilon}},$$

where the left-hand side is the probability that a worker has a utility greater than u, and the right-hand side is one minus the probability that the worker has a utility less than u for all possible pairs of residence and employment locations.

Therefore we have:

$$G(u) = e^{-\Psi_{\mathbb{M}} u^{-\epsilon}}, \qquad \Psi_{\mathbb{M}} = \sum_{k\in\mathbb{M}} \sum_{\ell\in\mathbb{M}} \Psi_{k\ell}. \tag{C.4}$$

Given this Fréchet distribution for utility, expected utility is:

$$\mathbb{E}\left[u\right] = \int_0^\infty \epsilon \Psi_{\mathbb{M}} u^{-\epsilon} e^{-\Psi_{\mathbb{M}} u^{-\epsilon}} du. \tag{C.5}$$

Now define the following change of variables:

$$y = \Psi_{\mathbb{M}} u^{-\epsilon}, \qquad dy = -\epsilon \Psi_{\mathbb{M}} u^{-(\epsilon+1)} du. \tag{C.6}$$

Using this change of variables, expected utility can be written as:

$$\mathbb{E}\left[u\right] = \int_0^\infty \Psi_{\mathbb{M}}^{1/\epsilon} y^{-1/\epsilon} e^{-y} dy, \tag{C.7}$$

which can be in turn written as:

$$\mathbb{E}\left[u\right] = \vartheta \Psi_{\mathbb{M}}^{1/\epsilon}, \qquad \vartheta = \Gamma\left(\frac{\epsilon - 1}{\epsilon}\right), \tag{C.8}$$

where $\Gamma(\cdot)$ is the Gamma function. Therefore we obtain the expression in equation (8) in the paper:

$$\mathbb{E}\left[u\right] = \vartheta \Psi_{\mathbb{M}}^{1/\epsilon} = \vartheta \left[\sum_{k\in\mathbb{M}} \sum_{\ell\in\mathbb{M}} \left(B_{k\ell} w_\ell\right)^\epsilon \left(\kappa_{k\ell} P_k^\alpha Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{1/\epsilon}. \tag{C.9}$$

C2 Residence and Workplace Choices

Using the distribution of utility for pairs of residence and employment locations, the probability that a worker chooses the bilateral commute from n to i out of all possible bilateral commutes is:

$$\begin{aligned}
\lambda_{ni} &= \Pr\left[u_{ni} \geq \max\{u_{k\ell}\}; \forall k, \ell\right], \qquad \text{(C.10)} \\
&= \int_0^\infty \prod_{\ell \neq i} G_{n\ell}(u) \left[\prod_{k\neq n} \prod_{\ell\in\mathbb{M}} G_{k\ell}(u)\right] g_{ni}(u) du, \\
&= \int_0^\infty \prod_{k\in\mathbb{M}} \prod_{\ell\in\mathbb{M}} \epsilon \Psi_{ni} u^{-(\epsilon+1)} e^{-\Psi_{k\ell} u^{-\epsilon}} du, \\
&= \int_0^\infty \epsilon \Psi_{ni} u^{-(\epsilon+1)} e^{-\Psi_{\mathbb{M}} u^{-\epsilon}} du.
\end{aligned}$$

Note that:

$$\frac{d}{du}\left[\frac{1}{\Psi_{\mathbb{M}}} e^{-\Psi_{\mathbb{M}} u^{-\epsilon}}\right] = \epsilon u^{-(\epsilon+1)} e^{-\Psi_{\mathbb{M}} u^{-\epsilon}}. \tag{C.11}$$

Using this result to evaluate the integral above, the probability that the worker chooses to live in location n and work in location i is:

$$\frac{L_{ni}}{L_{\mathbb{M}}} = \frac{\Psi_{ni}}{\Psi_{\mathbb{M}}} = \frac{\left(B_{ni} w_i\right)^\epsilon \left(\kappa_{ni} P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{M}} \sum_{\ell\in\mathbb{M}} \left(B_{k\ell} w_\ell\right)^\epsilon \left(\kappa_{k\ell} P_k^\alpha Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{C.12}$$

where L_{ni} is the measure of commuters from residence n to workplace i; $L_{\mathbb{M}}$ is the measure of workers in the economy as a whole; and we assume prohibitive commuting costs across the boundaries of Greater London, such that each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy.

Using this assumption, and summing across workplaces and residences in Greater London, the probability that a worker chooses a residence-workplace pair in Greater London is given by:

$$\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}}=\frac{\sum_{n\in\mathbb{N}}\sum_{i\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{M}}\sum_{\ell\in\mathbb{M}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}. \tag{C.13}$$

Dividing equation (C.12) by equation (C.13), the probability that a worker chooses to live in location n and work in location i (λ_{ni}) conditional on choosing a residence-workplace pair in Greater London is:

$$\lambda_{ni}=\frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}}=\frac{L_{ni}}{L_{\mathbb{N}}}=\frac{\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{C.14}$$

which corresponds to equation (6) in the paper.

Summing across workplaces i in equation (C.14), we obtain the probability that a worker in Greater London chooses to live in residence n, conditional on choosing a workplace-residence pair in Greater London ($R_n/L_{\mathbb{N}}$):

$$\lambda_n^R=\frac{R_n}{L_{\mathbb{N}}}=\frac{\sum_{i\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{C.15}$$

which corresponds to equation (7) in the paper.

Similarly, summing across residences n in equation (C.14), we obtain the probability that a worker in Greater London chooses workplace i, conditional on choosing a workplace-residence pair in Greater London ($L_i/L_{\mathbb{N}}$):

$$\lambda_i^L=\frac{L_i}{L_{\mathbb{N}}}=\frac{\sum_{n\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{C.16}$$

which corresponds to equation (7) in the paper.

For the measure of workers in location i (L_i), we can evaluate the conditional probability that they commute from location n (conditional on having chosen to work in location i):

$$\begin{aligned}\lambda_{ni|i}^L=\frac{\lambda_{ni}}{\lambda_i^L}&=\Pr\left[u_{ni}\geq\max\{u_{ri}\};\forall r\right], \\ &=\int_0^{\infty}\prod_{r\neq n}G_{ri}(u)g_{ni}(u)du, \\ &=\int_0^{\infty}e^{-\Psi_i^L u^{-\epsilon}}\epsilon\Psi_{ni}u^{-(\epsilon+1)}du.\end{aligned} \tag{C.17}$$

where

$$\Psi_i^L\equiv\sum_{k\in\mathbb{M}}\left(B_{ki}w_i\right)^{\epsilon}\left(\kappa_{ki}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}. \tag{C.18}$$

Using the result (C.11) to evaluate the integral in equation (C.17), the probability that a worker commutes from residence n to workplace i conditional on having chosen to work in location i is:

$$\lambda_{ni|i}^L=\frac{\lambda_{ni}}{\lambda_i^L}=\frac{\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{M}}\left(B_{ki}w_i\right)^{\epsilon}\left(\kappa_{ki}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}. \tag{C.19}$$

Under our assumption of prohibitive commuting costs between Greater London and the wider economy, we have:

$$\sum_{k\in\mathbb{M}}\left(B_{ki}w_i\right)^{\epsilon}\left(\kappa_{ki}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}=\sum_{k\in\mathbb{N}}\left(B_{ki}w_i\right)^{\epsilon}\left(\kappa_{ki}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}, \qquad \forall\, i\in\mathbb{N} \tag{C.20}$$

which implies that the commuting probability conditional on workplace in equation (C.19) is equal to:

$$\lambda^L_{ni|i} = \frac{\lambda_{ni}}{\lambda^L_i} = \frac{(B_{ni}w_i)^\epsilon \left(\kappa_{ni}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} (B_{ki}w_i)^\epsilon \left(\kappa_{ki}P_k^\alpha Q_k^{1-\alpha}\right)^{-\epsilon}}. \tag{C.21}$$

which further simplifies to:

$$\lambda^L_{ni|i} = \frac{B_{ni}^\epsilon \left(\kappa_{ni}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} B_{ki}^\epsilon \left(\kappa_{ki}P_k^\alpha Q_k^{1-\alpha}\right)^{-\epsilon}}. \tag{C.22}$$

For the measure of residents of location n (R_n), we can evaluate the conditional probability that they commute to location i (conditional on having chosen to live in location n):

$$\begin{aligned}
\lambda^R_{ni|n} = \frac{\lambda_{ni}}{\lambda^R_n} &= \Pr\left[u_{ni} \geq \max\{u_{n\ell}\}; \forall \ell\right], \qquad \text{(C.23)}\\
&= \int_0^\infty \prod_{\ell\neq i} G_{n\ell}(u) g_{ni}(u) du,\\
&= \int_0^\infty e^{-\Psi_n^R u^{-\epsilon}} \epsilon \Psi_{ni} u^{-(\epsilon+1)} du,
\end{aligned}$$

where

$$\Psi_n^R \equiv \sum_{\ell\in\mathbb{M}} (B_{n\ell}w_\ell)^\epsilon \left(\kappa_{n\ell}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}. \tag{C.24}$$

Using the result (C.11) to evaluate the integral in equation (C.23), the probability that a worker commutes to location i conditional on having chosen to live in location n is:

$$\lambda^R_{ni|n} = \frac{\lambda_{ni}}{\lambda^R_n} = \frac{(B_{ni}w_i)^\epsilon \left(\kappa_{ni}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{\ell\in\mathbb{M}} (B_{n\ell}w_\ell)^\epsilon \left(\kappa_{n\ell}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}, \tag{C.25}$$

Under our assumption of prohibitive commuting costs between Greater London and the wider economy, we have:

$$\sum_{\ell\in\mathbb{M}} (B_{n\ell}w_\ell)^\epsilon \left(\kappa_{n\ell}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon} = \sum_{\ell\in\mathbb{N}} (B_{n\ell}w_\ell)^\epsilon \left(\kappa_{n\ell}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}, \qquad \forall\, n\in\mathbb{N} \tag{C.26}$$

which implies that the commuting probability conditional on residence in equation (C.25) is equal to:

$$\lambda^R_{ni|n} = \frac{\lambda_{ni}}{\lambda^R_n} = \frac{(B_{ni}w_i)^\epsilon \left(\kappa_{ni}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{\ell\in\mathbb{N}} (B_{n\ell}w_\ell)^\epsilon \left(\kappa_{n\ell}P_n^\alpha Q_n^{1-\alpha}\right)^{-\epsilon}}, \tag{C.27}$$

which further simplifies to:

$$\lambda^R_{ni|n} = \frac{(B_{ni}w_i/\kappa_{ni})^\epsilon}{\sum_{\ell\in\mathbb{N}} (B_{n\ell}w_\ell/\kappa_{n\ell})^\epsilon}, \tag{C.28}$$

which corresponds to equation (14) in the paper.

Commuter market clearing requires that the measure of workers employed in each location i (L_i) equals the sum across all locations n of their measures of residents (R_n) times their conditional probabilities of commuting to i ($\lambda^R_{ni|n}$):

$$\begin{aligned}
L_i &= \sum_{n\in\mathbb{N}} \lambda^R_{ni|n} R_n \qquad \text{(C.29)}\\
&= \sum_{n\in\mathbb{N}} \frac{(B_{ni}w_i/\kappa_{ni})^\epsilon}{\sum_{\ell\in\mathbb{N}} (B_{n\ell}w_\ell/\kappa_{n\ell})^\epsilon} R_n,
\end{aligned}$$

where, since there is a continuous measure of workers residing in each location, there is no uncertainty in the supply of workers to each employment location.

Expected worker income conditional on living in location n equals the wages in all possible workplace locations weighted by the probabilities of commuting to those locations conditional on living in n:

$$\begin{aligned}\bar{v}_n &= \mathbb{E}\left[w|n\right] \\ &= \sum_{i\in\mathbb{N}} \lambda^R_{ni|n} w_i, \\ &= \sum_{i\in\mathbb{N}} \frac{\left(B_{ni}w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{\ell\in\mathbb{N}}\left(B_{n\ell}w_\ell/\kappa_{n\ell}\right)^{\epsilon}} w_i,\end{aligned} \tag{C.30}$$

where $\mathbb{E}$ denotes the expectations operator and the expectation is taken over the distribution for idiosyncratic amenities. Intuitively, expected worker income is high in locations that have low commuting costs (low κ_{ni}) to high-wage employment locations.

Another implication of the Fréchet distribution of utility is that the distribution of utility conditional on residing in location n and commuting to location i is the same across all bilateral pairs of locations with positive residents and employment, and is equal to the distribution of utility for the economy as a whole. To establish this result, note that the distribution of utility conditional on residing in location n and commuting to location i is:

$$\begin{aligned} &= \frac{1}{\lambda_{ni}} \int_0^u \prod_{s\neq i} G_{ns}(u) \left[\prod_{k\neq n}\prod_{\ell\in\mathbb{M}} G_{k\ell}(u)\right] g_{ni}(u)du, \\ &= \frac{1}{\lambda_{ni}} \int_0^u \left[\prod_{k\in\mathbb{M}}\prod_{\ell\in\mathbb{M}} e^{-\Psi_{k\ell}u^{-\epsilon}}\right] \epsilon\Psi_{ni}u^{-(\epsilon+1)}du, \\ &= \frac{\Psi_{\mathbb{M}}}{\Psi_{ni}} \int_0^u e^{-\Psi_{\mathbb{M}}u^{-\epsilon}} \epsilon\Psi_{ni}u^{-(\epsilon+1)}du, \\ &= e^{-\Psi_{\mathbb{M}}u^{\epsilon}}.\end{aligned} \tag{C.31}$$

On the one hand, lower land prices in location n or a higher wage in location i raise the utility of a worker with a given realization of idiosyncratic amenities b, and hence increase the expected utility of residing in n and working in i. On the other hand, lower land prices or a higher wage induce workers with lower realizations of idiosyncratic amenities b to reside in n and work in i, which reduces the expected utility of residing in n and working in i. With a Fréchet distribution of utility, these two effects exactly offset one another. Pairs of residence and employment locations with more attractive characteristics attract more commuters on the extensive margin until expected utility is the same across all pairs of residence and employment locations within the economy.

An implication of this result is that expected utility conditional on choosing a residence n and workplace i is the same across all residence-workplace pairs and equal to expected utility in the economy as a whole in equation (C.9):

$$\bar{U} = \vartheta\Psi_{\mathbb{M}}^{1/\epsilon} = \vartheta\left[\sum_{k\in\mathbb{M}}\sum_{\ell\in\mathbb{M}} \left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{1/\epsilon}. \tag{C.32}$$

Using the probability that a worker chooses a residence-workplace pair in Greater London in equation (C.13), this expression for expected utility conditional on choosing each residence-workplace pair ($\bar{U}$) in equation (C.32) can be re-written in the following form:

$$\bar{U}\left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}}\right)^{\frac{1}{\epsilon}} = \vartheta\left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}} \left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{C.33}$$

which corresponds to equation (9) in the paper, where only the limits of the summations differ on the right-hand sides of equations (C.32) and (C.33) .

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

D Isomorphisms

In this section of the online appendix, we show that our quantitative predictions for the impact of the removal of the railway network on workplace employment and commuting patterns hold in an entire class of urban models that satisfy the following three properties: (i) a gravity equation for bilateral commuting flows; (ii) land market clearing, such that income from the ownership of floor space equals the sum of payments for residential and commercial floor space use; (iii) Cobb-Douglas preferences and production technologies. When these three properties are satisfied, workplace income (the total income of all workers) is proportional to revenue and a sufficient statistic for payments for commercial floor space; residence income (the total income of all residents) is a sufficient statistic for payments for residential floor space; and commuting costs regulate the difference between workplace income and residence income.

An implication of this result is our baseline quantitative analysis does *not* require us to make assumptions about (1) whether productivity and amenities are exogenous or endogenous; (2) the underlying determinants of productivity and amenities including agglomeration forces; (3) whether the supply of floor space is exogenous or endogenous; (4) the underlying determinants of the supply of floor space; (5) the extent to which railways affect the cost of trading consumption goods; (6) expected utility in the wider economy. Regardless of the assumptions made about these other components of the model, we obtain the same predictions for the impact of the removal of the railway network on workplace employment and commuting patterns. The reason is that these predictions depend solely on the properties (i)-(iii) introduced immediately above, which hold regardless of these assumptions.

In Section D1, we derive these predictions from the canonical urban model with a single final good that is produced under conditions of perfect competition and constant returns to scale and costlessly traded between locations (as in Lucas and Rossi-Hansberg 2002 and Ahlfeldt et al. 2015). In Section D2, we show that these predictions continue to hold in an extension of this canonical urban model that incorporates non-traded goods. In Section D3, we derive these same predictions in a new economic geography model with monopolistic competition, increasing returns to scale and trade costs, as in Helpman (1998), Redding and Sturm (2008) and Monte, Redding and Rossi-Hansberg (2018). In Section D4, we show that this new economic geography model is isomorphic to a Ricardian spatial model with perfect competition, constant returns to scale and trade costs, as in Eaton and Kortum (2002) and Redding (2016). In Section D5, we demonstrate an analogous isomorphism to an Armington spatial model with neoclassical production and trade costs, as in Armington (1969), Allen and Arkolakis (2014) and Allen, Arkolakis and Li (2017). All of these different model structures satisfy a gravity equation for bilateral commuting flows; land market clearing; the requirement that payments for residential floor space are proportional to residence income; and the requirement that payments for commercial floor space are proportional to workplace income.

D1 Canonical Urban Model

We start by deriving our predictions for the impact of the removal of the railway network on workplace employment in a version of the canonical urban model following Lucas and Rossi-Hansberg (2002) and Ahlfeldt, Redding, Sturm and Wolf (2015). We consider a city (Greater London) that is embedded within a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations $\mathbb{M}$. Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and is indexed by t. The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws. With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and workplace equal the realized values.[1] Motivated by our empirical finding that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$. We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

D1.1 Preferences

Worker preferences are defined over consumption of a homogeneous final good and residential floor space. The indirect utility function is assumed to take the Cobb-Douglas form such that utility for a worker ω residing in n and working in i is given by:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega) w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}, \qquad 0 < \alpha < 1, \tag{D.1}$$

where we suppress the time subscript from now onwards, except where important; P_n is the price of the final good; Q_n is the price of residential floor space; w_i is the wage; κ_{ni} is an iceberg commuting cost; B_{ni} captures amenities from the bilateral commute from residence n to workplace i that are common across all workers; and $b_{ni}(\omega)$ is an idiosyncratic amenity draw that captures all the idiosyncratic factors that can cause an individual to live and work in particular locations within the city.

We assume that idiosyncratic amenities ($b_{ni}(\omega)$) are drawn from an independent extreme value (Fréchet) distribution for each residence-workplace pair and each worker:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{D.2}$$

where we normalize the Fréchet scale parameter in equation (D.2) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities B_{ni} from equation (D.1); the Fréchet shape parameter ϵ regulates the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller the shape parameter ϵ, the greater the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables.

We allow common amenities (B_{ni}) to vary bilaterally to capture the fact that the attractiveness of a given commute may depend on characteristics of both the workplace and the residence. In particular, we decompose this bilateral

[1] To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

parameter B_{ni} into a residence component that is common across all workplaces ($\mathcal{B}_n^R$), a workplace component that is common across all residences ($\mathcal{B}_i^L$), and an idiosyncratic component ($\mathcal{B}_{ni}^I$) that is specific to an individual residence-workplace pair:

$$B_{ni} = \mathcal{B}_n^R \mathcal{B}_i^L \mathcal{B}_{ni}^I, \qquad \mathcal{B}_n^R, \mathcal{B}_i^L, \mathcal{B}_{ni}^I > 0. \tag{D.3}$$

We allow the levels of $\mathcal{B}_n^R$, $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ to differ across residences n and workplaces i, although when we examine the impact of the construction of the railway network, we assume that $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ are time invariant. In contrast, we allow B_n^R to change over time, and for these changes to be potentially endogenous to the evolution of the surrounding concentration of economic activity through agglomeration forces.

D1.2 Production

The final good is produced under conditions of perfect competition and constant returns to scale using labor, machinery capital and commercial floor space, where commercial floor space includes both building capital and land. The production technology is assumed to take the Cobb-Douglas form with the following unit cost:

$$P_i = \frac{1}{A_i} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}, \qquad 0 < \beta^L, \beta^H, \beta^M < 1, \qquad \beta^L + \beta^H + \beta^M = 1, \tag{D.4}$$

where A_i denotes productivity in location i; q_i is the price of commercial floor space; and machinery is assumed to be perfectly mobile across locations with a common price r determined in the wider economy. Although an advantage of our empirical setting of 19th-century London is the absence of large-scale urban planning, we allow the price of commercial floor space (q_i) to potentially depart from the price of residential floor space (Q_i) in each location i through a wedge ξ_i, such that $q_i = \xi_i Q_i$. We also allow productivity (A_i) to potentially respond endogenously to changes in the surrounding concentration of economic activity through agglomeration forces.

Within Greater London, we assume that the homogeneous final good is costlessly traded such that:

$$P_i = P, \qquad \forall i \in \mathbb{N}. \tag{D.5}$$

Between Greater London and the rest of Great Britain, we allow for changes in trade costs for this good, which are reflected in changes in its price at the boundaries of Greater London (P).

From profit maximization and zero profits, we obtain the results in equations (10) and (12) in the paper that payments to labor, commercial floor space and machinery capital are constant shares of revenue:

$$w_i L_i = \beta^L X_i, \qquad q_i H_i^L = \beta^H X_i, \qquad r M_i = \beta^M X_i, \tag{D.6}$$

where L_i is workplace employment; X_i is revenue; and H_i^L denotes commercial floor space use. Therefore, payments for commercial floor space are proportional to workplace income:

$$q_i H_i^L = \frac{\beta^H}{\beta^L} w_i L_i. \tag{D.7}$$

Re-arranging equation (D.4), we obtain another key implication of profit maximization and zero profits for each location with positive production:

$$w_i = (P A_i)^{1/\beta^L} q_i^{-\beta^H/\beta^L} r^{-\beta^M/\beta^L}. \tag{D.8}$$

Intuitively, the maximum wage (w_i) that a location can afford to pay workers is increasing in the location's productivity (A_i) and the common price of the final good (P) and decreasing in the price of commercial floor space (q_i) and the common price of machinery capital (r).

D1.3 Land Market Clearing

Land market clearing implies that the total income received by landlords as owners of floor space (which equals rateable value ($\mathbb{Q}_n$) in our data) equals the sum of payments for the use of residential and commercial floor space:

$$\mathbb{Q}_n = Q_n H_n^R + q_n H_n^L = (1-\alpha)\, v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n. \tag{D.9}$$

where H_n^R is the quantity of residential floor space; v_n is the per capita income of location n's residents, as determined below as a function of commuting patterns; and R_n is the measure of these residents.

Importantly, we allow the supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) to be endogenous, and we allow the prices of residential and commercial floor space to potentially differ from one another through the wedge ξ_i ($q_i = \xi_i Q_i$). In our baseline quantitative analysis below, we are not required to make assumptions about these supplies of residential and commercial floor space or this wedge between commercial and residential floor prices. The reason is that we condition on the observed rateable values in the data ($\mathbb{Q}_n$) and the supplies and prices for residential and commercial floor space (H_n^R, H_n^L, Q_n, q_n) only enter the land market clearing condition (D.9) through these observed rateable values.

D1.4 Workplace and Residence Choices

Using indirect utility (D.1) and the Fréchet distribution of idiosyncratic amenities (D.2), this canonical urban model exhibits a gravity equation for commuting flows. Following the same analysis as in Section C of this online appendix, the probability that a worker chooses to reside in location $n \in \mathbb{N}$ and work in location $i \in \mathbb{N}$ conditional on choosing a residence-workplace pair in Greater London (λ_{ni}) is given by:

$$\lambda_{ni} = \frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{(B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} (B_{k\ell} w_{\ell})^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}}, \qquad n, i \in \mathbb{N}, \tag{D.10}$$

which corresponds to equation (6) in the paper.

Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker in lives in each residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^R = R_n/L_{\mathbb{N}}$). Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker in Greater London is employed in workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_i^L = L_i/L_{\mathbb{N}}$):

$$\lambda_n^R = \frac{\sum_{i\in\mathbb{N}} (B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} (B_{k\ell} w_{\ell})^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}}, \quad \lambda_i^L = \frac{\sum_{n\in\mathbb{N}} (B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} (B_{k\ell} w_{\ell})^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{D.11}$$

which corresponds to equation (7) in the paper.

From equations (D.10) and (D.11), the conditional probability that a worker commutes to location i conditional on residing in location n ($\lambda_{ni|n}^R$) also takes the same form as in equation (14) in the paper:

$$\lambda_{ni|n}^R = \frac{\lambda_{ni}}{\lambda_n^R} = \frac{(B_{ni} w_i/\kappa_{ni})^{\epsilon}}{\sum_{\ell\in\mathbb{N}} (B_{n\ell} w_{\ell}/\kappa_{n\ell})^{\epsilon}}. \tag{D.12}$$

Using this commuting probability conditional on residence ($\lambda_{ni|n}^R$) from equation (D.12), we obtain an identical expression for per capita income by residence as in equation (15) in the paper:

$$v_n = \sum_{i\in\mathbb{N}} \lambda_{ni|n}^R w_i. \tag{D.13}$$

Commuter market clearing implies that employment in each location (L_i) equals the measure of workers choosing to commute to that location. Using the commuting probabilities conditional on residence from equation (D.12), we obtain the same expression for this commuter market clearing condition as in equation (13) in the paper:

$$L_i = \sum_{n \in \mathbb{N}} \lambda^R_{ni|n} R_n. \tag{D.14}$$

Finally, using the Fréchet distribution of idiosyncratic amenities (D.2), expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is equalized across all residence-workplace pairs in the economy and takes the same form as in equation (8) in the paper:

$$\bar{U} = \vartheta \left[\sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} \left(B_{k\ell} w_\ell\right)^\epsilon \left(\kappa_{k\ell} P_k^\alpha Q_k^{1-\alpha}\right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \tag{D.15}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma((\epsilon - 1)/\epsilon)$; and $\Gamma(\cdot)$ is the Gamma function. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}}/L_{\mathbb{M}}$), we can rewrite this population mobility condition as:

$$\bar{U} \left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}} \right)^{\frac{1}{\epsilon}} = \vartheta \left[\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell} w_\ell\right)^\epsilon \left(\kappa_{k\ell} P_k^\alpha Q_k^{1-\alpha}\right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \tag{D.16}$$

where only the limits of the summations differ on the right-hand sides of equations (D.15) and (D.16).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

D1.5 Comparative Statics for Changes in Commuting Costs

We now show that this canonical urban model yields the same predictions for the impact of the removal of the railway network on workplace employment and commuting patterns as in the paper, once we condition on the observed variables in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values.

First, using equation (D.9), the land market clearing condition for any earlier year $\tau < t$ can be written in terms of the observed variables and model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt} \mathbb{Q}_{nt} = (1-\alpha) \hat{v}_{nt} v_{nt} \hat{R}_{nt} R_{nt} + \frac{\beta^H}{\beta^L} \hat{w}_{nt} w_{nt} \hat{L}_{nt} L_{nt}, \tag{D.17}$$

where recall that a hat above a variable denotes a relative change, such that $\hat{x}_t = x_\tau / x_t$.

Second, using equations (D.12) and (D.13), per capita income by residence (v_{nt}) for any earlier year $\tau < t$ can be written in the same form as equation (19) in the paper:

$$\hat{v}_{nt} v_{nt} = \sum_{i \in \mathbb{N}} \frac{\lambda^R_{nit|n} \hat{w}^\epsilon_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^\epsilon_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{w}_{it} w_{it}, \tag{D.18}$$

where ($\lambda^R_{nit|n}$, w_{it}, v_{nt}) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has cancelled from numerator and denominator of the fraction on the right-hand side of equation (D.18); and we assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Third, using equations (D.12) and (D.14), workplace employment (L_{it}) for any earlier year $\tau < t$ can be written in the same form as equation (18) in the paper:

$$\hat{L}_{it} L_{it} = \sum_{n \in \mathbb{N}} \frac{\lambda^R_{nit|n} \hat{w}^{\epsilon}_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^{\epsilon}_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{R}_{nt} R_{nt}, \tag{D.19}$$

where ($\lambda^R_{nit|n}$, L_{it}, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.19); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Finally, using equation (D.12), commuting flows ($\hat{L}_{nit}$) for any earlier year $\tau < t$ can be written in the same form as in equation (21) in the paper:

$$\hat{L}_{nit} L_{nit} = \frac{\lambda^R_{nit|n} \hat{w}^{\epsilon}_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^{\epsilon}_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{R}_{nt} R_{nt}, \tag{D.20}$$

where (L_{nit}, $\lambda^R_{nit|n}$, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from numerator and denominator of the fraction on the right-hand side of equation (D.20); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Note that equations (D.17), (D.18), (D.19) and (D.20) above are identical to equations (17), (19), (18) and (21) in the paper. Therefore, given the same observed variables in the initial equilibrium (L_{nt}, R_{nt}, $\mathbb{Q}_{nt}$, w_{nt}, v_{nt}, L_{nit}), the same observed changes in residents and rateable values ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$) and the same estimated changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), this canonical urban model predicts the same changes in workplace employment ($\hat{L}_{it}$) and commuting patterns ($\hat{L}_{nit}$) as in the paper.

D2 Non-traded Goods Extension of Canonical Urban Model

We now show that the same predictions for the impact of the removal of the railway network go through in an extension of the canonical urban model from the previous section to incorporate non-traded goods. We consider a city (Greater London) embedded within a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations ($\mathbb{M}$). Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and is indexed by t. The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws. With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and workplace equal the realized values.[2] Motivated by our empirical finding that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$. We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity in the traded and non-traded

[2] To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

sectors, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

D2.1 Preferences

Worker preferences are defined over consumption of a homogeneous traded good, a homogeneous non-traded good, and residential floor space. The indirect utility function is assumed to take the Cobb-Douglas form such that utility for a worker ω residing in n and working in i is given by:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega) w_i}{\kappa_{ni} \left(P_n^T\right)^{\alpha^T} \left(P_n^N\right)^{\alpha^N} Q_n^{1-\alpha^T-\alpha^N}}, \qquad 0 < \alpha^T, \alpha^N < 1, \qquad 0 < \alpha^T + \alpha^N < 1, \tag{D.21}$$

where we suppress the time subscript from now onwards, except where important; P_n^T is the price of the traded good; P_n^N is the price of the non-traded good; Q_n is the price of residential floor space; w_i is the wage; κ_{ni} is an iceberg commuting cost; B_{ni} captures amenities from the bilateral commute from residence n to workplace i that are common across all workers; and $b_{ni}(\omega)$ is an idiosyncratic amenity draw that captures all the idiosyncratic factors that can cause an individual to live and work in particular locations within the city.

We assume that idiosyncratic amenities ($b_{ni}(\omega)$) are drawn from an independent extreme value (Fréchet) distribution for each residence-workplace pair and each worker:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{D.22}$$

where we normalize the Fréchet scale parameter in equation (D.22) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities B_{ni} from equation (D.21); the Fréchet shape parameter ϵ regulates the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller the shape parameter ϵ, the greater the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables.

We allow common amenities (B_{ni}) to vary bilaterally to capture the fact that the attractiveness of a given commute may depend on characteristics of both the workplace and the residence. In particular, we decompose this bilateral parameter B_{ni} into a residence component that is common across all workplaces ($\mathcal{B}_n^R$), a workplace component that is common across all residences ($\mathcal{B}_i^L$), and an idiosyncratic component ($\mathcal{B}_{ni}^I$) that is specific to an individual residence-workplace pair:

$$B_{ni} = \mathcal{B}_n^R \mathcal{B}_i^L \mathcal{B}_{ni}^I, \qquad \mathcal{B}_n^R, \mathcal{B}_i^L, \mathcal{B}_{ni}^I > 0. \tag{D.23}$$

We allow the levels of $\mathcal{B}_n^R$, $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ to differ across residences n and workplaces i, although when we examine the impact of the construction of the railway network, we assume that $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ are time invariant. In contrast, we allow B_n^R to change over time, and for these changes to be potentially endogenous to the evolution of the surrounding concentration of economic activity through agglomeration forces,

D2.2 Production

The traded and non-traded goods are produced under conditions of perfect competition and constant returns to scale using labor, machinery capital and commercial floor space, where commercial floor space includes building capital and land. The production technology is assumed to take the Cobb-Douglas form with the following unit costs:

$$P_i^T = \frac{1}{A_i^T} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}, \qquad 0 < \beta^L, \beta^H, \beta^M < 1 \qquad \beta^L + \beta^H + \beta^M = 1, \tag{D.24}$$
$$P_i^N = \frac{1}{A_i^N} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M},$$

where A_i^T and A_i^N are the productivities of traded and non-traded production in location i; q_i is the price of commercial floor space; machinery is assumed to be perfectly mobile across locations with a common price r determined in the wider economy; and, for simplicity, we assume the same factor intensity in both sectors. Although an advantage of our empirical setting of 19th-century London is the absence of large-scale urban planning, we allow the price of commercial floor space (q_i) to potentially depart from the price of residential floor space (Q_i) in each location i through a wedge ξ_i, such that $q_i = \xi_i Q_i$. We also allow productivity in each sector (A_i^T, A_i^N) to potentially respond endogenously to changes in the surrounding concentration of economic activity through agglomeration forces.

Within Greater London, we assume that the traded good is costlessly traded such that:

$$P_i^T = P^T, \qquad \forall i \in \mathbb{N}. \tag{D.25}$$

Between Greater London and the rest of Great Britain, we allow for changes in trade costs for this good, which are reflected in changes in its price at the boundaries of Greater London (P^T).

From profit maximization and zero profits, we obtain the results in equations (10) and (12) in the paper that payments to labor, commercial floor space and machinery capital are constant shares of revenue in each sector:

$$w_i L_i^T = \beta^L X_i^T, \qquad q_i H_i^T = \beta^H X_i^T, \qquad r M_i^T = \beta^M X_i^T \tag{D.26}$$
$$w_i L_i^N = \beta^L X_i^N, \qquad q_i H_i^N = \beta^H X_i^N, \qquad r M_i^N = \beta^M X_i^N$$

where L_n^T and L_n^N denote employment in the traded and non-traded sectors respectively; X_i^T and X_i^N correspond to revenue in the two sectors; H_i^T and H_i^N represent commercial floor space use in the two sectors; and M_i^T and M_i^N are machinery inputs in the two sectors. Therefore, total payments for commercial floor space across the two sectors together are proportional to total workplace income:

$$q_i H_i^L = q_i \left[H_i^T + H_i^N\right] = \frac{\beta^H}{\beta^L} w_i \left[L_i^T + L_i^N\right] = \frac{\beta^H}{\beta^L} w_i L_i. \tag{D.27}$$

Re-arranging equation (D.24), we obtain another key implication of profit maximization and zero profits in each sector for each location with positive production:

$$w_i = \left(P^T A_i^T\right)^{1/\beta^L} q_i^{-\beta^H/\beta^L} r^{-\beta^M/\beta^L}, \tag{D.28}$$
$$w_i = \left(P_i^N A_i^N\right)^{1/\beta^L} q_i^{-\beta^H/\beta^L} r^{-\beta^M/\beta^L}.$$

Intuitively, the maximum wage (w_i) that a location can afford to pay in the traded sector is increasing in its productivity (A_i^T) and the price of the final good (P^T) and decreasing in the price of commercial floor space (q_i) and the common price of machinery capital (r). In equilibrium, each location produces both the traded and non-traded final good, and the wage (w_i) and the price of the non-traded good (P_i^N) adjust, such that zero profits are made in both sectors.

D2.3 Market Clearing

Land market clearing implies that the total income received by landlords as owners of floor space (which equals rateable value ($\mathbb{Q}_n$) in our data) equals the sum of payments for the use of residential and commercial floor space:

$$\mathbb{Q}_n = Q_n H_n^R + q_n H_n^L = (1-\alpha)\, v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n. \tag{D.29}$$

where H_n^R is residential floor space use; rateable values ($\mathbb{Q}_n$) equal the sum of prices times quantities for both residential and commercial floor space use; $\alpha = \alpha^T + \alpha^N$ is the overall share of expenditure on consumption goods; v_n is the per capita income of location n's residents, as determined below as a function of commuting patterns; and R_n is the measure of these residents.

Importantly, we allow the supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) to be endogenous, and we allow the prices of residential and commercial floor space to potentially differ from one another through the wedge ξ_i ($q_i = \xi_i Q_i$). In our baseline quantitative analysis below, we are not required to make assumptions about these supplies of residential and commercial floor space or this wedge between commercial and residential floor prices. The reason is that we condition on the observed rateable values in the data ($\mathbb{Q}_n$) and the supplies and prices for residential and commercial floor space (H_n^R, H_n^L, Q_n, q_n) only enter the land market clearing condition (D.29) through these observed rateable values.

D2.4 Workplace and Residence Choices

Using indirect utility (D.21) and the Fréchet distribution of idiosyncratic amenities (D.22), this extension of the canonical urban model exhibits a gravity equation for commuting flows. Following the same analysis as in Section C of this online appendix, the probability that a worker chooses to reside in location $n \in \mathbb{N}$ and work in location $i \in \mathbb{N}$ conditional on choosing a residence-workplace pair in Greater London (λ_{ni}) is given by:

$$\lambda_{ni} = \frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{(B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} \left(P_n^T\right)^{\alpha^T} \left(P_n^N\right)^{\alpha^N} Q_n^{1-\alpha^T-\alpha^N}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} (B_{k\ell} w_\ell)^{\epsilon} \left(\kappa_{k\ell} \left(P_k^T\right)^{\alpha^T} \left(P_k^N\right)^{\alpha^N} Q_k^{1-\alpha^T-\alpha^N}\right)^{-\epsilon}}, \qquad n, i \in \mathbb{N}, \tag{D.30}$$

which takes the same form as equation (6) in the paper, except that the consumption goods price index is now disaggregated into traded and non-traded goods.

Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker in Greater London lives in residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^R = R_n/L_{\mathbb{N}}$). Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker in Greater London is employed in workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^L = L_i/L_{\mathbb{N}}$):

$$\lambda_n^R = \frac{\sum_{i\in\mathbb{N}} (B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} \left(P_n^T\right)^{\alpha^T} \left(P_n^N\right)^{\alpha^N} Q_n^{1-\alpha^T-\alpha^N}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} (B_{k\ell} w_\ell)^{\epsilon} \left(\kappa_{k\ell} \left(P_k^T\right)^{\alpha^T} \left(P_k^N\right)^{\alpha^N} Q_k^{1-\alpha^T-\alpha^N}\right)^{-\epsilon}}, \tag{D.31}$$

$$\lambda_i^L = \frac{\sum_{n\in\mathbb{N}} (B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} \left(P_n^T\right)^{\alpha^T} \left(P_n^N\right)^{\alpha^N} Q_n^{1-\alpha^T-\alpha^N}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} (B_{k\ell} w_\ell)^{\epsilon} \left(\kappa_{k\ell} \left(P_k^T\right)^{\alpha^T} \left(P_k^N\right)^{\alpha^N} Q_k^{1-\alpha^T-\alpha^N}\right)^{-\epsilon}}. \tag{D.32}$$

Both expressions take the same form as in equation (7) in the paper, with the consumption goods price index disaggregated into traded and non-traded goods.

From equations (D.30) and (D.31), the conditional probability that a worker commutes to location i conditional on residing in location n also takes the same form as in equation (14) in the paper:

$$\lambda^R_{ni|n} = \frac{\lambda_{ni}}{\lambda^R_n} = \frac{\left(B_{ni} w_i / \kappa_{ni}\right)^{\epsilon}}{\sum_{\ell \in \mathbb{N}} \left(B_{n\ell} w_\ell / \kappa_{n\ell}\right)^{\epsilon}}. \tag{D.33}$$

Using this commuting probability conditional on residence ($\lambda^R_{ni|n}$) from equation (D.33), we obtain an identical expression for per capita income by residence as in equation (15) in the paper:

$$v_n = \sum_{i \in \mathbb{N}} \lambda^R_{ni|n} w_i. \tag{D.34}$$

Commuter market clearing implies that employment in each location (L_i) equals the measure of workers choosing to commute to that location. Using the commuting probabilities conditional on residence from equation (D.33), we obtain the same expression for this commuter market clearing condition as in equation (13) in the paper:

$$L_i = \sum_{n \in \mathbb{N}} \lambda^R_{ni|n} R_n. \tag{D.35}$$

Finally, using the Fréchet distribution for idiosyncratic amenities (D.22), expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is equalized across all residence-workplace pairs in the economy and takes the same form as in equation (8) in the paper:

$$\bar{U} = \vartheta \left[\sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} \left(B_{k\ell} w_\ell\right)^{\epsilon} \left(\kappa_{k\ell} \left(P^T_k\right)^{\alpha^T} \left(P^N_k\right)^{\alpha^N} Q_k^{1-\alpha^T-\alpha^N} \right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \tag{D.36}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma((\epsilon-1)/\epsilon)$; and $\Gamma(\cdot)$ is the Gamma function. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}}/L_{\mathbb{M}}$), we can re-write this population mobility condition as:

$$\bar{U} \left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}} \right)^{\frac{1}{\epsilon}} = \vartheta \left[\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell} w_\ell\right)^{\epsilon} \left(\kappa_{k\ell} \left(P^T_k\right)^{\alpha^T} \left(P^N_k\right)^{\alpha^N} Q_k^{1-\alpha^T-\alpha^N} \right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \tag{D.37}$$

where only the limits of the summations differ on the right-hand sides of equations (D.36) and (D.37).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

D2.5 Comparative Statics for Changes in Commuting Costs

We now show that this extension of the canonical urban model yields the same predictions for the impact of the removal of the railway network on workplace employment and commuting patterns as in the paper, once we condition on the observed variables in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values.

First, using equation (D.29), the land market clearing condition for any earlier year $\tau < t$ can be written in terms of the observed variables and model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt} \mathbb{Q}_{nt} = (1-\alpha) \hat{v}_{nt} v_{nt} \hat{R}_{nt} R_{nt} + \frac{\beta^H}{\beta^L} \hat{w}_{nt} w_{nt} \hat{L}_{nt} L_{nt}, \tag{D.38}$$

where recall that a hat above a variable denotes a relative change, such that $\hat{x}_t = x_\tau / x_t$.

Second, using equations (D.33) and (D.34), per capita income by residence (v_{nt}) for any earlier year $\tau < t$ can be written in the same form as equation (19) in the paper:

$$\hat{v}_{nt} v_{nt} = \sum_{i \in \mathbb{N}} \frac{\lambda^R_{nit|n} \hat{w}^\epsilon_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^\epsilon_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{w}_{it} w_{it}, \tag{D.39}$$

where ($\lambda^R_{nit|n}$, w_{it}, v_{nt}) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.39); and we assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Third, using equations (D.33) and (D.35), workplace employment (L_{it}) for any earlier year $\tau < t$ can be written in the same form as equation (18) in the paper:

$$\hat{L}_{it} L_{it} = \sum_{n \in \mathbb{N}} \frac{\lambda^R_{nit|n} \hat{w}^\epsilon_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^\epsilon_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{R}_{nt} R_{nt}, \tag{D.40}$$

where ($\lambda^R_{nit|n}$, L_{it}, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.40); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Finally, using equation (D.33), commuting flows ($\hat{L}_{nit}$) for any earlier year $\tau < t$ can be written in the same form as in equation (21) in the paper:

$$\hat{L}_{nit} L_{nit} = \frac{\lambda^R_{nit|n} \hat{w}^\epsilon_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^\epsilon_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{R}_{nt} R_{nt}, \tag{D.41}$$

where (L_{nit}, $\lambda^R_{nit|n}$, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.41); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Note that equations (D.38), (D.39), (D.40) and (D.41) above are identical to equations (17), (19), (18) and (21) in the paper. Therefore, given the same observed variables in the initial equilibrium (L_{nt}, R_{nt}, $\mathbb{Q}_{nt}$, w_{nt}, v_{nt}, L_{nit}), the same observed changes in residents and rateable values ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$) and the same estimated changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), this extension of the canonical urban model predicts the same changes in workplace employment ($\hat{L}_{it}$) and commuting patterns ($\hat{L}_{nit}$) as in the paper.

D3 New Economic Geography Model

We now derive our predictions for the impact of the removal of the railway network in a new economic geography model following Helpman (1998), Redding and Sturm (2008) and Monte, Redding and Rossi-Hansberg (2018). We consider a city (Greater London) that is embedded within a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations $\mathbb{M}$. Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and is indexed by t. The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws. With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and workplace equal the realized values.[3] Motivated by our empirical finding that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$. We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

D3.1 Preferences and Endowments

The preferences of a worker ω who lives in location n and works in location i are defined over final goods consumption ($C_n(\omega)$), residential floor space use ($H_n^R(\omega)$), iceberg commuting costs (κ_{ni}), common amenities for all workers (B_{ni}), and an idiosyncratic amenity draw for an individual worker for each residence-workplace pair ($b_{ni}(\omega)$), according to the following Cobb-Douglas functional form:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega)}{\kappa_{ni}} \left(\frac{C_n(\omega)}{\alpha} \right)^{\alpha} \left(\frac{H_n^R(\omega)}{1-\alpha} \right)^{1-\alpha}, \qquad 0 < \alpha < 1, \tag{D.42}$$

where we suppress the time subscript from now onwards, except where important. The idiosyncratic amenities shock for worker ω for each residence n and workplace i ($b_{ni}(\omega)$) is drawn from an independent Fréchet distribution:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{D.43}$$

where we normalize the Fréchet scale parameter in equation (D.43) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities B_{ni} from equation (D.42); the Fréchet shape parameter $\epsilon > 1$ regulates the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller is the shape parameter ϵ, the greater is the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables. All workers ω residing in location n and working in location i receive the same wage and make the same choices for consumption and residential floor space use. Therefore, we suppress the implicit dependence on ω, except where important.

We allow common amenities (B_{ni}) to vary bilaterally to capture the fact that the attractiveness of a given commute may depend on characteristics of both the workplace and the residence. In particular, we decompose this bilateral parameter B_{ni} into a residence component that is common across all workplaces ($\mathcal{B}_n^R$), a workplace component that is common across all residences ($\mathcal{B}_i^L$), and an idiosyncratic component ($\mathcal{B}_{ni}^I$) that is specific to an individual residence-workplace pair:

$$B_{ni} = \mathcal{B}_n^R \mathcal{B}_i^L \mathcal{B}_{ni}^I, \qquad \mathcal{B}_n^R, \mathcal{B}_i^L, \mathcal{B}_{ni}^I > 0. \tag{D.44}$$

We allow the levels of $\mathcal{B}_n^R$, $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ to differ across residences n and workplaces i, although when we examine the impact of the construction of the railway network, we assume that $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ are time invariant. In contrast, we allow B_n^R to change over time, and for these changes to be potentially endogenous to the evolution of the surrounding concentration of economic activity through agglomeration forces,

[3] To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

The goods consumption index in location n takes the constant elasticity of substitution (CES) or Dixit-Stiglitz form and is defined over a continuum of varieties sourced from each location i,

$$C_n = \left[\sum_{i \in \mathbb{M}} \int_0^{\mathcal{M}_i} c_{ni}(j)^{\rho} dj \right]^{\frac{1}{\rho}}, \qquad \sigma = \frac{1}{1-\rho} > 1, \tag{D.45}$$

where $c_{ni}(j)$ is consumption in location n of an individual variety j produced in location i; $\mathcal{M}_i$ is the mass of varieties produced in location i; and ρ is the CES parameter that determines the elasticity of substitution between varieties ($\sigma = 1/(1-\rho) > 1$).

Using the properties of CES preferences (D.45), the equilibrium consumption in location n of each variety j sourced from location i is determined by:

$$c_{ni}(j) = E_n P_n^{\sigma-1} p_{ni}(j)^{-\sigma}, \tag{D.46}$$

where $E_n = P_n C_n$ is total expenditure on consumption goods in location n; P_n is the price index dual to the consumption index (D.45); and $p_{ni}(j)$ is the "cost inclusive of freight" price of variety j produced in location i and consumed in location n.

Goods can be traded between locations subject to iceberg variable trade costs, such that $d_{ni} \geq 1$ units of a good must be shipped from location i in order for one unit to arrive in location n (where $d_{ni} > 1$ for $n \neq i$ and $d_{nn} = 1$). The "cost inclusive of freight" price of a variety in the location of consumption n ($p_{ni}(j)$) is thus a constant multiple of the "free on board" price of that variety in the location of production i ($p_i(j)$), with that multiple determined by these iceberg trade costs:

$$p_{ni}(j) = d_{ni} p_i(j). \tag{D.47}$$

D3.2 Production

Production is modelled as in the new economic geography literature following Krugman (1991) and Helpman (1998). Varieties are produced under conditions of monopolistic competition using labor, machinery capital, and commercial floor space, where commercial floor space includes both building capital and land. To produce a variety, a firm must incur both a fixed cost and a constant variable cost. We assume that these fixed and variable costs use the three factors of production with the same factor intensity, such that the production technology is homothetic. We allow the variable cost to vary with location productivity A_i, such that the total cost of producing $y_i(j)$ units of a variety j in location i is given by:

$$\Gamma_i(j) = \left(F + \frac{y_i(j)}{A_i} \right) w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}, \qquad 0 < \beta^L, \beta^H, \beta^M < 1, \qquad \beta^L + \beta^H + \beta^M = 1, \tag{D.48}$$

where w_i is the wage; q_i is the price of commercial floor space in location i; and machinery is assumed to be perfectly mobile across locations with a common price r determined in the wider economy. We also allow productivity (A_i) to potentially respond endogenously to changes in the surrounding concentration of economic activity through agglomeration forces. Profit maximization implies that equilibrium variety prices are a constant mark-up over marginal cost:

$$p_{ni}(j) = p_{ni} = \left(\frac{\sigma}{\sigma - 1} \right) \frac{d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}}{A_i}. \tag{D.49}$$

Profit maximization and zero profits imply that the equilibrium output of each variety is the same for all varieties produced in location i:

$$y_i(j) = \bar{y}_i = A_i F (\sigma - 1). \tag{D.50}$$

Using the equilibrium pricing rule (D.49) and zero profits (D.50), free on board revenue ($x_i(j) = p_i(j)y_i(j)$) for each variety j in location i can be written as:

$$p_i(j)y_i(j) = x_i(j) = \bar{x}_i = \sigma w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} F, \tag{D.51}$$

and the common equilibrium wage bill for each variety j in location i is given by:

$$w_i l_i(j) = w_i \bar{l}_i = \beta^L \bar{x}_i, \tag{D.52}$$

where $l_i(j) = \bar{l}_i$ is workplace employment for variety j in location i.

Aggregating across all varieties produced within location i, profit maximization and zero profits imply that payments for labor, commercial floor space and machinery capital are constant shares of revenue, as in equations (10) and (12) in the paper:

$$w_i L_i = \beta^L X_i, \qquad q_i H_i^L = \beta^H X_i \qquad r M_i = \beta^M X_i, \tag{D.53}$$

where L_i is total workplace employment; $X_i = \mathcal{M}_i \bar{x}_i$ is aggregate revenue; H_i^L denotes total commercial floor space use; and M_i is total machinery capital use. Therefore, payments for commercial floor space are proportional to workplace income ($w_i L_i$):

$$q_i H_i^L = \frac{\beta^H}{\beta^L} w_i L_i. \tag{D.54}$$

D3.3 Trade and Market Clearing

We assume that floor space is owned by landlords, who receive income from residents' and firms' expenditure on floor space, and consume only consumption goods where they live. Total expenditure on consumption goods equals the fraction α of the total income of residents plus the entire income of landlords. This income of landlords equals $(1-\alpha)$ times the total income of residents plus β^H times revenue (which equals β^H/β^L times the total income of workers). Therefore, total expenditure on consumption goods is:

$$E_n = P_n C_n = \alpha v_n R_n + (1-\alpha) v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n = v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n,$$

where v_n is the per capita income of location n's residents, as determined below as a function of commuting patterns, and R_n is the measure of these residents.

This new economic geography model implies a gravity equation for bilateral trade in goods between locations. Using CES demand in equation (D.46), and the fact that all varieties supplied from location i to location n charge the same price in equation (D.49), the share of location n's expenditure on goods produced in location i can be written as:

$$\pi_{ni} = \frac{\mathcal{M}_i p_{ni}^{1-\sigma}}{\sum_{k \in \mathbb{M}} \mathcal{M}_k p_{nk}^{1-\sigma}} = \frac{\mathcal{M}_i \left(d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} / A_i \right)^{1-\sigma}}{\sum_{k \in \mathbb{M}} \mathcal{M}_k \left(d_{nk} w_k^{\beta^L} q_k^{\beta^H} r^{\beta^M} / A_k \right)^{1-\sigma}}. \tag{D.55}$$

Therefore trade between locations n and i depends on bilateral trade costs (d_{ni}) in the numerator ("bilateral resistance") and on trade costs to all possible sources of supply k in the denominator ("multilateral resistance"). Goods market clearing and zero profits imply that payments to workers plus payments for commercial floor space use in each location equal expenditure on goods produced in that location:

$$w_i L_i + q_i H_i^L + r M_i = \left[1 + \frac{\beta^H}{\beta^L} + \frac{\beta^M}{\beta^L} \right] w_i L_i = \sum_{n \in \mathbb{N}} \pi_{ni} E_n. \tag{D.56}$$

Using equilibrium prices (D.49), the price index dual to the consumption index (D.45) can be rewritten as:

$$P_n = \left[\sum_{i \in \mathbb{M}} \mathcal{M}_i \left(\frac{\sigma}{\sigma - 1} \frac{d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}}{A_i} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} = \left(\frac{\mathcal{M}_n}{\pi_{nn}} \right)^{\frac{1}{1-\sigma}} \frac{\sigma}{\sigma - 1} \frac{w_n^{\beta^L} q_n^{\beta^H} r^{\beta^M}}{A_n}, \tag{D.57}$$

where the second equation uses the domestic trade share (π_{nn}) from equation (D.55) and $d_{nn} = 1$.

Labor market clearing implies that total payments to labor in each location equal the mass of varieties times labor payments for each variety. Using this relationship and the Cobb-Douglas production technology, the mass of varieties ($\mathcal{M}_i$) in each location can be written as a function of total labor payments ($w_i L_i$) and firm revenue ($\bar{x}_i$) in that location:

$$\mathcal{M}_i = \frac{w_i L_i}{w_i \bar{l}_i} = \frac{w_i L_i}{\beta^L \bar{x}_i}, \tag{D.58}$$

where L_i is total employment.

Land market clearing implies that the total income received by landlords as owners of floor space (which equals rateable value ($\mathbb{Q}_n$) in our data) equals the sum of payments for the use of residential and commercial floor space:

$$\mathbb{Q}_n = Q_n H_n^R + q_n H_n^L = (1 - \alpha) v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n, \tag{D.59}$$

where H_n^R is residential floor space use.

Importantly, we allow the supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) to be endogenous, and we allow the prices of residential and commercial floor space to potentially differ from one another through the wedge ξ_i ($q_i = \xi_i Q_i$). In our baseline quantitative analysis below, we are not required to make assumptions about these supplies of residential and commercial floor space or this wedge between commercial and residential floor prices. The reason is that we condition on the observed rateable values in the data ($\mathbb{Q}_n$) and the supplies and prices for residential and commercial floor space (H_n^R, H_n^L, Q_n, q_n) only enter the land market clearing condition (D.59) through these observed rateable values.

D3.4 Workplace and Residence Choices

Given the direct utility function (D.42), the corresponding indirect utility function for a worker ω residing in location n and working in location i is:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega) w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}, \tag{D.60}$$

which takes the same form as equation (3) in the paper and equation (D.1) in the canonical urban model in Section D1 of this online appendix. The only difference from the canonical urban model is in the underlying determinants of the price index for goods consumption (P_n), as now specified in equation (D.57).

Using indirect utility (D.60) and the Fréchet distribution of idiosyncratic amenities (D.43), this new economic geography model exhibits the same gravity equation predictions for commuting flows as in the paper and in the canonical urban model in Section D1 of this online appendix. Following the same analysis as in Section C of this online appendix, the probability that a worker chooses to reside in location $n \in \mathbb{N}$ and work in location $i \in \mathbb{N}$ conditional on choosing a residence-workplace pair in Greater London (λ_{ni}) is given by:

$$\lambda_{ni} = \frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{(B_{ni} w_i)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha} \right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell} w_{\ell})^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha} \right)^{-\epsilon}}, \qquad n, i \in \mathbb{N}, \tag{D.61}$$

which is identical to equation (6) in the paper, except that the price index for goods consumption (P_n) is now determined by equation (D.57).

Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker in Greater London lives in each residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^R = R_n / L_{\mathbb{N}}$). Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker in Greater London is employed in each workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_i^L = L_i / L_{\mathbb{N}}$):

$$\lambda_n^R = \frac{\sum_{i \in \mathbb{N}} \left(B_{ni} w_i\right)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell} w_\ell\right)^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}}, \quad \lambda_i^L = \frac{\sum_{n \in \mathbb{N}} \left(B_{ni} w_i\right)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell} w_\ell\right)^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{D.62}$$

where both expressions are the same as in equation (7) in the paper.

From equations (D.61) and (D.62), the conditional probability that a worker commutes to location i conditional on residing in location n takes the same form as in equation (14) in the paper:

$$\lambda_{ni|n}^R = \frac{\lambda_{ni}}{\lambda_n^R} = \frac{\left(B_{ni} w_i / \kappa_{ni}\right)^{\epsilon}}{\sum_{\ell \in \mathbb{N}} \left(B_{n\ell} w_\ell / \kappa_{n\ell}\right)^{\epsilon}}. \tag{D.63}$$

Using this commuting probability conditional on residence ($\lambda_{ni|n}^R$) from equation (D.63), we obtain an identical expression for per capita income conditional on living in location n as in equation (15) in the paper:

$$v_n = \sum_{i \in \mathbb{N}} \lambda_{ni|n}^R w_i. \tag{D.64}$$

Commuter market clearing again implies that employment in each location (L_i) equals the measure of workers choosing to commute to that location. Using the commuting probabilities conditional on residence from equation (D.63), we obtain the same expression for this commuter market clearing condition as in equation (13) in the paper:

$$L_i = \sum_{n \in \mathbb{N}} \lambda_{ni|n}^R R_n. \tag{D.65}$$

Finally, using the Fréchet distribution for idiosyncratic amenities (D.43), expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is equalized across all residence-workplace pairs in the economy and takes the same form as in equation (8) in the paper:

$$\bar{U} = \vartheta \left[\sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} \left(B_{k\ell} w_\ell\right)^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{D.66}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$; and $\Gamma(\cdot)$ is the Gamma function. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}} / L_{\mathbb{M}}$), we can re-write this population mobility condition as:

$$\bar{U} \left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}}\right)^{\frac{1}{\epsilon}} = \vartheta \left[\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell} w_\ell\right)^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{D.67}$$

where only the limits of the summations differ on the right-hand sides of equations (D.66) and (D.67).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}} / L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

D3.5 Comparative Statics for Changes in Commuting Costs

We now show that this new economic geography model yields the same predictions for the impact of the removal of the railway network on workplace employment and commuting as in the paper and the canonical urban model in Section D1 of this online appendix, once we condition on the observed values of variables in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values.

First, using equation (D.59), the land market clearing condition for any earlier year $\tau < t$ can be written in terms of the observed variables and model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\hat{v}_{nt}v_{nt}\hat{R}_{nt}R_{nt} + \frac{\beta^H}{\beta^L}\hat{w}_{nt}w_{nt}\hat{L}_{nt}L_{nt}, \tag{D.68}$$

where recall that a hat above a variable denotes a relative change, such that $\hat{x}_t = x_\tau / x_t$.

Second, using equations (D.63) and (D.64), expected income by residence (v_{nt}) for any earlier year $\tau < t$ can be written in the same form as equation (19) in the paper:

$$\hat{v}_{nt}v_{nt} = \sum_{i \in \mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it}w_{it}, \tag{D.69}$$

where ($\lambda^R_{nit|n}$, w_{it}, v_{nt}) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.69); and we assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Third, using equations (D.63) and (D.65), workplace employment (L_{it}) for any earlier year $\tau < t$ can be written in the same form as in equation (18) in the paper:

$$\hat{L}_{it}L_{it} = \sum_{n \in \mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{D.70}$$

where ($\lambda^R_{nit|n}$, L_{it}, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.70); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Finally, using equation (D.63), commuting flows ($\hat{L}_{nit}$) for any earlier year $\tau < t$ can be written in the same form as in equation (21) in the paper:

$$\hat{L}_{nit}L_{nit} = \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{D.71}$$

where (L_{nit}, $\lambda^R_{nit|n}$, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.71); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Note that equations (D.68), (D.69), (D.70) and (D.71) above are identical to equations (17), (19), (18) and (21) in the paper. Therefore, given the same observed variables in the initial equilibrium (L_{nt}, R_{nt}, $\mathbb{Q}_{nt}$, w_{nt}, v_{nt}, L_{nit}), the same observed changes in residents and rateable values ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$) and the same estimated changes in commuting costs

($\hat{\kappa}_{nit}$), this new economic geography model predicts the same changes in workplace employment ($\hat{L}_{it}$) and commuting patterns ($\hat{L}_{nit}$) as in the paper and in the canonical urban model in Section D1 of this online appendix.

D4 Ricardian Spatial Model

We next derive our predictions for the impact of the removal of the railway network on workplace employment and commuting in a Ricardian spatial model following Eaton and Kortum (2002) and Redding (2016). We consider a city (Greater London) that is embedded in a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations $\mathbb{M}$. Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and is indexed by t. The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws. With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and workplace equal the realized values.[4] Motivated by our empirical finding that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$. We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

D4.1 Preferences and Endowments

The preferences of a worker ω who lives in location n and works in location i are defined over final goods consumption ($C_n(\omega)$), residential floor space use ($H_n^R(\omega)$), iceberg commuting costs (κ_{ni}), common amenities for all workers (B_{ni}), and an idiosyncratic amenity draw for an individual worker for each residence-workplace pair ($b_{ni}(\omega)$), according to the following Cobb-Douglas functional form:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega)}{\kappa_{ni}} \left(\frac{C_n(\omega)}{\alpha}\right)^{\alpha} \left(\frac{H_n^R(\omega)}{1-\alpha}\right)^{1-\alpha}, \qquad 0 < \alpha < 1, \tag{D.72}$$

where we suppress the time subscript from now onwards, except where important. The idiosyncratic amenities shock for worker ω for each residence n and workplace i ($b_{ni}(\omega)$) is drawn from an independent Fréchet distribution:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{D.73}$$

where we normalize the Fréchet scale parameter in equation (D.73) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities B_{ni} from equation (D.72); the Fréchet shape parameter $\epsilon > 1$ regulates the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller is the shape parameter ϵ, the greater is the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables. All workers ω residing in location n and working in location i receive the same wage and make the same choices for consumption and residential floor space use. Therefore, we suppress the implicit dependence on ω, except where important.

[4] To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

We allow common amenities (B_{ni}) to vary bilaterally to capture the fact that the attractiveness of a given commute may depend on characteristics of both the workplace and the residence. In particular, we decompose this bilateral parameter B_{ni} into a residence component that is common across all workplaces ($\mathcal{B}_n^R$), a workplace component that is common across all residences ($\mathcal{B}_i^L$), and an idiosyncratic component ($\mathcal{B}_{ni}^I$) that is specific to an individual residence-workplace pair:

$$B_{ni} = \mathcal{B}_n^R \mathcal{B}_i^L \mathcal{B}_{ni}^I, \qquad \mathcal{B}_n^R, \mathcal{B}_i^L, \mathcal{B}_{ni}^I > 0. \tag{D.74}$$

We allow the levels of $\mathcal{B}_n^R$, $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ to differ across residences n and workplaces i, although when we examine the impact of the construction of the railway network, we assume that $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ are time invariant. In contrast, we allow B_n^R to change over time, and for these changes to be potentially endogenous to the evolution of the surrounding concentration of economic activity through agglomeration forces,

The goods consumption index for location n takes the constant elasticity of substitution (CES) form and is defined over a fixed continuum of goods $j \in [0, 1]$:

$$C_n = \left[\int_0^1 c_n(j)^\rho dj\right]^{\frac{1}{\rho}}, \tag{D.75}$$

where $c_n(j)$ is consumption of good j in country n; the CES parameter (ρ) determines the elasticity of substitution between goods ($\sigma = 1/(1-\rho) > 1$). The corresponding dual price index for goods consumption (P_n) is:

$$P_n = \left[\int_0^1 p_n(j)^{1-\sigma} dj\right]^{\frac{1}{1-\sigma}}, \qquad \sigma = \frac{1}{1-\rho} > 1, \tag{D.76}$$

where $p_n(j)$ is the price of good j in country n.

D4.2 Production

Each good j can be produced in each location i under conditions of perfect competition using labor, machinery capital and commercial floor space, where commercial floor space includes both building capital and land. The production technology is assumed to take the Cobb-Douglas form. If a good is produced by a location, the requirement of zero profits implies that the good's "free on board" price must equal its constant unit cost of production:

$$p_i(j) = \frac{w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}}{z_i(j)}, \qquad 0 < \beta^L, \beta^H, \beta^M < 1, \qquad \beta^L + \beta^H + \beta^M = 1, \tag{D.77}$$

where w_i denotes the wage; q_i represents the price of commercial floor space in location i; machinery is assumed to be perfectly mobile across locations with a common price r determined in the wider economy; $z_i(j)$ is productivity; and to focus on Ricardian reasons for trade, we assume that factor intensity is the same for all goods, as controlled by (β^L, β^H, β^M).

Each location i draws an idiosyncratic productivity $z_i(j)$ for each good j from an independent Fréchet distribution:

$$F_i(z) = e^{-A_i z^{-\theta}}, \qquad A_i > 0, \quad \theta > 1, \tag{D.78}$$

where the scale parameter A_i determines average productivity for location i and the shape parameter θ controls the dispersion of productivity across goods. We allow this scale parameter that determines each location's average productivity to potentially respond endogenously to changes in the surrounding concentration of economic activity through agglomeration forces.

Goods can be traded between locations subject to iceberg variable trade costs, such that $d_{ni} \geq 1$ units of a good must be shipped from location i in order for one unit to arrive in location n (where $d_{ni} > 1$ for $n \neq i$ and $d_{nn} = 1$). The "cost inclusive of freight" price of a good in the location of consumption n ($p_{ni}(j)$) is thus a constant multiple of the "free on board" price of that good in the location of production i ($p_i(j)$) with that multiple determined by the iceberg trade costs:

$$p_{ni}(j) = d_{ni} p_i(j). \tag{D.79}$$

Combining equations (D.77) and (D.79), the cost to a consumer in location n of purchasing one unit of good j from location i is given by:

$$p_{ni}(j) = \frac{d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M}}{z_i(j)}. \tag{D.80}$$

From profit maximization and zero profits, we obtain the results in equations (10) and (12) in the paper that payments to labor, commercial floor space and machinery capital are constant shares of revenue:

$$w_i L_i = \beta^L X_i, \qquad q_i H_i^L = \beta^H X_i, \qquad r M_i = \beta^M X_i, \tag{D.81}$$

where L_i is workplace employment; X_i denotes revenue; H_i^L represents commercial use of floor space; and M_i is machinery capital use. Therefore, payments for commercial floor space are proportional to workplace income:

$$q_i H_i^L = \frac{\beta^H}{\beta^L} w_i L_i. \tag{D.82}$$

D4.3 Trade and Market Clearing

We assume that floor space is owned by landlords, who receive income from residents' expenditure on floor space, and consume only consumption goods where they live. Total expenditure on consumption goods equals the fraction α of the total income of residents plus the entire income of landlords. This income of landlords equals $(1-\alpha)$ times the total income of residents plus β^H times revenue (which equals β^H/β^L times the total income of workers). Therefore total expenditure on consumption goods is:

$$E_n = P_n C_n = \alpha v_n R_n + (1-\alpha) v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n = v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n,$$

where v_n is the average income of location n's residents, as determined below as a function of commuting patterns, and R_n is the measure of these residents.

This Ricardian spatial model also implies a gravity equation for bilateral trade in goods between locations. Goods are homogeneous in the sense that one unit of a given good is the same as any other unit of that good. Therefore, the representative consumer in a given location sources each good from the lowest-cost supplier to that location. Using equilibrium prices (D.80) and the properties of the Fréchet distribution following Eaton and Kortum (2002), the share of the expenditure of location n on goods produced by location i is:

$$\pi_{ni} = \frac{A_i \left(d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} \right)^{-\theta}}{\sum_{k \in \mathbb{M}} A_k \left(d_{nk} w_k^{\beta^L} q_k^{\beta^H} r^{\beta^M} \right)^{-\theta}}, \tag{D.83}$$

where the elasticity of trade flows to trade costs is determined by the Fréchet shape parameter for productivity θ.

Goods market clearing and zero profits imply that payments to workers plus payments for commercial floor space use plus payments for machinery use in each location equal expenditure on goods produced in that location:

$$w_i L_i + q_i H_i^L + r M_i = \sum_{n \in \mathbb{M}} \pi_{ni} E_n. \tag{D.84}$$

Using equilibrium prices (D.80) and the properties of the Fréchet distribution, the consumption goods price index in equation (D.76) can be rewritten as:

$$P_n = \gamma \left[\sum_{i \in \mathbb{M}} A_i \left(d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} \right)^{-\theta} \right]^{-\frac{1}{\theta}}, \tag{D.85}$$

where $\gamma \equiv \left[\Gamma \left(\frac{\theta - (\sigma - 1)}{\theta} \right) \right]^{\frac{1}{1-\sigma}}$; $\Gamma(\cdot)$ denotes the Gamma function; and we require $\theta > \sigma - 1$ to ensure a finite value for the price index.

Using the trade share (D.83), and noting that $d_{nn} = 1$, the consumption goods price index in equation (D.85) can be further rewritten solely in terms of the domestic trade share (π_{nn}), wages (w_n), the price of commercial floor space (q_n), the common price of machinery capital (r) and parameters:

$$P_n = \gamma \left(\frac{A_n}{\pi_{nn}} \right)^{-\frac{1}{\theta}} \left(w_n^{\beta^L} q_n^{\beta^H} r^{\beta^M} \right). \tag{D.86}$$

Land market clearing implies that the total income received by landlords as owners of floor space (which equals rateable value ($\mathbb{Q}_n$) in our data) equals the sum of payments for the use of residential and commercial floor space:

$$\mathbb{Q}_n = Q_n H_n^R + q_n H_n^L = (1 - \alpha) v_n R_n + \left(\frac{\beta^H}{\beta^L} \right) w_n L_n, \tag{D.87}$$

where H_n^R denotes residential floor space use.

Importantly, we allow the supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) to be endogenous, and we allow the prices of residential and commercial floor space to potentially differ from one another through the wedge ξ_i ($q_i = \xi_i Q_i$). In our baseline quantitative analysis below, we are not required to make assumptions about these supplies of residential and commercial floor space or this wedge between commercial and residential floor prices. The reason is that we condition on the observed rateable values in the data ($\mathbb{Q}_n$) and the supplies and prices for residential and commercial floor space (H_n^R, H_n^L, Q_n, q_n) only enter the land market clearing condition (D.87) through these observed rateable values.

D4.4 Workplace and Residence Choices

Given the direct utility function (D.72), the corresponding indirect utility function for a worker ω residing in location n and working in location i is:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega) w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}, \tag{D.88}$$

which takes the same form as equation (3) in the paper and equation (D.1) in the canonical urban model in Section D1 of this online appendix. The only difference from the canonical urban model is in the underlying determinants of the price index for goods consumption (P_n), as now specified in equation (D.85).

Using indirect utility (D.88) and the Fréchet distribution of idiosyncratic amenities (D.73), this Ricardian spatial model exhibits the same gravity equation predictions for commuting flows as in the paper and in the canonical urban model in Section D1 of this online appendix. Following the same analysis as in Section C of this online appendix, the probability that a worker chooses to reside in location $n \in \mathbb{N}$ and work in location $i \in \mathbb{N}$ conditional on choosing a residence-workplace pair in Greater London (λ_{ni}) is given by:

$$\lambda_{ni} = \frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{(B_{ni}w_i)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}(B_{k\ell}w_{\ell})^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \qquad n,i \in \mathbb{N}, \tag{D.89}$$

which is identical to equation (6) in the paper, except that the price index for goods consumption (P_n) is now determined by equation (D.85).

Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker in Greater London lives in each residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^R = R_n/L_{\mathbb{N}}$). Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker in Greater London is employed in each workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_i^L = L_i/L_{\mathbb{N}}$):

$$\lambda_n^R = \frac{\sum_{i\in\mathbb{N}}(B_{ni}w_i)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}(B_{k\ell}w_{\ell})^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \quad \lambda_i^L = \frac{\sum_{n\in\mathbb{N}}(B_{ni}w_i)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}(B_{k\ell}w_{\ell})^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{D.90}$$

where both expressions are the same as in equation (7) in the paper.

From equations (D.89) and (D.90), the conditional probability that a worker commutes to location i conditional on residing in location n takes the same form as in equation (14) in the paper:

$$\lambda_{ni|n}^R = \frac{\lambda_{ni}}{\lambda_n^R} = \frac{(B_{ni}w_i/\kappa_{ni})^{\epsilon}}{\sum_{\ell\in\mathbb{N}}(B_{n\ell}w_{\ell}/\kappa_{n\ell})^{\epsilon}}. \tag{D.91}$$

Using this commuting probability conditional on residence ($\lambda_{ni|n}^R$) from equation (D.91), we obtain an identical expression for per capita income conditional on living in location n as in equation (15) in the paper:

$$v_n = \sum_{i\in\mathbb{N}} \lambda_{ni|n}^R w_i. \tag{D.92}$$

Commuter market clearing again implies that employment in each location (L_i) equals the measure of workers choosing to commute to that location. Using the commuting probabilities conditional on residence from equation (D.91), we obtain the same expression for this commuter market clearing condition as in equation (13) in the paper:

$$L_i = \sum_{n\in\mathbb{N}} \lambda_{ni|n}^R R_n. \tag{D.93}$$

Finally, using the Fréchet distribution for idiosyncratic amenities (D.73), expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is equalized across all residence-workplace pairs in the economy and takes the same form as in equation (8) in the paper:

$$\bar{U} = \vartheta\left[\sum_{k\in\mathbb{M}}\sum_{\ell\in\mathbb{M}}(B_{k\ell}w_{\ell})^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{D.94}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$; and $\Gamma(\cdot)$ is the Gamma function. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}}/L_{\mathbb{M}}$), we can re-write this population mobility condition as:

$$\bar{U}\left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}}\right)^{\frac{1}{\epsilon}} = \vartheta\left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}(B_{k\ell}w_{\ell})^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{D.95}$$

where only the limits of the summations differ on the right-hand sides of equations (D.94) and (D.95).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

D4.5 Comparative Statics for Changes in Commuting Costs

We now show that this Ricardian spatial model yields the same predictions for the impact of removing the railway network on workplace employment and commuting as in the paper and the canonical urban model in Section D1 of this online appendix, once we condition on the observed variables in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values.

First, using equation (D.87), the land market clearing condition for any earlier year $\tau < t$ can be written in terms of observed variables and model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\hat{v}_{nt}v_{nt}\hat{R}_{nt}R_{nt} + \frac{\beta^H}{\beta^L}\hat{w}_{nt}w_{nt}\hat{L}_{nt}L_{nt}, \tag{D.96}$$

where recall that a hat above a variable denotes a relative change, such that $\hat{x}_t = x_\tau/x_t$.

Second, using equations (D.91) and (D.92), expected income by residence (v_{nt}) for any earlier year $\tau < t$ can be written in the form as equation (19) in the paper:

$$\hat{v}_{nt}v_{nt} = \sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^\epsilon_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^\epsilon_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it}w_{it}, \tag{D.97}$$

where ($\lambda^R_{nit|n}$, w_{it}, v_{nt}) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.97); and we assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Third, using equations (D.91) and (D.93), workplace employment (L_{it}) for any earlier year $\tau < t$ can be written in the same form as equation (18) in the paper:

$$\hat{L}_{it}L_{it} = \sum_{n\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^\epsilon_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^\epsilon_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{D.98}$$

where ($\lambda^R_{nit|n}$, L_{it}, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.98); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Finally, using equation (D.91), commuting flows ($\hat{L}_{nit}$) for any earlier year $\tau < t$ can be written in an analogous form as follows:

$$\hat{L}_{nit}L_{nit} = \frac{\lambda^R_{nit|n}\hat{w}^\epsilon_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^\epsilon_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{D.99}$$

where (L_{nit}, $\lambda^R_{nit|n}$, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.99); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}_{it}^{L} = 1$, $\hat{\mathcal{B}}_{nit}^{I} = 1$).

Note that equations (D.96), (D.97), (D.98) and (D.99) are identical to equations (17), (19), (18) and (21) in the paper. Therefore, given the same observed variables in the initial equilibrium (L_{nt}, R_{nt}, $\mathbb{Q}_{nt}$, w_{nt}, v_{nt}, L_{nit}), the same observed changes in residents and rateable values ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$) and the same estimated changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), this Ricardian spatial model predicts the same changes in workplace employment ($\hat{L}_{it}$) and commuting patterns ($\hat{L}_{nit}$) as in the paper and the canonical urban model in Section D1 of this online appendix.

D5 Armington Model

Finally, we derive our predictions for the impact of the removal of the railway network in an Armington spatial model following Armington (1969), Allen and Arkolakis (2014) and Allen, Arkolakis and Li (2017). We consider a city (Greater London) that is embedded within a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations $\mathbb{M}$. Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and indexed by t. The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws. With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and workplace equal the realized values.[5] Motivated by our empirical finding that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$. We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

D5.1 Preferences and Endowments

The preferences of a worker ω who lives in location n and works in location i are defined over final goods consumption ($C_n(\omega)$), residential floor space use ($H_n^R(\omega)$), iceberg commuting costs (κ_{ni}), common bilateral amenities for all workers (B_{ni}), and an idiosyncratic amenity draw for an individual worker for each residence-workplace pair ($b_{ni}(\omega)$), according to the Cobb-Douglas functional form:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega)}{\kappa_{ni}} \left(\frac{C_n(\omega)}{\alpha} \right)^{\alpha} \left(\frac{H_n^R(\omega)}{1-\alpha} \right)^{1-\alpha}, \qquad 0 < \alpha < 1, \tag{D.100}$$

where we suppress the time subscript from now onwards, except where important. The idiosyncratic amenities shock for worker ω for each residence n and workplace i ($b_{ni}(\omega)$) is drawn from an independent Fréchet distribution:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{D.101}$$

where we normalize the Fréchet scale parameter in equation (D.101) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities (B_{ni}) from equation (D.100); the Fréchet shape parameter $\epsilon > 1$

[5] To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

regulates the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller is the shape parameter ϵ, the greater is the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables. All workers ω residing in location n and working in location i receive the same wage and make the same choices for consumption and residential floor space use. Therefore, we suppress the implicit dependence on ω from now onwards, except where important.

We allow common amenities (B_{ni}) to vary bilaterally to capture the fact that the attractiveness of a given commute may depend on characteristics of both the workplace and the residence. In particular, we decompose this bilateral parameter B_{ni} into a residence component that is common across all workplaces ($\mathcal{B}_n^R$), a workplace component that is common across all residences ($\mathcal{B}_i^L$), and an idiosyncratic component ($\mathcal{B}_{ni}^I$) that is specific to an individual residence-workplace pair:

$$B_{ni} = \mathcal{B}_n^R \mathcal{B}_i^L \mathcal{B}_{ni}^I, \qquad \mathcal{B}_n^R, \mathcal{B}_i^L, \mathcal{B}_{ni}^I > 0. \tag{D.102}$$

We allow the levels of $\mathcal{B}_n^R$, $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ to differ across residences n and workplaces i, although when we examine the impact of the construction of the railway network, we assume that $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ are time invariant. In contrast, we allow B_n^R to change over time, and for these changes to be potentially endogenous to the evolution of the surrounding concentration of economic activity through agglomeration forces,

Consumption goods are assumed to be differentiated by location of origin according to the constant elasticity of substitution (CES) functional form. Therefore the consumption index in location n is:

$$C_n = \left[\sum_{i \in \mathbb{M}} c_{ni}^{\rho}\right]^{\frac{1}{\rho}}, \tag{D.103}$$

where c_{ni} denotes consumption in location n of the good produced by location i; and the CES parameter (ρ) determines the elasticity of substitution between the goods produced by each location ($\sigma = 1/(1-\rho) > 1$).

In this specification with differentiation by location of origin, the CES functional form implies that the marginal utility of consuming a location's good approaches infinity as consumption of that good converges to zero. Therefore, in equilibrium, each location consumes the goods produced by all locations. Using the properties of the CES functional form, the corresponding dual price index for goods consumption (P_n) is:

$$P_n = \left[\sum_{i \in \mathbb{M}} p_{ni}^{1-\sigma}\right]^{\frac{1}{1-\sigma}}, \qquad \sigma = \frac{1}{1-\rho} > 1, \tag{D.104}$$

where p_{ni} denotes the price in country n of the good produced by country i.

D5.2 Production

Goods from each location of origin are produced under conditions of perfect competition using labor, machinery capital and commercial floor space, where commercial floor space includes both building capital and land. We assume that the production technology takes the Cobb-Douglas form. Using zero profits and the fact that the goods of all locations are consumed and produced in equilibrium, the "free on board" price of each location's good equals its constant unit cost of production:

$$p_i = w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} / A_i, \qquad 0 < \beta^L, \beta^H, \beta^M < 1, \qquad \beta^L + \beta^H + \beta^M = 1, \tag{D.105}$$

where A_i denotes productivity; w_i is the wage; q_i corresponds to the price of commercial floor space; and machinery is assumed to be perfectly mobile across locations with a common price r determined in the wider economy. We allow productivity A_i in each location to respond endogenously to changes in the surrounding concentration of economic activity through agglomeration forces.

Goods can be traded between locations subject to iceberg variable trade costs, such that $d_{ni} \geq 1$ units of a good must be shipped from location i in order for one unit to arrive in location n (where $d_{ni} > 1$ for $n \neq i$ and $d_{nn} = 1$). The "cost inclusive of freight" price of a good in the location of consumption n (p_{ni}) is thus a constant multiple of the "free on board" price of that good in the location of production i (p_i) with that multiple determined by the iceberg trade costs:

$$p_{ni} = d_{ni} p_i. \tag{D.106}$$

Combining equations (D.105) and (D.106), the cost to the consumer in location n of purchasing the good produced by location i is:

$$p_{ni} = d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} / A_i. \tag{D.107}$$

From profit maximization and zero profits, we obtain the results in equations (10) and (12) in the paper that payments to labor, commercial floor space and machinery capital are constant shares of revenue:

$$w_i L_i = \beta^L X_i, \qquad q_i H_i^L = \beta^H X_i, \qquad r M_i = \beta^M X_i, \tag{D.108}$$

where L_i is employment; X_i is revenue; H_i^L denotes commercial floor space use; and M_i corresponds to machinery capital use. Therefore, payments for commercial floor space are proportional to workplace income ($w_i L_i$):

$$q_i H_i^L = \frac{\beta^H}{\beta^L} w_i L_i. \tag{D.109}$$

D5.3 Trade and Market Clearing

We assume that floor space is owned by landlords, who receive income from residents' expenditure on floor space, and consume only consumption goods where they live. Total expenditure on consumption goods equals the fraction α of the total income of residents plus the entire income of landlords. This income of landlords equals $(1-\alpha)$ times the total income of residents plus β^H times revenue (which equals β^H/β^L times the total income of workers). Therefore, total expenditure on consumption goods is:

$$E_n = P_n C_n = \alpha v_n R_n + (1-\alpha) v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n = v_n R_n + \frac{\beta^H}{\beta^L} w_n L_n, \tag{D.110}$$

where v_n is the average income of location n's residents, as determined below as a function of commuting patterns, and R_n is the measure of these residents.

This Armington model also implies a gravity equation for bilateral trade in goods between locations. Using again the properties of CES preferences, the share of expenditure of location n on goods produced by location i is:

$$\pi_{ni} = \frac{\left(d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} / A_i\right)^{1-\sigma}}{\sum_{k \in \mathbb{M}} \left(d_{nk} w_k^{\beta^L} q_k^{\beta^H} r^{\beta^M} / A_k\right)^{1-\sigma}}, \tag{D.111}$$

where the elasticity of trade to trade costs $(1-\sigma)$ is now determined by the elasticity of substitution between the goods produced by each location.

Goods market clearing and zero profits imply that payments to workers plus payments for commercial floor space use plus payments for machinery capital use in each location equal expenditure on goods produced in that location:

$$w_i L_i + q_i H_i^L + r M_i = \sum_{n \in \mathbb{N}} \pi_{ni} X_n. \tag{D.112}$$

We now use the expression for the equilibrium price of each location's good in equation (D.107) to rewrite the consumption goods price index in equation (D.104) as follows:

$$P_n = \left[\sum_{i \in \mathbb{M}} \left(d_{ni} w_i^{\beta^L} q_i^{\beta^H} r^{\beta^M} / A_i \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \tag{D.113}$$

Using the trade share (D.111), and noting that $d_{nn} = 1$, the consumption goods price index in equation (D.113) can be further rewritten solely in terms of the domestic trade share (π_{nn}), wages (w_n), the price of commercial floor space (q_i), the common price of machinery (r), and parameters:

$$P_n = \left(\frac{1}{\pi_{nn}} \right)^{\frac{1}{1-\sigma}} \left(\frac{w_n^{\beta^L} q_n^{\beta^H} r^{\beta^M}}{A_n} \right). \tag{D.114}$$

Land market clearing implies that the total income received by landlords as owners of floor space (which equals rateable value ($\mathbb{Q}_n$) in our data) equals the sum of payments for the use of residential and commercial floor space:

$$\mathbb{Q}_n = Q_n H_n^R + q_n H_n^L = (1-\alpha) v_n R_n + \left(\frac{\beta^H}{\beta^L} \right) w_n L_n, \tag{D.115}$$

where H_n^R denotes residential floor space use.

Importantly, we allow the supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) to be endogenous, and we allow the prices of residential and commercial floor space to potentially differ from one another through the wedge ξ_i ($q_i = \xi_i Q_i$). In our baseline quantitative analysis below, we are not required to make assumptions about these supplies of residential and commercial floor space or this wedge between commercial and residential floor prices. The reason is that we condition on the observed rateable values in the data ($\mathbb{Q}_n$) and the supplies and prices for residential and commercial floor space (H_n^R, H_n^L, Q_n, q_n) only enter the land market clearing condition (D.115) through these observed rateable values.

D5.4 Workplace and Residence Choices

Given the direct utility function (D.100), the corresponding indirect utility function for a worker ω residing in location n and working in location i is:

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega) w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}, \tag{D.116}$$

which takes the same form as equation (3) in the paper and equation (D.1) in the canonical urban model in Section D1 of this online appendix. The only difference from the canonical urban model is in the underlying determinants of the price index for goods consumption (P_n), as now determined by equation (D.113).

Using indirect utility (D.116) and the Fréchet distribution of idiosyncratic amenities (D.101), this Armington spatial model exhibits the same gravity equation predictions for commuting flows as in the paper and the canonical urban model in Section D1 of this online appendix. Following the same analysis as in Section C of this online appendix, the probability that a worker chooses to reside in location $n \in \mathbb{N}$ and work in location $i \in \mathbb{N}$ conditional on choosing a residence-workplace pair in Greater London (λ_{ni}) is given by:

$$\lambda_{ni} = \frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \qquad n,i\in\mathbb{N}, \tag{D.117}$$

which is identical to equation (6) in the paper, except that the price index for goods consumption (P_n) is now determined by equation (D.113).

Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker in Greater London lives in each residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^R = R_n/L_{\mathbb{N}}$). Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker in Greater London is employed in each workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^L = L_i/L_{\mathbb{N}}$):

$$\lambda_n^R = \frac{\sum_{i\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \quad \lambda_i^L = \frac{\sum_{n\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{D.118}$$

where both expressions are the same as in equation (7) in the paper.

From equations (D.117) and (D.118), the conditional probability that a worker commutes to location i conditional on residing in location n takes the same form as in equation (14) in the paper:

$$\lambda_{ni|n}^R = \frac{\lambda_{ni}}{\lambda_n^R} = \frac{\left(B_{ni}w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{\ell\in\mathbb{N}}\left(B_{n\ell}w_{\ell}/\kappa_{n\ell}\right)^{\epsilon}}. \tag{D.119}$$

Using this commuting probability conditional on residence ($\lambda_{ni|n}^R$) from equation (D.119), we obtain an identical expression for per capita income conditional on living in location n as in equation (15) in the paper:

$$v_n = \sum_{i\in\mathbb{N}} \lambda_{ni|n}^R w_i. \tag{D.120}$$

Commuter market clearing again implies that employment in each location (L_i) equals the measure of workers choosing to commute to that location. Using the commuting probabilities conditional on residence from equation (D.119), we obtain the same expression for this commuter market clearing condition as in equation (13) in the paper:

$$L_i = \sum_{n\in\mathbb{N}} \lambda_{ni|n}^R R_n. \tag{D.121}$$

Finally, using the Fréchet distribution for idiosyncratic amenities (D.101), expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is equalized across all residence-workplace pairs in the economy and takes the same form as in equation (8) in the paper:

$$\bar{U} = \vartheta\left[\sum_{k\in\mathbb{M}}\sum_{\ell\in\mathbb{M}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{D.122}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$; and $\Gamma(\cdot)$ is the Gamma function. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}}/L_{\mathbb{M}}$), we can re-write this population mobility condition as:

$$\bar{U}\left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}}\right)^{\frac{1}{\epsilon}} = \vartheta\left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_{\ell}\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{D.123}$$

where only the limits of the summations differ on the right-hand sides of equations (D.122) and (D.123).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

D5.5 Comparative Statics for Changes in Commuting Costs

We now show that this Armington model yields exactly the same predictions for the impact of the removal of the railway network on workplace employment and commuting as in the paper and the canonical urban model in Section D1 of this online appendix, once we condition on the observed variables in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values.

First, using equation (D.115), the land market clearing condition for any earlier year $\tau < t$ can be written in terms of observed variables and model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\hat{v}_{nt}v_{nt}\hat{R}_{nt}R_{nt} + \frac{\beta^H}{\beta^L}\hat{w}_{nt}w_{nt}\hat{L}_{nt}L_{nt}, \tag{D.124}$$

where recall that a hat above a variable denotes a relative change, such that $\hat{x}_t = x_\tau/x_t$.

Second, using equations (D.119) and (D.120), expected income by residence (v_{nt}) for any earlier year $\tau < t$ can be written in the same form as equation (19) in the paper:

$$\hat{v}_{nt}v_{nt} = \sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it}w_{it}, \tag{D.125}$$

where ($\lambda^R_{nit|n}$, w_{it}, v_{nt}) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.125); and we assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Third, using equations (D.119) and (D.121), workplace employment (L_{it}) for any earlier year $\tau < t$ can be written in the same form as equation (18) in the paper:

$$\hat{L}_{it}L_{it} = \sum_{n\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{D.126}$$

where ($\lambda^R_{nit|n}$, L_{it}, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.126); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Finally, using equation (D.119), commuting flows ($\hat{L}_{nit}$) for any earlier year $\tau < t$ can be written in the same form as in equation (21) in the paper:

$$\hat{L}_{nit}L_{nit} = \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}, \tag{D.127}$$

where (L_{nit}, $\lambda^R_{nit|n}$, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (D.127); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Note that equations (D.124), (D.125), (D.126) and (D.127) are identical to equations (17), (19), (18) and (21) in the paper. Therefore, given the same observed variables in the initial equilibrium (L_{nt}, R_{nt}, $\mathbb{Q}_{nt}$, w_{nt}, v_{nt}, L_{nit}), the same observed changes in residents and rateable values ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$) and the same estimated changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), this Armington spatial model predicts the same changes in workplace employment ($\hat{L}_{it}$) and commuting patterns ($\hat{L}_{nit}$) as in the paper and the canonical urban model in Section D1 of this online appendix.

E Effective Units of Labor

In this section of the online appendix, we consider an alternative interpretation of the idiosyncratic shock to worker commuting decisions. In our baseline specification in the paper, we interpret this shock as capturing amenities or preferences. In contrast, we now consider an alternative specification, in which we interpret this shock as capturing effective units of labor or worker ability. Although the relationship between worker income and wages is different in this alternative specification, we show that our baseline quantitative analysis takes a similar form as in the paper.

Under this alternative specification, the idiosyncratic draw ($b_{nit}(\omega)$) no longer enters the direct utility function. We assume that the preferences of worker ω residing in n and working in i at time t are given by:

$$U_{nit}(\omega) = \frac{B_{ni}}{\kappa_{nit}} \left(\frac{C_{nt}(\omega)}{\alpha} \right)^{\alpha} \left(\frac{H^R_{nt}(\omega)}{1-\alpha} \right)^{1-\alpha}, \tag{E.1}$$

where $C_{nt}(\omega)$ is a consumption goods index; $H^R_{nt}(\omega)$ is residential use of floor space; and B_{ni} captures amenities from the bilateral commute from residence n to workplace i that are common across all workers. Nevertheless, the idiosyncratic draw continues to enter the indirect utility function in exactly the same form as in our baseline specification in equation (3) in the paper, because worker income now depends on the wage per unit of ability (w_{it}) times the realization for worker ability ($b_{nit}(\omega)$):

$$U_{nit}(\omega) = \frac{B_{ni} b_{nit}(\omega) w_{it}}{\kappa_{nit} P^{\alpha}_{nt} Q^{1-\alpha}_{n}}. \tag{E.2}$$

We assume that idiosyncratic worker ability ($b_{ni}(\omega)$) is drawn from an independent extreme value (Fréchet) distribution for each residence-workplace pair and each worker:

$$G_{ni}(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{E.3}$$

where we normalize the Fréchet scale parameter in equation (E.3) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities B_{ni} from equation (E.2); the Fréchet shape parameter ϵ regulates the dispersion of idiosyncratic worker ability, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller the shape parameter ϵ, the greater the heterogeneity in idiosyncratic worker ability, and the less sensitive are worker location decisions to economic variables.

The main difference between this alternative specification in equations (E.2) and (E.3) and our baseline specification in the paper is the relationship between worker income and wages. In our baseline specification in the paper, the income of a worker employed in workplace i is equal to her wage (w_{it}). In contrast, in this alternative specification, income for worker ω residing in n and employed in i at time t ($\tilde{w}_{nit}(\omega)$) is equal to her wage per unit of ability at her workplace i

(w_{it}) times her idiosyncratic draw for ability $(b_{nit}(\omega))$:

$$\tilde{w}_{nit}(\omega) = w_{it} b_{nit}(\omega). \tag{E.4}$$

Using the monotonic relationship between utility and idiosyncratic worker ability in equation (E.2), the distribution of utility for residence n and workplace i takes the same form as in our baseline specification in the paper:

$$G_{nit}(u) = e^{-\Psi_{nit} u^{-\epsilon}}, \qquad \Psi_{nit} \equiv (B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} P_{nt}^{\alpha} Q_{nt}^{1-\alpha}\right)^{-\epsilon}. \tag{E.5}$$

Following an analogous analysis as for idiosyncratic amenities, the distribution of utility conditional on choosing each residence-workplace pair also takes a similar form as in our baseline specification in the paper and is again equal to the distribution of utility across all possible residence-workplace pairs:

$$G_t(u) = e^{-\Psi_{\mathbb{M}t} u^{-\epsilon}}, \qquad \Psi_{\mathbb{M}t} = \sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} \Psi_{k\ell t}. \tag{E.6}$$

Given these distributions for utility, the probability that a worker commutes from residence n to workplace i is:

$$\frac{L_{nit}}{L_{\mathbb{M}t}} = \frac{(B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} P_{nt}^{\alpha} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} P_{kt}^{\alpha} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}, \tag{E.7}$$

where L_{nit} is the measure of commuters from residence n to workplace i; $L_{\mathbb{M}t}$ is the measure of workers in the economy as a whole; and we assume prohibitive commuting costs across the boundaries of Greater London, such that each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy.

Summing across workplaces and residences in Greater London, the probability that a worker chooses a residence-workplace pair in Greater London is given by:

$$\frac{L_{\mathbb{N}t}}{L_{\mathbb{M}t}} = \frac{\sum_{n \in \mathbb{N}} \sum_{i \in \mathbb{N}} (B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} P_{nt}^{\alpha} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} P_{kt}^{\alpha} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}. \tag{E.8}$$

Dividing equation (E.7) by equation (E.8), we obtain the probability that a worker chooses to live in location n and work in location i, conditional on choosing a residence-workplace pair in Greater London (λ_{nit}):

$$\lambda_{nit} = \frac{L_{nit}/L_{\mathbb{M}t}}{L_{\mathbb{N}t}/L_{\mathbb{M}t}} = \frac{L_{nit}}{L_{\mathbb{N}t}} = \frac{(B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} P_{nt}^{\alpha} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} P_{kt}^{\alpha} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}, \qquad n, i \in \mathbb{N}, \tag{E.9}$$

which takes the same form as in equation (6) in the paper.

Summing across workplaces $i \in \mathbb{N}$ in equation (E.9), we obtain the probability that a worker in Greater London chooses to live in residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London $(R_{nt}/L_{\mathbb{N}t})$:

$$\lambda_{nt}^{R} = \frac{R_{nt}}{L_{\mathbb{N}t}} = \frac{\sum_{\ell \in \mathbb{N}} (B_{n\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{n\ell t} P_{nt}^{\alpha} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} P_{kt}^{\alpha} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}, \tag{E.10}$$

which takes the same form as in equation (7) in the paper. Similarly, summing across residences $n \in \mathbb{N}$ in equation (E.9), we obtain the probability that a worker in Greater London chooses to work in workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London $(L_{it}/L_{\mathbb{N}t})$:

$$\lambda_{it}^{L} = \frac{L_{it}}{L_{\mathbb{N}t}} = \frac{\sum_{k \in \mathbb{N}} (B_{kit} w_{it})^{\epsilon} \left(\kappa_{kit} P_{kt}^{\alpha} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} P_{kt}^{\alpha} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}, \tag{E.11}$$

which takes the same form as in equation (7) in the paper.

Using equations (E.9), (E.10) and (E.11), the conditional probability of commuting to workplace i conditional on living in residence n also takes the same form as in the paper:

$$\lambda^R_{nit|n} = \frac{\lambda_{nit}}{\lambda^R_{nt}} = \frac{(B_{nit} w_{it}/\kappa_{nit})^{\epsilon}}{\sum_{\ell \in \mathbb{N}} (B_{n\ell t} w_{\ell t}/\kappa_{n\ell t})^{\epsilon}}. \tag{E.12}$$

Similarly, the conditional probability of commuting from residence n to workplace i conditional on working in workplace i takes the same form as in the paper:

$$\lambda^L_{nit|i} = \frac{\lambda_{nit}}{\lambda^L_{it}} = \frac{\left(B_{nit}\kappa_{nit} P^{\alpha}_{nt} Q^{1-\alpha}_{nt}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \left(B_{kit}\kappa_{kit} P^{\alpha}_{kt} Q^{1-\alpha}_{kt}\right)^{-\epsilon}}. \tag{E.13}$$

Using the indirect utility function (E.2) and the distribution of utility conditional on choosing each residence-workplace pair (E.6), the distribution of ability conditional on choosing a given workplace i and residence n is:

$$G_{nit}(b) = e^{-\Phi_{nit} b^{-\epsilon}}, \tag{E.14}$$

$$\Phi_{nit} \equiv \left[\frac{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} P^{\alpha}_{kt} Q^{1-\alpha}_{kt}\right)^{-\epsilon}}{(B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} P^{\alpha}_{nt} Q^{1-\alpha}_{nt}\right)^{-\epsilon}} \right] = \frac{L_{\mathbb{M}t}}{L_{nit}}. \tag{E.15}$$

Given this distribution, expected worker ability conditional on choosing workplace i and residence n is:

$$\mathbb{E}_{nit}\left[b\right] = \int_0^{\infty} \epsilon \Phi_{nit} b^{-\epsilon} e^{-\Phi_{nit} b^{-\epsilon}} db. \tag{E.16}$$

Now we define the following change of variables:

$$y = \Phi_{nit} b^{-\epsilon}, \qquad dy = -\epsilon \Phi_{nit} b^{-(\epsilon+1)} db. \tag{E.17}$$

Using this change of variables, expected worker ability conditional on choosing workplace i and residence n can be written as:

$$\mathbb{E}_{nit}\left[b\right] = \int_0^{\infty} \Phi^{1/\epsilon}_{nit} y^{-1/\epsilon} e^{-y} dy, \tag{E.18}$$

which can be in turn written as:

$$\mathbb{E}_{nit}\left[b\right] = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \Phi^{1/\epsilon}_{nit} = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \left[\frac{L_{\mathbb{M}t}}{L_{nit}}\right]^{\frac{1}{\epsilon}} = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right) \left[\frac{L_{\mathbb{M}t}}{L_{\mathbb{N}t}} \frac{1}{\lambda_{nit}}\right]^{\frac{1}{\epsilon}}, \tag{E.19}$$

where $\Gamma(\cdot)$ is the Gamma function. Intuitively, the larger the share of workers in Greater London commuting from residence n to workplace i (λ_{nit}), and the larger the share of the economy's workers in Greater London ($L_{\mathbb{N}t}/L_{\mathbb{M}t}$), the lower the average ability of these workers in equation (E.19).

Using equation (E.19), we can also compute the corresponding value of average worker income conditional on choosing workplace i and residence n as:

$$\tilde{w}_{nit} = w_{it} \mathbb{E}_{nit}\left[b\right], \tag{E.20}$$

From equation (E.19) and the commuting probabilities conditional on residence (E.13), overall average worker ability (across all residences) conditional on working in workplace i is:

$$\mathbb{E}_{it}\left[b\right] = \sum_{n \in \mathbb{N}} \lambda^L_{nit|i} \mathbb{E}_{nit}\left[b\right]. \tag{E.21}$$

Using equation (E.21), we can also compute the corresponding value of overall average worker income (across all residences) conditional on working in workplace i as:

$$\tilde{w}_{it} = w_{it} \left[\sum_{n \in \mathbb{N}} \lambda^L_{nit|i} \mathbb{E}_{nit}\left[b\right]\right]. \tag{E.22}$$

Despite the different relationship between a worker's income and her wage in this alternative specification with idiosyncratic shocks to worker ability, our combined land and commuter market clearing condition takes a similar form to equation (16) in our baseline specification in the paper:

$$\mathbb{Q}_{nt} = (1-\alpha)\left[\sum_{i\in\mathbb{N}} \lambda^R_{nit|n} \mathbb{E}_{nit}\left[b\right] w_{it}\right] R_{nt} + \frac{\beta^H}{\beta^L}\left[\sum_{i\in\mathbb{N}} \lambda^L_{int|n} \mathbb{E}_{int}\left[b\right]\right] w_{nt} L_{nt}, \tag{E.23}$$

where wages are now adjusted by average worker ability, as determined in equation (E.19).

F Dynamic Model

In this section of the online appendix, we develop a dynamic model, in which adjustment costs for investments in durable building capital introduce gradual adjustment in response to changes in the transport network. We introduce these investments in durable building capital following the standard approach to intertemporal saving and investment decisions in Obstfeld and Rogoff (1996). If these adjustment costs for investments in durable building capital are the sole source of dynamics, we show that worker commuting decisions conditional on the price of floor space and wages remain the same as those in our static model in the paper. The reason is that the price of floor space and wages fully summarize the impact of these investments in durable building capital on worker commuting decisions. An important implication is that our baseline quantitative analysis remains unchanged relative to our static model, because it conditions on the observed rateable values and employment by residence in the data, which capture these dynamics in the price of floor space and wages. Only when we undertake counterfactuals for changes in the transport network, does the introduction of adjustment costs for investments in durable building capital affect the quantitative predictions of our approach, because we no longer condition on the observed rateable values and employment by residence in the data.

We consider a city (Greater London) that is embedded within a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations ($\mathbb{M}$). Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and is indexed by t. The economy consists of two types of infinitely-lived agents: workers and landlords. Workers are assumed not to have access to a saving or borrowing technology and are modeled as "hand to mouth," as in Kaplan and Violante (2014). The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws. With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and workplace equal the realized values.[6] Motivated by our empirical finding that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$.

Landlords are geographically immobile and own the stock of building capital (O_{nt}) and land (K_n) in the location in which they live. Building capital (O_{nt}) and land (K_n) are used to produce floor space (H_{nt}). Whereas land is in fixed supply, landlords can invest in accumulating building capital over time. At the beginning of each period t, the representative landlord in each location n inherits a durable stock of building capital in that location. She chooses each period how much to invest in building capital (I_{nt}) to augment this durable stock and how much to save or borrow using

[6] To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

a riskless storage technology (bonds). Once building capital has been accumulated in a given location, it depreciates at a constant rate ψ, such that in the absence of any investment, the durable stock of building capital in a given location gradually depreciates away. We assume that investment in building capital is subject to adjustment costs, which imply sluggish adjustment in the stock of building capital over time. We also assume that the investment technology for building capital is putty-clay, in the sense that once investments have been made to create durable building capital in a given location, this building capital cannot be used elsewhere. Therefore, investment in building capital net of depreciation in each location must be non-negative.

We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

F1 Workers

The intertemporal preferences of worker ω living in residence n and working in workplace i at time t are additive separable and isoelastic:

$$\mathbb{U}_{nit}^{R}(\omega)=\sum_{s=t}^{\infty}\rho^{s-t}\frac{U_{nit}^{R}(\omega)^{1-\frac{1}{\zeta}}}{1-\frac{1}{\zeta}},\qquad \zeta>0, \tag{F.1}$$

where ρ is the subjective rate of time discount; ζ is the intertemporal elasticity of substitution; and $U_{nit}^{R}(\omega)$ is the worker's instantaneous utility function.

The instantaneous utility for worker ω residing in location n and working in location i is assumed to depend on consumption of a single tradeable homogeneous final good ($C_{nit}^{R}(\omega)$), consumption of residential floor space ($H_{nit}^{R}(\omega)$), commuting costs (κ_{ni}), common amenities for all workers (B_{ni}), and an idiosyncratic amenity draw ($b_{nit}(\omega)$) for each individual worker, according to the following Cobb-Douglas functional form:

$$U_{nit}^{R}(\omega)=\frac{B_{nit}b_{nit}(\omega)}{\kappa_{nit}}\left(\frac{C_{nit}^{R}(\omega)}{\alpha}\right)^{\alpha}\left(\frac{H_{nit}^{R}(\omega)}{1-\alpha}\right)^{1-\alpha},\qquad 0<\alpha<1. \tag{F.2}$$

Throughout this section, we assume for simplicity that common amenities (B_{nit}) are exogenous, although the analysis here can be extended to make these amenities endogenous to the surrounding concentration of economic activity through the introduction of agglomeration forces, as in the paper.

Idiosyncratic amenities ($b_{nit}(\omega)$) are assumed to be drawn from an independent extreme value (Fréchet) distribution each period for each residence-workplace pair, as in the paper:

$$G_{nit}(b)=e^{-b^{-\epsilon}},\qquad \epsilon>1, \tag{F.3}$$

where we normalize the Fréchet scale parameter in equation (F.3) to one, because it enters worker choice probabilities isomorphically to common amenities B_{ni} from equation (F.2); the Fréchet shape parameter ϵ regulates the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller the shape parameter ϵ, the greater the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables.

Without access to a savings or borrowing technology, workers choose their residence, workplace, consumption of the final and residential floor space use each period to maximize their instantaneous utility (F.2). From the first-order conditions to this instantaneous utility maximization problem, each worker allocates a constant share of expenditure to the final good and residential floor space:

$$C_{nit}^{R}(\omega) = \alpha \frac{w_{it}}{P_{nt}}, \tag{F.4}$$

$$H_{nit}^{R}(\omega) = (1-\alpha) \frac{w_{it}}{Q_{nt}}, \tag{F.5}$$

where P_{nt} is the price of the final good and Q_{nt} is the price of floor space. Throughout this section, we assume for simplicity that there are no frictions to the allocation of floor space between residential and commercial use. Therefore, no arbitrage ensures that there is a single price of floor space for both residential and commercial use (Q_{nt}). We assume that the final good can be costlessly traded throughout the economy and choose it as our numeraire, such that:

$$P_{nt} = P_t = 1, \qquad \forall n. \tag{F.6}$$

Substituting for equilibrium consumption and residential floor space use in equation (F.2), the instantaneous indirect utility function takes the same form as in equation (3) in the paper:

$$U_{nit}^{R}(\omega) = \frac{B_{nit} b_{nit}(\omega) w_{it}}{\kappa_{nit} Q_{nt}^{1-\alpha}}, \qquad 0 < \alpha < 1, \tag{F.7}$$

where we have used our choice of numeraire ($P_{nt} = 1$).

The instantaneous utility function (F.7) and the specification of instantaneous amenities (F.3) take the same form as in the paper. Therefore, using the fact that workers have no access to a saving or borrowing technology and maximize instantaneous utility, their choices of workplace and residence also take exactly the same form as in the paper. Conditional on choosing a residence-workplace pair in Greater London, the probability that a worker chooses to live in location n and work in location i is given by:

$$\lambda_{nit} = \frac{L_{nit}/L_{\mathbb{M}t}}{L_{\mathbb{N}t}/L_{\mathbb{M}t}} = \frac{L_{nit}}{L_{\mathbb{N}t}} = \frac{(B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}, \qquad n, i \in \mathbb{N}, \tag{F.8}$$

where we have used our choice of numeraire ($P_{nt} = 1$). Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker lives in each residence $n \in \mathbb{N}$ ($\lambda_{it}^{R} = R_{nt}/L_{\mathbb{N}t}$), conditional on choosing a residence-workplace pair in Greater London:

$$\lambda_{nt}^{R} = \frac{\sum_{i \in \mathbb{N}} (B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}. \tag{F.9}$$

Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker lives in each workplace $i \in \mathbb{N}$ ($\lambda_{nt}^{L} = L_{it}/L_{\mathbb{N}t}$), conditional on choosing a residence-workplace pair in Greater London:

$$\lambda_{it}^{L} = \frac{\sum_{n \in \mathbb{N}} (B_{nit} w_{it})^{\epsilon} \left(\kappa_{nit} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell t} w_{\ell t})^{\epsilon} \left(\kappa_{k\ell t} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}. \tag{F.10}$$

Using the commuting probability (λ_{nit}) from equation (F.8) and the residence probability (λ_{nt}^{R}) from equation (F.9), the conditional probability that a worker commutes to location i conditional on residing in location n is:

$$\lambda_{nit|n}^{R} = \frac{\lambda_{nit}}{\lambda_{nt}^{R}} = \frac{(B_{nit} w_{it}/\kappa_{nit})^{\epsilon}}{\sum_{\ell \in \mathbb{N}} (B_{n\ell t} w_{\ell t}/\kappa_{n\ell t})^{\epsilon}}. \tag{F.11}$$

Using this commuting probability conditional on residence from equation (F.11), per capita income by residence in location n is:

$$v_{nt} = \sum_{i \in \mathbb{N}} \lambda_{nit|n}^{R} w_{it}. \tag{F.12}$$

Commuter market clearing requires that the measure of employment in each location (L_{it}) equals the measure of workers choosing to commute to that location. Using the commuting probabilities conditional on residence from equation (F.11), this commuter market clearing condition can be written as:

$$L_{it} = \sum_{n \in \mathbb{N}} \lambda^{R}_{nit|n} R_{nt}. \tag{F.13}$$

Finally, using the Fréchet distribution for idiosyncratic amenities (F.3), expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is equalized across all residence-workplace pairs in the economy:

$$\bar{U} = \vartheta \left[\sum_{k \in \mathbb{M}} \sum_{\ell \in \mathbb{M}} \left(B_{k\ell t} w_{\ell t} \right)^{\epsilon} \left(\kappa_{k\ell t} Q_{kt}^{1-\alpha} \right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \tag{F.14}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$; and $\Gamma(\cdot)$ is the Gamma function. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}t}/L_{\mathbb{M}t}$), we can re-write this population mobility condition as:

$$\bar{U} \left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}} \right)^{\frac{1}{\epsilon}} = \vartheta \left[\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell t} w_{\ell t} \right)^{\epsilon} \left(\kappa_{k\ell t} Q_{kt}^{1-\alpha} \right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \tag{F.15}$$

where only the limits of the summation signs differ on the right-hand sides of equations (F.14) and (F.15).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

While workplace and residence choices in equations (F.8)-(F.15) take exactly the same form as in the paper conditional on wages and the price of floor space (using our choice of numeraire $P_{nt} = 1$), the key difference in this dynamic model is that equilibrium wages (w_{nt}) and the price of floor space (Q_{nt}) will be influenced by sluggish adjustment in the stock of durable floor space, as shown below.

F2 Landlords

The intertemporal preferences of the representative landlord in location n at time t are assumed to take the same additive separable and isoelastic form:

$$\mathbb{U}^{H}_{nt} = \sum_{s=t}^{\infty} \rho^{s-t} \frac{\left(U^{H}_{nt}\right)^{1-\frac{1}{\zeta}}}{1-\frac{1}{\zeta}}, \qquad \zeta > 0, \tag{F.16}$$

where U^{H}_{nt} is the representative landlord's instantaneous utility function. We assume for simplicity that instantaneous utility for the representative landlord depends solely on consumption of the tradeable homogeneous final good (C^{H}_{nt}):[7]

$$U^{H}_{nt} = C^{H}_{nt}. \tag{F.17}$$

The representative landlord in each location has access to a storage technology (bonds) and an investment technology for augmenting the durable stock of building capital in that location. The storage technology can be used to borrow and save in terms of the final good at an exogenous rate of return r that is determined in the wider economy. The

[7] Allowing landlords to consume residential floor space as well as the tradeable final good slightly complicates some expressions, but leaves the analysis largely unchanged.

investment technology is assumed to take the following form: if I_{nt} units of the final good are invested in building capital in location n in period t, the durable stock of building capital in period $t+1$ (O_{nt+1}) is given by:

$$O_{nt+1} = I_{nt} + (1-\psi)O_{nt}, \tag{F.18}$$

where ψ is the rate of depreciation. We assume that this investment in building capital is subject to quadratic adjustment costs, such that achieving an investment I_{nt} incurs the following cost in terms of the final good:

$$\frac{\chi}{2}\frac{I_{nt}^2}{O_{nt}}, \tag{F.19}$$

where χ parameterizes the magnitude of these adjustment costs.

Denoting the stock of bonds held by the representative landlord in location n in period t by Z_{nt}, the resulting period-by-period budget constraint is:

$$J_{nt}O_{nt} + G_{nt}K_n + (1+r)Z_{nt} = C_{nt}^H + I_{nt} + \frac{\chi}{2}\frac{I_{nt}^2}{O_{nt}} + Z_{nt+1}, \tag{F.20}$$

where J_{nt} is the price of building capital and G_{nt} is the price of land. Therefore, the first term on the left-hand side ($J_{nt}O_{nt}$) captures the representative household's income from ownership of building capital in period t; the second term on the left-hand side ($G_{nt}K_n$) corresponds to income from ownership of land in period t; the third term on the left-hand side ($(1+r)Z_{nt}$) is the value of the representative landlord's stock of bonds inclusive of the rate of return on these bonds in period t; the first term on the right-hand side (C_{nt}^H) is the representative landlord's consumption of the final good in period t; the second term on the right-hand side (I_{nt}) is investment in building capital in period t; the third term on the right-hand side ($(\chi/2)(I_{nt}^2/O_{nt})$) is the adjustment cost incurred in achieving this level of investment in period t; and the final term on the right-hand side (Z_{nt+1}) is the value of the representative landlord's stock of bonds carried forward into period $t+1$.

Using this period-by-period budget constraint (F.20), we can write the representative landlord's intertemporal budget constraint as:

$$\sum_{s=t}^{\infty}\left(\frac{1}{1+r}\right)^{s-t}(J_{ns}O_{ns} + G_{ns}K_n) + (1+r)\,Z_{nt} = \sum_{s=t}^{\infty}\left(\frac{1}{1+r}\right)^{s-t}\left(C_{ns}^H + I_{ns} + \frac{\chi}{2}\frac{I_{nt}^2}{O_{nt}}\right), \tag{F.21}$$

where we assume that the following transversality condition holds:

$$\lim_{T\to\infty}\left(\frac{1}{1+r}\right)^T Z_{t+T+1} = 0. \tag{F.22}$$

F3 Production

The homogeneous final good is produced under conditions of perfect competition and constant returns to scale. For simplicity, we abstract from the use of machinery capital as a factor of production in this section. We assume that output of the final good (Y_{it}) is produced using inputs of labor (L_{it}) and commercial floor space (H_{it}^L) according to the following Cobb-Douglas technology:

$$Y_{it} = A_{it}\left(\frac{L_{it}}{\beta^L}\right)^{\beta^L}\left(\frac{H_{it}^L}{\beta^H}\right)^{\beta^H}, \qquad 0 < \beta^L, \beta^H < 1, \qquad \beta^L + \beta^H = 1, \tag{F.23}$$

where A_{it} denotes final goods productivity. Throughout this section, we also assume for simplicity that productivity (A_{it}) is exogenous, although the analysis here can be extended to make it endogenous to the surrounding concentration of economic activity through the introduction of agglomeration forces, as in the paper.

From profit maximization and zero profits, payments for labor and commercial floor space are constant shares of revenue ($X_{it} = P_{it}Y_{it} = Y_{it}$) in each location with positive production of the final good:

$$w_{it}L_{it} = \beta^L X_{it}, \tag{F.24}$$

$$Q_{it}H_{it}^L = \beta^H X_{it}. \tag{F.25}$$

Recall that we assume for simplicity in this section that there are no frictions to the allocation of floor space between residential and commercial use. Therefore, no arbitrage ensures that there is a single price of floor space for both residential and commercial use (Q_{nt}).

Combining equations (F.24) and (F.25), total payments for commercial floor space are a constant multiple of total labor income by workplace, which equals the wage (w_{nt}) times employment (L_{nt}):

$$Q_{it}H_{it}^L = \frac{\beta^H}{\beta^L} w_{it}L_{it}. \tag{F.26}$$

Using equations (F.24) and (F.25) to substitute for employment (L_{it}) and commercial floor space use (H_{it}^L) in the production technology, we obtain the requirement that the price of the final good ($P_{it} = 1$) must equal its unit cost in order for zero profits to be made and the final good to be produced:

$$1 = \frac{1}{A_{it}} w_{it}^{\beta^L} Q_{it}^{\beta^H}. \tag{F.27}$$

Re-arranging this relationship yields the following expression for wages as a function of the price of floor space:

$$w_{it} = A_{it}^{\frac{1}{\beta^L}} Q_{it}^{-\frac{\beta^H}{\beta^L}}. \tag{F.28}$$

Intuitively, the higher productivity (A_{it}) and the lower the price of floor space (Q_{it}), the higher the wages (w_{it}) that can be paid while maintaining zero profits in equilibrium.

F4 Construction Sector

Floor space (H_{nt}) is produced by a competitive construction sector using the stock of durable capital (O_{nt}) and land (K_n) according to a constant returns to scale technology. We assume for simplicity that this construction technology takes the following Cobb-Douglas form:

$$H_{nt} = \left(\frac{O_{nt}}{\mu}\right)^{\mu} \left(\frac{K_n}{1-\mu}\right)^{1-\mu}, \qquad 0 < \mu < 1. \tag{F.29}$$

Perfect competition and zero profits in construction imply that the price of floor space (Q_{nt}) is equal to its unit cost of production, as determined by the price of building capital (J_{nt}) and the price of land (G_{nt}):

$$Q_{nt} = J_{nt}^{\mu} G_{nt}^{1-\mu}. \tag{F.30}$$

Perfect competition and zero profits also imply that payments for building capital and land are constant shares of payments for floor space:

$$J_{nt}O_{nt} = \mu Q_{nt}H_{nt}, \tag{F.31}$$

$$G_{nt}K_n = (1-\mu)\, Q_{nt}H_{nt}, \tag{F.32}$$

where payments for floor space include those for both residential and commercial use.

F5 Building Capital Market Clearing

Building capital market clearing implies that the income received by landlords from ownership of building capital equals the sum of payments for the use of building capital for residential and commercial floor space:

$$J_{nt}O_{nt} = \mu\left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda^{R}_{nit|n}w_{it}\right)R_{nt} + \frac{\beta^{H}}{\beta^{L}}w_{nt}L_{nt}\right]. \tag{F.33}$$

The first term inside the square parentheses on the right-hand side captures payments for the use of building capital for residential floor space, which equals a constant share of total income by residence, where total income by residence equals the number of residents (R_{nt}) times average per capita income (v_{nt}, as determined by the commuting probabilities conditional on residence in equation (F.12)). The second term inside these square parentheses captures payments for the use of building capital for commercial floor space, which equals a constant multiple of total income by workplace from equation (F.26), where total income by workplace equals employment (L_{nt}) times wages (w_{nt}).

F6 Land Market Clearing

Similarly, land market clearing requires that the income received by landlords from ownership of land equals the sum of payments for the use of land for residential and commercial floor space:

$$G_{nt}K_{n} = (1-\mu)\left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda^{R}_{nit|n}w_{it}\right)R_{nt} + \frac{\beta^{H}}{\beta^{L}}w_{nt}L_{nt}\right]. \tag{F.34}$$

Following the same logic as for building capital above, the first term inside the square parentheses on the right-hand side captures payments for the use of land for residential floor space, which equals a constant share of total income by residence. The second term inside these square parentheses captures payments for the use of land for commercial floor space, which equals a constant multiple of total income by workplace.

F7 Floor Space Market Clearing

Summing equations (F.33) and (F.34), we obtain the same combined land and commuter market clearing condition as in equation (16) in our static model in the paper:

$$Q_{nt}H_{nt} = \left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda^{R}_{nit|n}w_{it}\right)R_{nt} + \frac{\beta^{H}}{\beta^{L}}w_{nt}L_{nt}\right]. \tag{F.35}$$

Analogous to the market clearing conditions for building capital and land above, the first and second terms inside the square parentheses on the right-hand side capture payments for the use of residential and commercial floor space respectively, which are constant multiples of income by residence and income by workplace respectively.

F8 Saving and Investment

We now characterize the optimal savings and investment decisions of the representative landlord in each location n in each time period t. In an interior equilibrium with positive net investment ($I_{nt} = O_{nt+1} - (1-\psi)O_{nt} > 0$), we can re-write the period-by-period budget constraint (F.20) as:

$$C^{H}_{nt} = (1+r)Z_{nt} - Z_{nt+1} + (J_{nt}O_{nt} + G_{nt}K_{n}) - (O_{nt+1} - (1-\psi)O_{nt}) - \frac{\chi}{2}\frac{(O_{nt+1} - (1-\psi)O_{nt})^{2}}{O_{nt}}. \tag{F.36}$$

Substituting this expression for consumption (F.36) into the representative landlord's intertemporal utility function (F.16), we obtain:

$$\mathbb{U}_{nt}^{H}=\sum_{s=t}^{\infty}\frac{\rho^{s-t}\left[\begin{array}{c}(1+r)Z_{ns}-Z_{ns+1}+(J_{ns}O_{ns}+G_{ns}K_{n})\\ -\left(O_{ns+1}-(1-\psi)O_{ns}\right)-\frac{\chi}{2}\frac{(O_{ns+1}-(1-\psi)O_{ns})^{2}}{O_{ns}}\end{array}\right]^{1-\frac{1}{\zeta}}}{1-\frac{1}{\zeta}}. \tag{F.37}$$

The first-order condition with respect to Z_{ns+1} is:

$$\rho^{s+1-t}(1+r)\left(C_{ns+1}^{H}\right)^{-1/\zeta}-\rho^{s-t}\left(C_{ns}^{H}\right)^{-1/\zeta}=0, \tag{F.38}$$

which can be re-written as the following Euler equation:

$$C_{ns+1}^{H}=\rho^{\zeta}\left(1+r\right)^{\zeta}C_{ns}^{H}. \tag{F.39}$$

We thus obtain the conventional result that the consumption profile of the representative landlord is constant over time if the rate of return to bonds equals the discount rate ($\rho=1/(1+r)$); rises over time if the rate of return on bonds exceeds the discount rate ($\rho>1/(1+r)$; and falls over time if the rate of return on bonds is less than the discount rate ($\rho<1/(1+r)$. The corresponding first-order condition with respect to O_{ns+1} is:

$$\begin{aligned}\rho^{s+1-t}\left(C_{ns+1}^{H}\right)^{-1/\zeta}\left[J_{ns+1}+(1-\psi)-\frac{\chi}{2}\frac{\partial}{\partial O_{ns+1}}\left(\frac{\left(O_{ns+2}-(1-\psi)O_{ns+1}\right)^{2}}{O_{ns+1}}\right)\right]\\ -\rho^{s-t}\left(C_{ns}^{H}\right)^{-1/\zeta}\left[1+\frac{\chi}{2}\frac{\partial}{\partial O_{ns+1}}\left(\frac{\left(O_{ns+1}-(1-\psi)O_{ns}\right)^{2}}{O_{ns}}\right)\right]=0,\end{aligned} \tag{F.40}$$

where we have:

$$\frac{\partial}{\partial O_{ns+1}}\left(\frac{\left(O_{ns+2}-(1-\psi)O_{ns+1}\right)^{2}}{O_{ns+1}}\right)=-2(1-\psi)\frac{I_{ns+1}}{O_{ns+1}}-\left(\frac{I_{ns+1}}{O_{ns+1}}\right)^{2}, \tag{F.41}$$

$$\frac{\partial}{\partial O_{ns+1}}\left(\frac{\left(O_{ns+1}-(1-\psi)O_{ns}\right)^{2}}{O_{ns}}\right)=\frac{2I_{ns}}{O_{ns}}. \tag{F.42}$$

Using these relationships, the first-order condition with respect to O_{ns+1} (F.40) can be re-written as:

$$\begin{aligned}\rho^{s+1-t}\left(C_{ns+1}^{H}\right)^{-1/\zeta}\left[J_{ns+1}+(1-\psi)+\chi(1-\psi)\frac{I_{ns+1}}{O_{ns+1}}+\frac{\chi}{2}\left(\frac{I_{ns+1}}{O_{ns+1}}\right)^{2}\right]\\ -\rho^{s-t}\left(C_{ns}^{H}\right)^{-1/\zeta}\left[1+\chi\frac{I_{ns}}{O_{ns}}\right]=0.\end{aligned} \tag{F.43}$$

Using the Euler equation (F.39), we can rewrite this first-order condition with respect to O_{ns+1} as the following no-arbitrage condition that determines the investment rate (I_{ns}/O_{ns}) as a function of the price of building capital (J_{ns+1}) relative to rate of return on bonds (r) and the depreciation rate (ψ):

$$\left[J_{ns+1}+(1-\psi)+\chi(1-\psi)\frac{I_{ns+1}}{O_{ns+1}}+\frac{\chi}{2}\left(\frac{I_{ns+1}}{O_{ns+1}}\right)^{2}\right]=(1+r)\left[1+\chi\frac{I_{ns}}{O_{ns}}\right],\qquad I_{ns}>0. \tag{F.44}$$

From these first-order conditions, the representative landlord chooses the overall level of investments in bonds and building capital to achieve consumption smoothing, as captured by the Euler equation (F.39). In an interior equilibrium with positive net investment, the representative landlord chooses the investment rate for building capital to equate its rate of return net of depreciation and adjustment costs with the rate of return on bonds, as captured by the no-arbitrage condition (F.44), Finally, if the rate of return on investments in building capital is less than the rate of return on bonds for all positive values of net investment in building capital ($J_{ns+1}<r+\psi$), we have a corner solution, in which the representative landlord sets investment in building capital equal to zero ($I_{ns}=0$).

F9 Steady-State Equilibrium

We now characterize the model's steady-state equilibrium, before examining its transition dynamics in response to a shock (such as a transport improvement) in the next section. The steady-state equilibrium is referenced by a constant vector of endogenous variables in each location (λ_n^{R*}, λ_n^{L*}, w_n^*, J_n^*, G_n^*, O_n^*, I_n^*) and a constant aggregate city population ($L_{\mathbb{N}}^*$). All other endogenous variable of the model can be written in terms of this equilibrium vector and scalar. We focus on an interior steady-state equilibrium with a positive steady-state investment in building capital ($I_n^* > 0$). In such an interior steady-state equilibrium, a constant stock of durable building capital in each location requires that the investment rate exactly offsets depreciation in each location:

$$\frac{I_n^*}{O_n^*} = \psi. \tag{F.45}$$

Using this constant investment rate in the no-arbitrage condition between the rate of return to bonds and investment in building capital, the steady-state price of durable building capital is the same across all locations:

$$J_n^* = J^* = r\left(1 + \chi\psi\right) + \psi\left(1 + \frac{\chi}{2}\psi\right), \qquad \text{for all} \qquad n. \tag{F.46}$$

Intuitively, no-arbitrage across alternative forms of investment ensures that the common steady-state price of building capital across all locations (J^*) is pinned down by the rate of return on bonds (r), the depreciation rate for building capital (ψ) and adjustment costs (χ).

Using this common steady-state price of building capital (F.46) in the zero-profit condition for the construction sector (F.30), we find that the steady-state price of floor space (Q_n^*) only varies across locations because of variation in the steady-state price of land (G_n^*):

$$Q_n^* = \left(J^*\right)^{\mu}\left(G_n^*\right)^{1-\mu}. \tag{F.47}$$

Using this solution for the steady-state price of floor space (F.47) in the zero profit condition for production (F.28), we find that the steady-state wage (w_n^*) in each location with positive production is determined by location productivity (A_n), the steady-state price of land in that location (G_n^*), and the common steady-state price of building capital across all locations (J^*):

$$w_n^* = A_n^{\frac{1}{\beta^L}}\left(J^*\right)^{-\mu\frac{\beta^H}{\beta^L}}\left(G_n^*\right)^{-(1-\mu)\frac{\beta^H}{\beta^L}}. \tag{F.48}$$

Therefore, steady-state differences in wages across locations are driven by differences in productivity and variation in the steady-state price of land across locations.

The steady-state price of land in each location (G_n^*) is determined by the land market clearing condition (F.34) and depends upon the steady-state values of income by workplace and residence:

$$G_n^* = \frac{(1-\mu)\left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda_{nit|n}^{R*}w_i^*\right)\lambda_n^{R*} + \frac{\beta^H}{\beta^L}w_n^*\lambda_n^{L*}\right]L_{\mathbb{N}}^*}{K_n}, \tag{F.49}$$

where $L_{\mathbb{N}}^*$ is the steady-state aggregate city population.

The steady-state durable stock of building capital in each location (O_n^*) is determined by the capital market clearing condition (F.33) using the steady-state price of building capital (F.46), and also depends on the steady-state values of income by workplace and residence:

$$O_n^* = \frac{\mu\left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda_{ni|n}^{R*}w_i^*\right)\lambda_n^{R*} + \frac{\beta^H}{\beta^L}w_n^*\lambda_n^{L*}\right]L_{\mathbb{N}}^*}{J^*}. \tag{F.50}$$

The steady-state residence and workplace probabilities (λ_n^{R*} and λ_n^{L*} respectively) can be determined from equations (F.9) and (F.10) using these steady-state values of building capital prices (F.46), wages (F.48), and land (F.49):

$$\lambda_n^{R*} = \left[\frac{\sum_{\ell \in \mathbb{N}} (B_{n\ell} w_\ell^*)^\epsilon \left(\kappa_{n\ell} (G_n^*)^{-(1-\mu)(1-\alpha)} \right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell} w_\ell^*)^\epsilon \left(\kappa_{k\ell} (G_k^*)^{-(1-\mu)(1-\alpha)} \right)^{-\epsilon}} \right], \tag{F.51}$$

$$\lambda_n^{L*} = \left[\frac{\sum_{k \in \mathbb{N}} (B_{kn} w_n^*)^\epsilon \left(\kappa_{kn} (G_k^*)^{-(1-\mu)(1-\alpha)} \right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell} w_\ell^*)^\epsilon \left(\kappa_{k\ell} (G_k^*)^{-(1-\mu)(1-\alpha)} \right)^{-\epsilon}} \right]. \tag{F.52}$$

Similarly, the steady-state commuting probabilities conditional on residence ($\lambda_{ni|n}^{R*}$) can be obtained from equation (F.11) using steady-state wages (F.48) and are given by:

$$\lambda_{ni|n}^{R*} = \frac{\lambda_{ni}^*}{\lambda_n^{R*}} = \frac{(B_{ni} w_i^* / \kappa_{ni})^\epsilon}{\sum_{\ell \in \mathbb{N}} (B_{n\ell} w_\ell^* / \kappa_{n\ell})^\epsilon}. \tag{F.53}$$

Using these steady-state commuting probabilities conditional on residence ($\lambda_{ni|n}^{R*}$), steady-state per capita income by residence in location n is:

$$v_n^* = \sum_{i \in \mathbb{N}} \lambda_{ni|n}^{R*} w_i^*. \tag{F.54}$$

Finally, the steady-state aggregate city population ($L_\mathbb{N}^*$) is determined by the population mobility condition that expected utility for each bilateral commute is equal to the value in the wider economy ($\bar{U}$):

$$\bar{U} \left(\frac{L_\mathbb{N}^*}{L_\mathbb{M}} \right)^{\frac{1}{\epsilon}} = \vartheta \left[\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} (B_{k\ell} w_\ell^*)^\epsilon \left(\kappa_{k\ell} (J^*)^{\mu(1-\alpha)} (G_k^*)^{(1-\mu)(1-\alpha)} \right)^{-\epsilon} \right]^{\frac{1}{\epsilon}}. \tag{F.55}$$

The steady-state equilibrium vector of six variables in each location (R_n^*, L_n^*, w_n^*, G_n^*, O_n^*, I_n^*) and the two scalars (J^*, $L_\mathbb{N}^*$) are uniquely determined by the system of eight equations corresponding to: (i) constant aggregate stock of building capital (F.45); (ii) no-arbitrage between alternative forms of investment (F.46); (iii) zero profits in production (F.48); (iv) land market clearing (F.49); (v) capital market clearing (F.50); (vi) the residence choice probabilities (F.51); (vii) the workplace choice probabilities (F.52); and (viii) population mobility (F.55).

Proposition F.1 *Assume exogenous, finite and strictly positive location characteristics ($A_n \in (0,\infty)$, $B_{ni} \in (0,\infty) \times (0,\infty)$, $\kappa_{ni} \in (0,\infty) \times (0,\infty)$, $K_n \in (0,\infty)$). Under these assumptions, there exists a unique steady-state equilibrium vector of six variables in each location (R_n^*, L_n^*, w_n^*, G_n^*, O_n^*, I_n^*) and the two scalars (J^*, $L_\mathbb{N}^*$), given expected utility ($\bar{U}$) and total population ($L_\mathbb{M}$) in the wider economy.*

Proof. Assume exogenous, finite and strictly positive location characteristics ($A_n \in (0,\infty)$, $B_{ni} \in (0,\infty) \times (0,\infty)$, $\kappa_{ni} \in (0,\infty) \times (0,\infty)$, $K_n \in (0,\infty)$). Under these assumptions, all locations are incompletely specialized as both workplaces and residences, because the support of the Fréchet distribution for idiosyncratic amenities is unbounded from above. Using the probability of residing in a location (equation (F.51)), the probability of working in a location (equation (F.52)), the zero-profit condition in equation (F.48), and the population mobility condition between the city and the larger economy in equation (F.55), the fraction of workers residing in location n can be written as:

$$\lambda_n^{R*} = \frac{R_n^*}{L_\mathbb{N}^*} = \left(\frac{\vartheta}{\bar{U}} \right)^\epsilon \left(\frac{L_\mathbb{M}}{L_\mathbb{N}^*} \right) \sum_{\ell \in \mathbb{N}} B_{n\ell}^\epsilon A_\ell^{\epsilon/\beta^L} (J^*)^{-\epsilon\mu[\beta^H/\beta^L + (1-\alpha)]} \kappa_{n\ell}^{-\epsilon} (G_\ell^*)^{-\epsilon(1-\mu)\beta^H/\beta^L} (G_n^*)^{-\epsilon(1-\mu)(1-\alpha)},$$

while the fraction of workers employed in location n can be written as:

$$\lambda_n^{L*} = \frac{L_n^*}{L_{\mathbb{N}}^*} = \left(\frac{\vartheta}{\bar{U}}\right)^{\epsilon}\left(\frac{L_{\mathbb{M}}}{L_{\mathbb{N}}^*}\right)\sum_{k\in\mathbb{N}} B_{kn}^{\epsilon} A_n^{\epsilon/\beta^L}\left(J^*\right)^{-\epsilon\mu[\beta^H/\beta^L+(1-\alpha)]}\kappa_{kn}^{-\epsilon}\left(G_n^*\right)^{-\epsilon(1-\mu)\beta^H/\beta^L}\left(G_k^*\right)^{-\epsilon(1-\mu)(1-\alpha)},$$

and expected worker income conditional on residing in block n from equation (F.54) can be written as:

$$v_n^* = \sum_{i\in\mathbb{N}}\frac{B_{ni}^{\epsilon}A_i^{\epsilon/\beta^L}\kappa_{ni}^{-\epsilon}\left(G_i^*\right)^{-\epsilon(1-\mu)\beta^H/\beta^L}}{\sum_{\ell\in\mathbb{N}}B_{n\ell}^{\epsilon}A_\ell^{\epsilon/\beta^L}\kappa_{n\ell}^{-\epsilon}\left(G_\ell^*\right)^{-\epsilon(1-\mu)\beta^H/\beta^L}}\left[A_i^{1/\beta^L}\left(J^*\right)^{-\mu\beta^H/\beta^L}\left(G_i^*\right)^{-(1-\mu)\beta^H/\beta^L}\right],$$

and the land market clearing condition from equation (F.49) can be written as:

$$\frac{\beta^H}{\beta^L}(1-\mu)\frac{w_n^*\lambda_n^{L*}}{G_n^*} + (1-\alpha)(1-\mu)\frac{v_n^*\lambda_n^{R*}}{G_n^*} = \frac{K_n}{L_{\mathbb{N}}^*}.$$

Combining the above relationships, this land market clearing condition can be re-expressed as:

$$D_n(\boldsymbol{G^*}) = \frac{\beta^H}{\beta^L}(1-\mu)\left[\frac{A_n^{1/\beta^L}(J^*)^{-\mu\beta^H/\beta^L}}{(G_n^*)^{1+(1-\mu)\beta^H/\beta^L}}\right]\left[\sum_{k\in\mathbb{N}}\frac{B_{kn}^{\epsilon}A_n^{\epsilon/\beta^L}(J^*)^{-\epsilon\mu\left[\beta^H/\beta^L+(1-\alpha)\right]}\kappa_{kn}^{-\epsilon}}{(G_n^*)^{\epsilon(1-\mu)\beta^H/\beta^L}(G_k^*)^{\epsilon(1-\mu)(1-\alpha)}}\right]$$
$$+\frac{(1-\alpha)(1-\mu)}{G_n^*}\left[\sum_{i\in\mathbb{N}}\left(\frac{B_{ni}^{\epsilon}A_i^{\epsilon/\beta^L}\kappa_{ni}^{-\epsilon}(G_i^*)^{-\epsilon(1-\mu)\beta^H/\beta^L}}{\sum_{\ell\in\mathbb{N}}B_{n\ell}^{\epsilon}A_\ell^{\epsilon/\beta^L}\kappa_{n\ell}^{-\epsilon}(G_\ell^*)^{-\epsilon(1-\mu)\beta^H/\beta^L}}\right)\frac{A_i^{1/\beta^L}(J^*)^{-\mu\beta^H/\beta^L}}{(G_i^*)^{(1-\mu)\beta^H/\beta^L}}\right]\left[\sum_{\ell\in\mathbb{N}}\frac{B_{n\ell}^{\epsilon}A_\ell^{\epsilon/\beta^L}(J^*)^{-\epsilon\mu\left[\beta^H/\beta^L+(1-\alpha)\right]}\kappa_{n\ell}^{-\epsilon}}{(G_\ell^*)^{\epsilon(1-\mu)\beta^H/\beta^L}(G_n^*)^{\epsilon(1-\mu)(1-\alpha)}}\right] = K_n,$$

for all $n \in R$, where we have chosen units in which to measure utility such that $(\bar{U}/\vartheta)^{\epsilon}/L_{\mathbb{M}} = 1$. The above land market clearing condition provides a system of equations for the N boroughs in terms of the N unknown steady-state land prices G_n^*, which has the following properties:

$$\lim_{G_n^*\to 0} D_n(\boldsymbol{G^*}) = \infty > K_n, \qquad \lim_{G_n^*\to\infty} D_n(\boldsymbol{G^*}) = 0 < K_n,$$

$$\frac{dD_n(\boldsymbol{G^*})}{dG_n^*} < 0, \qquad \frac{dD_n(\boldsymbol{G^*})}{dG_i^*} < 0, \qquad \left|\frac{dD_n(\boldsymbol{G^*})}{dG_n^*}\right| > \left|\frac{dD_n(\boldsymbol{G^*})}{dG_i^*}\right|.$$

It follows that there exists a unique vector of steady-state land prices $\boldsymbol{G^*}$ that solves this system of land market clearing conditions. Having solved for steady-state land prices in each location (G_n^*), steady-state wages (w_n^*) follow immediately from the zero-profit condition for production in equation (F.48). Given steady-state land prices (G_n^*) and wages (w^*), the steady-state residence probability (λ_n^{R*}) follows immediately from equation (F.51); the steady-state workplace probability (λ_n^{L*}) follows immediately from equation (F.52); and the steady-state conditional commuting probability conditional on residence ($\lambda_{ni|n}^{R*}$) follows immediately from equation (F.53). Having solved for (λ_n^{R*}, λ_n^{L*}, w_n^*, G_n^*), we recover the steady-state measure of workers residing in the city using the population mobility condition (F.55) and our choice of units in which to measure utility ($(\bar{U}/\vartheta)^{\epsilon}/L_{\mathbb{M}} = 1$):

$$L_{\mathbb{N}}^* = \left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}} B_{k\ell}^{\epsilon}A_\ell^{\epsilon/\beta^L}\left(J^*\right)^{-\epsilon\mu\left[\beta^H/\beta^L+(1-\alpha)\right]}\kappa_{k\ell}^{-\epsilon}\left(G_\ell^*\right)^{-\epsilon(1-\mu)\beta^H/\beta^L}\left(G_k^*\right)^{-\epsilon(1-\mu)(1-\alpha)}\right].$$

Combining the residence and workplace probabilities (λ_n^{R*} and λ_n^{L*} respectively) with total city population ($L_{\mathbb{N}}^*$), we obtain employment by residence ($R_n^* = \lambda_n^{R*}L_{\mathbb{N}}^*$) and employment by workplace ($L_n^* = \lambda_n^{L*}L_{\mathbb{N}}^*$). Using these steady-state solutions (λ_n^{R*}, λ_n^{L*}, w_n^*, $L_{\mathbb{N}}^*$) in the capital market clearing condition (F.50), we obtain the steady-state stock of building capital (O_n^*). From this solution for the steady-state stock of building capital (O_n^*) and the steady-state stability condition (F.45), we obtain the steady-state investment level (I_n^*). Finally, the no-arbitrage condition (F.46) determines the steady-state rate of return on building capital J^*. This completes the characterization of the steady-state equilibrium vector of six variables in each location (R_n^*, L_n^*, w_n^*, G_n^*, O_n^*, I_n^*) and the two scalars (J^*, $L_{\mathbb{N}}^*$). ■

Having determined this unique steady-state equilibrium vector, we can solve for all other endogenous variables of the model. In particular, steady-state per capita income (v_n^*) is:

$$v_n^* = \sum_{i \in \mathbb{N}} \lambda_{ni|n}^{R*} w_i^*. \tag{F.56}$$

Using steady-state residents (R_n^*), employment (L_n^*), wages (w_n^*) and per capita income by residence (v_n^*), we can derive for the steady-state shares of building capital, land and floor space allocated to commercial and residential use. Under our assumption of a Cobb-Douglas construction sector, the steady-state commercial share varies across locations, but is the same for building capital, land and floor space, and is given by:

$$\theta_n^{M*} = \theta_n^{K*} = \theta_n^{H*} = \frac{1}{1 + \frac{1-\alpha}{\beta^H/\beta^L} \frac{v_n^*}{w_n^*} \frac{R_n^*}{L_n^*}}. \tag{F.57}$$

The corresponding residential share equals one minus the commercial shares, and hence also varies across locations, but is the same for building capital, land and floor space:

$$1 - \theta_n^{M*} = 1 - \theta_n^{K*} = 1 - \theta_n^{H*} = \frac{\frac{1-\alpha}{\beta^H/\beta^L} \frac{v_n^*}{w_n^*} \frac{R_n^*}{L_n^*}}{1 + \frac{1-\alpha}{\beta^H/\beta^L} \frac{v_n^*}{w_n^*} \frac{R_n^*}{L_n^*}}. \tag{F.58}$$

F10 Transition Dynamics

We now examine the transition dynamics of the model from period t onwards, given an initial stock of building capital (O_t), both with adjustment costs ($\chi > 0$) and without ($\chi = 0$). In the special case of the model with no adjustment costs ($\chi = 0$), if the initial stock of building capital is less than its steady-state value in all locations ($O_{nt} < O_n^*$ for all n), the city converges immediately to the steady-state value of the stock of building capital in each location (O_n^*). The reason is that landlords can borrow or save in bond markets in the larger economy at the constant rate of return r. Therefore, with no adjustment costs, landlords borrow as much as required to finance investment to increase the stock of building capital to its steady-state value.

However, even in this special case of the model with no adjustment costs ($\chi = 0$), if the initial stock of building capital is greater than its steady-state value in some locations ($O_{nt} > O_n^*$ for some $n = \ell$), the spatial distribution of economic activity in the city exhibits transition dynamics. The reason is our assumption that the stock of building capital is putty-clay, which implies that the investment rate net of depreciation must be non-negative. Therefore, in those locations in which the initial stock of building capital is greater than its steady-state value ($O_{nt} > O_n^*$ for some $n = \ell$), the stock of building capital gradually depreciates towards its steady-state level ($O_{\ell t} = (1 - \psi) O_{\ell t-1}$). This gradual depreciation in some locations induces transition dynamics in the entire spatial equilibrium of economic activity, because this spatial equilibrium depends on the supply of building capital in all locations.

In the presence of adjustment costs ($\chi > 0$), these transition dynamics in general occur whenever the initial stock of building capital differs from its steady-state value in some locations ($O_{nt} \neq O_n^*$ for some $n = \ell$), and we now characterize this general case. Given an initial stock of building capital in each location n in period t (O_{nt}), as determined by investments in period $t - 1$ and earlier, a sequential equilibrium of the model is referenced by a vector of endogenous variables in each location ($R_{ns}, L_{ns}, w_{ns}, J_{ns}, G_{ns}, O_{ns}, I_{ns}$) and a scalar for aggregate city population ($L_{\mathbb{N}s}$) for each period $s \geq t$.

To characterize the model's transition dynamics from the initial conditions in period $t - 1$, we start from a period $T > t$ sufficiently far in the future that the economy is in steady-state, with all endogenous variables equal to their steady-state equilibrium values (R_n^*, L_n^*, w_n^*, J_n^*, G_n^*, O_n^*, I_n^*, $L_{\mathbb{N}}^*$). Given these steady-state equilibrium values in period T, we use the general equilibrium conditions of the model to solve recursively for the endogenous variables in each earlier year $s < T$, until we arrive at the initial conditions in period $t-1$ (R_{nt-1}, L_{nt-1}, w_{nt-1}, J_{nt-1}, G_{nt-1}, O_{nt-1}, I_{nt-1}, $L_{\mathbb{N}t-1}$). In particular, along this transition path, the investment rate for each earlier year $s < T$ must satisfy the no-arbitrage condition (F.44) that equates the rate of return to building investment net of depreciation and adjustment costs with the rate of return on bonds. Therefore, given the price of building capital ($J_{nT} = J^*$), investment in building capital ($I_{nT} = I_n^*$) and the stock of building capital ($O_{nT} = O_n^*$) in period T, we can use this no-arbitrage condition (F.44) and the investment technology (F.18) to solve for investment in building capital (I_{nT-1}) and the stock of building capital (O_{nT-1}) in period $T-1$:

$$\left[J_n^* + (1-\psi) + \chi(1-\psi)\frac{I_n^*}{O_n^*} + \frac{\chi}{2}\left(\frac{I_n^*}{O_n^*}\right)^2\right] = (1+r)\left[1 + \chi\frac{I_{nT-1}}{O_{nT-1}}\right]. \tag{F.59}$$

Given the resulting solution for the stock of building capital (O_{nT-1}) in period $T-1$, we can solve for all other endogenous variables of the model from the remaining static general equilibrium conditions of the model for period $T-1$. The equilibrium price of building capital (J_{nT-1}) is determined by the capital market clearing condition (F.33), which depends on income by residence and workplace, and can be re-written as follows:

$$J_{nT-1} = \frac{\mu\left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda_{niT-1|n}^R w_{iT-1}\right)\lambda_{nT-1}^R + \frac{\beta^H}{\beta^L}w_{nT-1}\lambda_{nT-1}^L\right]L_{\mathbb{N}T-1}}{O_{nT-1}}, \tag{F.60}$$

where $L_{\mathbb{N}T-1}$ is the total city population.

The corresponding equilibrium price of land (G_{nT-1}) in period $T-1$ is determined by the land market clearing condition (F.34), which also depends on income by residence and workplace, and can be re-written as follows:

$$G_{nT-1} = \frac{(1-\mu)\left[(1-\alpha)\left(\sum_{i\in\mathbb{N}}\lambda_{niT-1|n}^R w_{iT-1}\right)\lambda_{nT-1}^R + \frac{\beta^H}{\beta^L}w_{nT-1}\lambda_{nT-1}^L\right]L_{\mathbb{N}T-1}}{K_n}. \tag{F.61}$$

Using these equilibrium prices of building capital (F.60) and land (F.61) in the zero-profit condition for the construction sector (F.30), we can determine the equilibrium price of floor space (Q_{nT-1}) in period $T-1$:

$$Q_{nT-1} = J_{nT-1}^{\mu} G_{nT-1}^{1-\mu}. \tag{F.62}$$

Using this equilibrium price of floor space (F.62) in the zero profit condition for production (F.28), we obtain the equilibrium wage (w_{nT-1}) in period $T-1$:

$$w_{nT-1} = A_{nT-1}^{\frac{1}{\beta^L}} J_{nT-1}^{-\mu\frac{\beta^H}{\beta^L}} G_{nT-1}^{-(1-\mu)\frac{\beta^H}{\beta^L}}. \tag{F.63}$$

The equilibrium residence and workplace probabilities (λ_{nT-1}^R and λ_{nT-1}^L respectively) in period $T-1$ can be determined from equations (F.9) and (F.10) using these equilibrium prices of building capital (F.60), land (F.61) and labor (F.63):

$$\lambda_{nT-1}^R = \left[\frac{\sum_{\ell\in\mathbb{N}}\left(B_{n\ell T-1}w_{\ell T-1}\right)^{\epsilon}\left(\kappa_{n\ell}J_{nT-1}^{-\mu(1-\alpha)}G_{nT-1}^{-(1-\mu)(1-\alpha)}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell T-1}w_{\ell T-1}\right)^{\epsilon}\left(\kappa_{k\ell T-1}J_{kT-1}^{-\mu(1-\alpha)}G_{kT-1}^{-(1-\mu)(1-\alpha)}\right)^{-\epsilon}}\right], \tag{F.64}$$

$$\lambda_{nT-1}^L = \left[\frac{\sum_{r\in\mathbb{N}}\left(B_{rnT-1}w_{nT-1}\right)^{\epsilon}\left(\kappa_{rnT-1}J_{rT-1}^{-\mu(1-\alpha)}G_{rT-1}^{-(1-\mu)(1-\alpha)}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell T-1}w_{\ell T-1}\right)^{\epsilon}\left(\kappa_{k\ell T-1}J_{kT-1}^{-\mu(1-\alpha)}G_{kT-1}^{-(1-\mu)(1-\alpha)}\right)^{-\epsilon}}\right]. \tag{F.65}$$

Similarly, the equilibrium commuting probabilities conditional on residence ($\lambda^R_{nis|n}$) can be obtained from equation (F.11) using equilibrium wages (F.63) and are given by:

$$\lambda^R_{niT-1|n} = \frac{\left(B_{niT-1} w_{iT-1} / \kappa_{niT-1}\right)^{\epsilon}}{\sum_{\ell \in \mathbb{N}} \left(B_{n\ell T-1} w_{\ell T-1} / \kappa_{n\ell T-1}\right)^{\epsilon}}. \tag{F.66}$$

Finally, the equilibrium aggregate city population ($L_{\mathbb{N}T-1}$) is determined by the population mobility condition that expected utility for each bilateral commute equals its value in the wider economy ($\bar{U}$):

$$\bar{U}\left(\frac{L_{\mathbb{N}T-1}}{L_{\mathbb{M}}}\right)^{\frac{1}{\epsilon}} = \vartheta\left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell T-1} w_{\ell T-1}\right)^{\epsilon}\left(\kappa_{k\ell T-1} J_{kT-1}^{\mu(1-\alpha)} G_{kT-1}^{(1-\mu)(1-\alpha)}\right)^{-\epsilon}\right]. \tag{F.67}$$

Given these solutions for the equilibrium values of the endogenous variables of the model in period $T-1$ (R_{nT-1}, L_{nT-1}, w_{nT-1}, J_{nT-1}, G_{nT-1}, O_{nT-1}, I_{nT-1}, $L_{\mathbb{N}T-1}$), we can repeat this recursive procedure to solve for the endogenous variables of the model in period $T-2$ and earlier, until we arrive at the initial conditions in period $t-1$ (R_{nt-1}, L_{nt-1}, w_{nt-1}, J_{nt-1}, G_{nt-1}, O_{nt-1}, I_{nt-1}, $L_{\mathbb{N}t-1}$). We thus obtain a recursive characterization of the entire transition path of the economy from the initial conditions to the new steady-state equilibrium.

F10.1 Comparative Statics for Changes in Commuting Costs

Remarkably, we now show that this dynamic model with adjustment costs for durable investments in building capital yields exactly the same predictions for the impact of the removal of the railway network on workplace employment and commuting as in our baseline quantitative analysis in the paper, once we condition on the observed data in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values. The reason is that durable investments in building capital affect workers' commuting decisions through changes in wages and the price of floor space. By conditioning on the observed data in the initial equilibrium in our baseline year and the observed historical changes in residence employment and rateable values, we control for these changes in wages and the price of floor space, and hence control for the impact of these durable investments on workers' commuting decisions.

First, using equation (F.35), the land market clearing condition for any earlier year $\tau < t$ can be written in terms of observed variables and model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\hat{v}_{nt} v_{nt} \hat{R}_{nt} R_{nt} + \frac{\beta^H}{\beta^L}\hat{w}_{nt} w_{nt} \hat{L}_{nt} L_{nt}, \tag{F.68}$$

where recall that a hat above a variable denotes a relative change, such that $\hat{x}_t = x_\tau / x_t$.

Second, using equations (F.11) and (F.12), expected income by residence (v_{nt}) for any earlier year $\tau < t$ can be written in the same form as in equation (19) in the paper:

$$\hat{v}_{nt} v_{nt} = \sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it} w_{it}, \tag{F.69}$$

where ($\lambda^R_{nit|n}$, w_{it}, v_{nt}) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has cancelled from the numerator and denominator of the fraction on the right-hand side of equation (F.69); and we assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Third, using equations (F.11) and (F.13), workplace employment (L_{it}) for any earlier year $\tau < t$ can be written in the same form as in equation (18) in the paper:

$$\hat{L}_{it} L_{it} = \sum_{n \in \mathbb{N}} \frac{\lambda^R_{nit|n} \hat{w}^{\epsilon}_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^{\epsilon}_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{R}_{nt} R_{nt}, \tag{F.70}$$

where ($\lambda^R_{nit|n}$, L_{it}, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (F.70); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Finally, using equation (F.11), commuting flows ($\hat{L}_{nit}$) for any earlier year $\tau < t$ can be written in the same form as in equation (21) in the paper:

$$\hat{L}_{nit} L_{nit} = \frac{\lambda^R_{nit|n} \hat{w}^{\epsilon}_{it} \hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell \in \mathbb{N}} \lambda^R_{n\ell t|n} \hat{w}^{\epsilon}_{\ell t} \hat{\kappa}^{-\epsilon}_{n\ell t}} \hat{R}_{nt} R_{nt}, \tag{F.71}$$

where (L_{nit}, $\lambda^R_{nit|n}$, R_{nt}, $\hat{R}_{nt}$) are observed or have been solved for; we estimate the change in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$); the change in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt}$) has again cancelled from the numerator and denominator of the fraction on the right-hand side of equation (F.71); and we continue to assume that the workplace and bilateral components of amenities are constant over time ($\hat{\mathcal{B}}^L_{it} = 1$, $\hat{\mathcal{B}}^I_{nit} = 1$).

Note that equations (F.68), (F.69), (F.70) and (F.71) take exactly the same form as equations (17), (19), (18) and (21) in the paper. Therefore, given the same observed variables in the initial equilibrium in our baseline year (L_{nt}, R_{nt}, $\mathbb{Q}_{nt}$, w_{nt}, v_{nt}, L_{nit}), the same observed historical changes in residents and rateable values ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$), and the same estimated changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), this dynamic model predicts the same changes in workplace employment ($\hat{L}_{it}$) and commuting patterns ($\hat{L}_{nit}$) as the static model in the paper.

G Productivity, Amenities and the Supply of Floor Space

In this section of the online appendix, we report the derivations of the results in Section VII. of the paper, in which we choose one urban model from our class in order to recover the supply of floor space, productivity and amenities. In particular, we choose an extension of the canonical urban model to incorporate non-traded goods, as developed in Section D2 of this online appendix. This framework permits a particularly tractable and transparent approach to recovering productivity and amenities and estimating the strength of agglomeration forces. It also allows us to undertake counterfactuals for removing the railway network under a range of alternative assumptions about the floor space supply elasticity and the strength of agglomeration forces.

In Section G1, we separate rateable values into the price and supply of floor space. In Section G2, we recover productivity and amenities for each location. In Section G3, we use these solutions for the supply of the floor space, productivity and amenities to decompose the observed reorganization of economic activity in Greater London into the contributions of changes in commuting costs and changes in each of these other determinants of economic activity. In Section G4, we separate changes in productivity and amenities into the contributions of agglomeration forces and changes in locational fundamentals.

G1 Supply of Floor Space

We separate rateable values into the price and supply of floor space by making the following two additional assumptions. First, we assume no-arbitrage between commercial and residential floor space use, which is consistent with the positive values of residents and workers for all boroughs in our data and the absence of large-scale urban planning in 19th-century London. Therefore, the price of commercial floor space (q_{nt}) is equal to the price of residential floor space (Q_{nt}), such that $q_{nt} = Q_{nt}$, and hence rateable values can be written as:

$$\mathbb{Q}_{nt} = Q_{nt} H^R_{nt} + q_{nt} H^L_{nt} = Q_{nt} H_{nt}. \tag{G.1}$$

Second, we model the supply of floor space (H_{nt}) as depending on geographical land area (K_n) and a constant elasticity function of the price of floor space (Q_{nt}) following Saiz (2010), as in equation (25) in the paper:

$$H_{nt} = h K_n Q^{\mu}_{nt}, \tag{G.2}$$

where h is a constant; $\mu \geq 0$ is the floor space supply elasticity; and $\mu = 0$ corresponds to the special case of a perfectly inelastic supply of floor space.

Substituting this supply function for floor space (G.2) into our expression for rateable values (G.1), we obtain:

$$\mathbb{Q}_{nt} = h K_n Q^{1+\mu}_{nt}. \tag{G.3}$$

Re-arranging this relationship (G.3), we obtain the closed-form expression for the price of floor space (Q_{nt}) in terms of rateable values per unit of land area ($\mathbb{Q}_{nt}/K_n$) from equation (26) in the paper:

$$Q_{nt} = \left(\frac{\mathbb{Q}_{nt}}{h K_n} \right)^{\frac{1}{1+\mu}}. \tag{G.4}$$

Substituting this expression for the price of floor space (G.4) into the floor space supply function (G.2), we obtain the closed-form expression for the supply of floor space (H_{nt}) in terms of rateable values ($\mathbb{Q}_{nt}$) and land area (K_n) from equation (26) in the paper:

$$H_{nt} = h K_n \left(\frac{\mathbb{Q}_{nt}}{h K_n} \right)^{\frac{\mu}{1+\mu}}. \tag{G.5}$$

Dividing the price of floor space for each location n in equation (G.4) by its value for the City of London (indexed by ℓ), and dividing the supply of floor space for each location n in equation (G.5) by its value for the City of London (indexed by ℓ), we obtain the following expressions for the relative price and supply of floor space:

$$\frac{Q_{nt}}{Q_{\ell t}} = \left(\frac{\mathbb{Q}_{nt}/K_n}{\mathbb{Q}_{\ell t}/K_\ell} \right)^{\frac{1}{1+\mu}}, \tag{G.6}$$

$$\frac{H_{nt}}{H_{\ell t}} = \left(\frac{K_n}{K_\ell} \right)^{\frac{1}{1+\mu}} \left(\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{\ell t}} \right)^{\frac{\mu}{1+\mu}}, \tag{G.7}$$

where the constant h has cancelled from both expressions.

Taking relative changes over time in equations (G.4) and (G.5), we also obtain the following expressions for relative changes in the price of floor space ($\hat{Q}_{nt}$) and the supply of floor space ($\hat{H}_{nt}$) as a function of the relative change in rateable values ($\hat{\mathbb{Q}}_{nt}$):

$$\hat{Q}_{nt} = \hat{\mathbb{Q}}^{\frac{1}{1+\mu}}_{nt}, \tag{G.8}$$

$$\hat{H}_{nt} = \hat{\mathbb{Q}}^{\frac{\mu}{1+\mu}}_{nt}, \tag{G.9}$$

where the terms in land area have differenced out, because they are constant over time.

G2 Productivity and Amenities

To recover productivity and amenities, we use these solutions for the price of floor space in an extension of the canonical urban model to incorporate non-traded goods, as developed in Section D2 of this online appendix. Relative to our baseline quantitative analysis in earlier sections of the paper, we now impose additional assumptions on preferences, production technology and market structure. Starting with preferences in equation (3) in the paper, we assume that the consumption goods price index (P_{nt}) is a Cobb-Douglas function of the price of a homogeneous traded good (P_{nt}^{T}) and a homogeneous non-traded good (P_{nt}^{N}), as in equation (27) in the paper:

$$P_{nt} = \left(P_{nt}^{T}\right)^{\nu} \left(P_{nt}^{N}\right)^{1-\nu}, \qquad 0 < \nu < 1. \tag{G.10}$$

Within Greater London, we assume that the homogeneous traded good is costlessly traded, such that

$$P_{nt}^{T} = P_{t}^{T}, \qquad \forall n \in \mathbb{N}. \tag{G.11}$$

Between Greater London and the rest of Great Britain, we allow for changes in trade costs for the traded good, which are reflected in changes in the common price of this good within Greater London (P_{t}^{T}).

Turning now to production technology and market structure, we assume that both traded and non-traded goods are produced with labor, machinery capital and commercial floor space under conditions of perfect competition. We assume for simplicity the same production technology in both sectors. Given our specification of the consumption goods price index in equation (G.10), the marginal utility of consuming the non-traded good converges to infinity as consumption of that good approaches zero. Therefore, in equilibrium, the non-traded good is consumed in all populated locations, which in turn implies that it must be produced in all populated locations. Consistent with the substantial workplace employment observed for all boroughs in our data, we assume that the traded good is also produced in all locations, as can be ensured by the appropriate choice of productivity in the traded sector for each location. Using profit maximization and zero profits, the following two zero-profit conditions must be satisfied for the traded and non-traded sectors in each populated location:

$$P_{t}^{T} = \frac{1}{A_{nt}^{T}} w_{nt}^{\beta^{L}} Q_{nt}^{\beta^{H}} r_{t}^{\beta^{M}}, \tag{G.12}$$

$$P_{nt}^{N} = \frac{1}{A_{nt}^{N}} w_{nt}^{\beta^{L}} Q_{nt}^{\beta^{H}} r_{t}^{\beta^{M}}, \tag{G.13}$$

$$0 < \beta^{L}, \beta^{H}, \beta^{M} < 1, \qquad \beta^{L} + \beta^{H} + \beta^{M} = 1.$$

We first use these additional assumptions on preferences, production technology and market structure to recover a composite measure of traded productivity ($\mathbb{A}_{nt}^{T}$) for each location. Re-arranging the traded zero-profit condition (G.12), we obtain the following expression for composite traded productivity ($\mathbb{A}_{nt}^{T}$), as in equation (28) in the paper:

$$\mathbb{A}_{nt}^{T} = w_{nt}^{\beta^{L}} Q_{nt}^{\beta^{H}}, \qquad \mathbb{A}_{nt}^{T} \equiv P_{t}^{T} A_{nt}^{T} r_{t}^{-\beta^{M}}, \tag{G.14}$$

where we already determined wages (w_{nt}) in our baseline quantitative analysis and the price of floor space (Q_{nt}) in the previous section; this composite traded productivity ($\mathbb{A}_{nt}^{T}$) captures traded productivity (A_{nt}^{T}), the common price of the traded good (P_{t}^{T}), and the common price of machinery (r_t). Taking relative changes over time in equation (G.14), the relative change in composite traded productivity ($\hat{\mathbb{A}}_{nt}^{T}$) for each location is thus:

$$\hat{\mathbb{A}}_{nt}^{T} = \hat{w}_{nt}^{\beta^{L}} \hat{Q}_{nt}^{\beta^{H}}. \tag{G.15}$$

We next use these additional assumptions on preferences, production technology and market structure to recover a composite measure of changes in amenities. We focus on changes in amenities, because our baseline quantitative analysis uses estimates of changes in commuting costs from the construction of the railway network ($\hat{\kappa}_{nit}^{-\epsilon}$) and controls for unobserved determinants of the level of commuting costs in our baseline year using the initial commuting probabilities conditional on residence ($\lambda_{nit|n}^{R}$). Using these estimates of changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), we now show that we can recover changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) from the residence choice probabilities in equation (7) in the paper. We start by substituting the price index for consumption goods (P_{nt}) from equation (G.10) into these residence choice probabilities:

$$\lambda_{nt}^{R} = \frac{\sum_{\ell \in \mathbb{N}} \left(B_{n\ell t} w_{\ell t}\right)^{\epsilon} \left(\kappa_{n\ell t} \left(P_{nt}^{T}\right)^{\alpha\nu} \left(P_{nt}^{N}\right)^{\alpha(1-\nu)} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell t} w_{\ell t}\right)^{\epsilon} \left(\kappa_{k\ell t} \left(P_{kt}^{T}\right)^{\alpha\nu} \left(P_{kt}^{N}\right)^{\alpha(1-\nu)} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}. \tag{G.16}$$

Using our assumption that the homogeneous traded good is costlessly ($P_{nt}^{T} = P_{t}^{T}$) traded from equation (G.11), these residence choice probabilities simplify to:

$$\lambda_{nt}^{R} = \frac{\sum_{\ell \in \mathbb{N}} \left(B_{n\ell t} w_{\ell t}\right)^{\epsilon} \left(\kappa_{n\ell t} \left(P_{nt}^{N}\right)^{\alpha(1-\nu)} Q_{nt}^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k\ell t} w_{\ell t}\right)^{\epsilon} \left(\kappa_{k\ell t} \left(P_{kt}^{N}\right)^{\alpha(1-\nu)} Q_{kt}^{1-\alpha}\right)^{-\epsilon}}. \tag{G.17}$$

Using the zero-profit condition for the non-traded sector (G.13) to substitute for the price of the non-traded good (P_{nt}^{N}), these residence choice probabilities can be re-expressed as:

$$\lambda_{nt}^{R} = \frac{\sum_{\ell \in \mathbb{N}} B_{n\ell t}^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{n\ell t}^{-\epsilon} \left(A_{nt}^{N}\right)^{\epsilon\alpha(1-\nu)} w_{nt}^{-\epsilon\alpha(1-\nu)\beta^{L}} Q_{nt}^{-\epsilon\left[\alpha(1-\nu)\beta^{H}+(1-\alpha)\right]}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} B_{k\ell t}^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{k\ell t}^{-\epsilon} \left(A_{kt}^{N}\right)^{\epsilon\alpha(1-\nu)} w_{kt}^{-\epsilon\alpha(1-\nu)\beta^{L}} Q_{kt}^{-\epsilon\left[\alpha(1-\nu)\beta^{H}+(1-\alpha)\right]}}, \tag{G.18}$$

where the term in the common price of machinery capital r_t has cancelled from the numerator and denominator. Using the zero-profit condition for the traded sector (G.12) to substitute for the wage by residence (w_{nt} in the numerator and w_{kt} in the denominator), these residence choice probabilities can be further re-written as:

$$\lambda_{nt}^{R} = \frac{\sum_{\ell \in \mathbb{N}} B_{n\ell t}^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{n\ell t}^{-\epsilon} \left(A_{nt}^{N}\right)^{\epsilon\alpha(1-\nu)} \left(A_{nt}^{T}\right)^{-\epsilon\alpha(1-\nu)} Q_{nt}^{-\epsilon(1-\alpha)}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} B_{k\ell t}^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{k\ell t}^{-\epsilon} \left(A_{kt}^{N}\right)^{\epsilon\alpha(1-\nu)} \left(A_{kt}^{T}\right)^{-\epsilon\alpha(1-\nu)} Q_{kt}^{-\epsilon(1-\alpha)}}, \tag{G.19}$$

where again the term in the common price of machinery capital r_t has cancelled from the numerator and denominator. Decomposing bilateral amenities (B_{nit}) into a residence component ($\mathcal{B}_{n}^{R}$), a workplace component ($\mathcal{B}_{i}^{L}$) and an idiosyncratic component ($\mathcal{B}_{nit}^{I}$), we can re-write these residence choice probabilities as:

$$\lambda_{nt}^{R} = \frac{\sum_{\ell \in \mathbb{N}} \left(\mathcal{B}_{nt}^{R} \mathcal{B}_{\ell t}^{L} \mathcal{B}_{n\ell t}^{I}\right)^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{n\ell t}^{-\epsilon} \left(A_{nt}^{N}\right)^{\epsilon\alpha(1-\nu)} \left(A_{nt}^{T}\right)^{-\epsilon\alpha(1-\nu)} Q_{nt}^{-\epsilon(1-\alpha)}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(\mathcal{B}_{kt}^{R} \mathcal{B}_{\ell t}^{L} \mathcal{B}_{k\ell t}^{I}\right)^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{k\ell t}^{-\epsilon} \left(A_{kt}^{N}\right)^{\epsilon\alpha(1-\nu)} \left(A_{kt}^{T}\right)^{-\epsilon\alpha(1-\nu)} Q_{kt}^{-\epsilon(1-\alpha)}}. \tag{G.20}$$

Defining a measure of residents' commuting market access (RMA_n) as follows:

$$RMA_{nt} \equiv \left[\sum_{\ell \in \mathbb{N}} \left(\mathcal{B}_{\ell t}^{L} \mathcal{B}_{n\ell t}^{I}\right)^{\epsilon} w_{\ell t}^{\epsilon} \kappa_{n\ell}^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{G.21}$$

and a measure of composite residential amenities as:

$$\mathbb{B}_{nt} \equiv \mathcal{B}_{nt}^{R} \left(A_{nt}^{N}\right)^{\alpha(1-\nu)} \left(A_{nt}^{T}\right)^{-\alpha(1-\nu)}, \tag{G.22}$$

we can further re-write these residence choice probabilities as:

$$\lambda_{nt}^{R} = \frac{\mathbb{B}_{nt}^{\epsilon} Q_{nt}^{-\epsilon(1-\alpha)} RMA_{nt}^{\epsilon}}{\sum_{k \in \mathbb{N}} \mathbb{B}_{kt}^{\epsilon} Q_{kt}^{-\epsilon(1-\alpha)} RMA_{kt}^{\epsilon}}. \tag{G.23}$$

In relative changes, we can write these residence choice probabilities as:

$$\hat{\lambda}^R_{nt}\lambda^R_{nt} = \frac{\lambda^R_{nt}\hat{\mathbb{B}}^{\epsilon}_{nt}\hat{Q}^{-\epsilon(1-\alpha)}_{nt}\widehat{RMA}^{\epsilon}_{nt}}{\sum_{k\in\mathbb{N}}\lambda^R_{kt}\hat{\mathbb{B}}^{\epsilon}_{kt}\hat{Q}^{-\epsilon(1-\alpha)}_{kt}\widehat{RMA}^{\epsilon}_{kt}}, \tag{G.24}$$

which corresponds to equation (29) in the paper, where

$$\widehat{RMA}_{nt} = \left[\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}\right]^{\frac{1}{\epsilon}}, \tag{G.25}$$

and we assume constant workplace and idiosyncratic components of amenities ($\hat{\mathcal{B}}^L_{it} = 1$ and $\hat{\mathcal{B}}^I_{nit} = 1$).

Given the observed choice probabilities in the initial equilibrium in our baseline year (λ^R_{nt}, $\lambda^R_{nit|n}$), our estimated changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), the observed historical changes in choice probabilities ($\hat{\lambda}^R_{nt}$) and our solutions for changes in the price of floor space ($\hat{Q}_{nt}$) and wages ($\hat{w}_{nt}$), we use equation (G.24) to recover unique changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) for each location up to a choice of units in which to measure these amenities.

G3 Model-Based Decompositions

From the quantitative analysis in the previous two sections, we recover relative changes in the supply of floor space, traded productivity and residential amenities ($\hat{H}_{nt}$, $\hat{\mathbb{A}}^T_{nt}$, $\hat{\mathbb{B}}_{nt}$). In this section of the online appendix, we first show that these variables are structural residuals that exactly replicate the observed changes in residence employment ($\hat{R}_{nt}$) and rateable values ($\hat{\mathbb{Q}}_{nt}$) and the changes in wages ($\hat{w}_{nt}$), workplace employment ($\hat{L}_{nt}$) and commuting flows ($\hat{L}_{nit}$) from our baseline quantitative analysis as equilibrium outcomes. We next use this result to undertake model-based decompositions in which we change each of these structural residuals individually or jointly and examine the impact on the spatial distribution of economic activity. While we use these decompositions to assess the relative quantitative magnitude of different mechanisms in the model, it should be kept in mind that these different mechanisms influence one another. In particular, these decompositions treat productivity and amenities as exogenous. However, in the presence of agglomeration forces, productivity and amenities can respond endogenously to the reorganization of economic activity induced by the change in commuting costs. In our counterfactuals in Section VIII. in the paper and Section H of this online appendix, we examine these endogenous responses through agglomeration forces.

We begin by establishing that ($\hat{H}_{nt}$, $\hat{\mathbb{A}}^T_{nt}$, $\hat{\mathbb{B}}_{nt}$) are indeed structural residuals that exactly rationalize the observed data and the solutions from our baseline quantitative analysis. From our solution for the supply of floor space in equation (G.5) and the zero-profit condition in the traded sector (G.14), the changes in composite traded productivity ($\hat{\mathbb{A}}^T_{nt}$) and the supply of floor space ($\hat{H}_{nt}$) ensure that the changes in wages from our baseline quantitative analysis ($\hat{w}_{nt}$) and the observed changes in rateable values ($\hat{\mathbb{Q}}_{nt}$) are consistent with zero-profits in the traded sector:

$$\hat{\mathbb{A}}^T_{nt} = \hat{w}^{\beta^L}_{nt}\hat{Q}^{\beta^H}_{nt} = \hat{w}^{\beta^L}_{nt}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{\beta^H} \tag{G.26}$$

From our solution for the supply of floor space in equation (G.5), the zero-profit condition in the traded sector (G.14), and the residence choice probabilities (G.25), the changes in composite traded productivity ($\hat{\mathbb{A}}^T_{nt}$), the supply of floor space ($\hat{H}_{nt}$) and amenities ($\hat{\mathbb{B}}_{nt}$) ensure that the observed changes in rateable values ($\hat{\mathbb{Q}}_{nt}$) and residence choice probabilities ($\hat{\lambda}^R_{nt}$) are consistent with utility maximization, given our estimated changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$) and the observed choice probabilities in the initial equilibrium in our baseline year (λ_{nit}, λ_{nt}^{R}):

$$\begin{aligned}
\hat{\lambda}_{nt}^{R}\lambda_{nt}^{R} &= \frac{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}\hat{Q}_{nt}^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}}. \\
&= \frac{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}. \\
&= \frac{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t}\hat{\mathbb{B}}_{nt}^{\epsilon}\left(\hat{\mathbb{A}}_{\ell t}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{\ell t}}{\hat{H}_{\ell t}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{n\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\left(\hat{\mathbb{A}}_{\ell t}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{\ell t}}{\hat{H}_{\ell t}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}.
\end{aligned} \tag{G.27}$$

Similarly, these changes in composite traded productivity ($\hat{\mathbb{A}}_{nt}^{T}$), the supply of floor space ($\hat{H}_{nt}$) and amenities ($\hat{\mathbb{B}}_{nt}$) ensure that the observed changes in rateable values ($\hat{\mathbb{Q}}_{nt}$) and the changes in workplace choice probabilities ($\hat{\lambda}_{nt}^{L}$) from our baseline quantitative analysis are consistent with utility maximization, given our estimated changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and the observed choice probabilities in the initial equilibrium in our baseline year (λ_{nit}, λ_{it}^{L}):

$$\begin{aligned}
\hat{\lambda}_{it}^{L}\lambda_{it}^{L} &= \frac{\sum_{k\in\mathbb{N}}\lambda_{kit}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{kit}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}}. \\
&= \frac{\sum_{k\in\mathbb{N}}\lambda_{kit}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{kit}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}. \\
&= \frac{\sum_{\ell\in\mathbb{N}}\lambda_{rit}\hat{\mathbb{B}}_{rt}^{\epsilon}\left(\hat{\mathbb{A}}_{it}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{it}}{\hat{H}_{it}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{rit}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{rt}}{\hat{H}_{rt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\left(\hat{\mathbb{A}}_{\ell t}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{\ell t}}{\hat{H}_{\ell t}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}.
\end{aligned} \tag{G.28}$$

Additionally, these changes in composite traded productivity ($\hat{\mathbb{A}}_{nt}^{T}$), the supply of floor space ($\hat{H}_{nt}$) and amenities ($\hat{\mathbb{B}}_{nt}$) ensure that the observed changes in rateable values ($\hat{\mathbb{Q}}_{nt}$) and the changes in bilateral commuting flows ($\hat{L}_{nit}$) from our baseline quantitative analysis are consistent with utility maximization, given our estimated changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and the observed choice probabilities (λ_{nit}) in the initial equilibrium in our baseline year:

$$\begin{aligned}
\hat{\lambda}_{nit}\lambda_{nit} &= \frac{\lambda_{nit}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{nit}^{-\epsilon}\hat{Q}_{nt}^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}}. \\
&= \frac{\lambda_{nit}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{nit}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}. \\
&= \frac{\lambda_{nit}\hat{\mathbb{B}}_{nt}^{\epsilon}\left(\hat{\mathbb{A}}_{it}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{it}}{\hat{H}_{it}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{nit}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\left(\hat{\mathbb{A}}_{\ell t}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{\ell t}}{\hat{H}_{\ell t}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}.
\end{aligned} \tag{G.29}$$

Given that the model exactly rationalizes the relative changes in residence and workplace choice probabilities ($\hat{\lambda}_{nt}^{R}$, $\hat{\lambda}_{nt}^{L}$), we ensure that it rationalizes the relative changes in levels of residence and workplace employment ($\hat{R}_{nt}$, $\hat{L}_{nt}$) using the observed relative changes in total city employment in the data ($\hat{L}_{\mathbb{N}t}$):

$$\hat{R}_{nt} = \hat{\lambda}_{nt}^{R}\hat{L}_{\mathbb{N}t}, \tag{G.30}$$

$$\hat{L}_{nt} = \hat{\lambda}_{nt}^{L}\hat{L}_{\mathbb{N}t}. \tag{G.31}$$

Finally, these changes in composite traded productivity ($\hat{\mathbb{A}}_{nt}^{T}$), the supply of floor space ($\hat{H}_{nt}$) and amenities ($\hat{\mathbb{B}}_{nt}$) ensure that the observed changes in rateable values ($\hat{\mathbb{Q}}_{nt}$) and residence employment ($\hat{R}_{nt}$) are consistent with our combined land and commuter market clearing condition (20), given our estimated changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and the rateable values ($\mathbb{Q}_{nt}$), wages (w_{nt}) and residents (R_{nt}) in our initial year:

$$\begin{aligned}\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda_{nit|n}^{R}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{nit}^{-\epsilon}}{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t|n}^{R}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}}\hat{w}_{it}w_{it}\right]\hat{R}_{nt}R_{nt} \\ + \frac{\beta^{H}}{\beta^{L}}\hat{w}_{nt}w_{nt}\left[\sum_{i\in\mathbb{N}} \frac{\lambda_{int|i}^{R}\hat{w}_{nt}^{\epsilon}\hat{\kappa}_{int}^{-\epsilon}}{\sum_{\ell\in\mathbb{N}}\lambda_{i\ell t|i}^{R}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{i\ell t}^{-\epsilon}}\hat{R}_{it}R_{it}\right].\end{aligned} \tag{G.32}$$

$$\begin{aligned}\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda_{nit|n}^{R}\left(\hat{\mathbb{A}}_{it}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{it}}{\hat{H}_{it}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{nit}^{-\epsilon}}{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t|n}^{R}\left(\hat{\mathbb{A}}_{\ell t}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{\ell t}}{\hat{H}_{\ell t}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{n\ell t}^{-\epsilon}}\left(\hat{A}_{it}\right)^{\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{it}}{\hat{H}_{it}}\right)^{-\beta^{H}/\beta^{L}} w_{it}\right]\hat{R}_{nt}R_{nt} \\ + \frac{\beta^{H}}{\beta^{L}}\left(\hat{\mathbb{A}}_{nt}^{T}\right)^{1/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\beta^{H}/\beta^{L}} w_{nt}\left[\sum_{i\in\mathbb{N}} \frac{\lambda_{int|i}^{R}\left(\hat{\mathbb{A}}_{nt}^{T}\right)^{\epsilon}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{int}^{-\epsilon}}{\sum_{\ell\in\mathbb{N}}\lambda_{i\ell t|i}^{R}\left(\hat{\mathbb{A}}_{\ell t}^{T}\right)^{\epsilon/\beta^{L}}\left(\frac{\hat{\mathbb{Q}}_{\ell t}}{\hat{H}_{\ell t}}\right)^{-\epsilon\beta^{H}/\beta^{L}}\hat{\kappa}_{i\ell t}^{-\epsilon}}\hat{R}_{it}R_{it}\right].\end{aligned}$$

We have thus established that ($\hat{H}_{nt}$, $\hat{\mathbb{A}}_{nt}^{T}$, $\hat{\mathbb{B}}_{nt}$) are structural residuals that exactly rationalize the observed data and solutions from our baseline quantitative analysis ($\hat{\mathbb{Q}}_{nt}$ $\hat{w}_{nt}$, $\hat{\lambda}_{nt}^{R}$, $\hat{\lambda}_{nt}^{L}$, $\hat{\lambda}_{nit}$, $\hat{R}_{nt}$, $\hat{L}_{nt}$) given our estimated changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), the values of variables in the initial equilibrium in our baseline year ($\mathbb{Q}_{nt}$ w_{nt}, λ_{nt}^{R}, λ_{nt}^{L}, λ_{nit}, $\lambda_{nit|n}^{R}$, R_{nt}, L_{nt}, $L_{\mathbb{N}t}$), and observed changes in total city employment ($\hat{L}_{\mathbb{N}t}$).

Using these properties, if we start at our initial equilibrium in year $t = 1921$ and undertake a counterfactual in which we change estimated commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and all our structural residuals simultaneously ($\hat{H}_{nt}$, $\hat{\mathbb{A}}_{nt}^{T}$, $\hat{\mathbb{B}}_{nt}$), we exactly replicate the observed changes in residence employment ($\hat{R}_{nt}$) and rateable values ($\hat{\mathbb{Q}}_{nt}$) and the changes in wages ($\hat{w}_{nt}$), workplace employment ($\hat{L}_{nt}$) and commuting ($\hat{L}_{nit}$) from our baseline quantitative analysis. In particular, we undertake this counterfactual using our solutions for ($\hat{H}_{nt}$, $\hat{\mathbb{A}}_{nt}^{T}$, $\hat{\mathbb{B}}_{nt}$), the variables in the initial equilibrium ($\mathbb{Q}_{nt}$, w_{nt}, λ_{nt}^{R}, λ_{nt}^{L}, $\lambda_{nit|n}^{R}$, $\bar{L}_{t}$), our estimated change in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and the observed change in total city employment ($\hat{\bar{L}}_{t}$), by solving the following system of five equations for the five endogenous changes in rateable values ($\hat{\mathbb{Q}}_{nt}$), wages ($\hat{w}_{nt}$), residence probabilities ($\hat{\lambda}_{nt}^{R}$), workplace probabilities ($\hat{\lambda}_{nt}^{L}$) and bilateral commuting probabilities ($\hat{\lambda}_{nit}$):

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = \left\{(1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda_{nit|n}^{R}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{nit}^{-\epsilon}}{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t|n}^{R}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}}\hat{w}_{it}w_{it}\right]\hat{\lambda}_{nt}^{R}\lambda_{nt}^{R} + \frac{\beta^{H}}{\beta^{L}}\hat{w}_{nt}w_{nt}\hat{\lambda}_{nt}^{L}\lambda_{nt}^{L}\right\}\hat{\bar{L}}\bar{L}_{t}, \tag{G.33}$$

$$\hat{\mathbb{Q}}_{nt} = \hat{H}_{nt}\left(\hat{\mathbb{A}}_{nt}^{T}\right)^{1/\beta^{H}}\hat{w}_{nt}^{-\beta^{L}/\beta^{H}}, \tag{G.34}$$

$$\hat{\lambda}_{nt}^{L}\lambda_{nt}^{L} = \frac{\sum_{i\in\mathbb{N}}\lambda_{int}\hat{\mathbb{B}}_{it}^{\epsilon}\hat{w}_{nt}^{\epsilon}\hat{\kappa}_{int}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{it}}{\hat{H}_{it}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}, \tag{G.35}$$

$$\hat{\lambda}_{nt}^{R}\lambda_{nt}^{R} = \frac{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{nt}}{\hat{H}_{nt}}\right)^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\left(\frac{\hat{\mathbb{Q}}_{kt}}{\hat{H}_{kt}}\right)^{-\epsilon(1-\alpha)}}. \tag{G.36}$$

In this specification, in which the supply of floor space (H_{nt}), productivity ($\mathbb{A}_{nt}^{T}$) and amenities ($\mathbb{B}_{nt}$) are changed exogenously, there are no agglomeration forces in the model and the supply of land is perfectly inelastic, which ensures the existence of a unique equilibrium. Therefore, these counterfactuals yield unique predictions for the impact of changes in commuting costs and the structural residuals on the spatial distribution of economic activity.

As well as undertaking a counterfactual in which we change commuting costs ($\hat{\kappa}_{nit}$) and all our structural residuals ($\hat{H}_{nt}$, $\hat{\mathbb{A}}_{nt}^{T}$, $\hat{\mathbb{B}}_{nt}$) simultaneously, we can also undertake counterfactuals in which we change only some of these structural residuals, while setting the relative changes in the other structural residuals equal to one. We use these counterfactuals to assess the relative importance of changes in commuting costs ($\hat{\kappa}_{nit}$), floor space ($\hat{H}_{nt}$), productivity ($\hat{\mathbb{A}}_{nt}^{T}$), and amenities ($\mathbb{B}_{nt}$) for the reorganization of economic activity within Greater London during our sample period. In a first counterfactual, we change only commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), which reveals the pure impact of the change in commuting costs on the spatial distribution of economic activity. In a second counterfactual, we change both commuting costs and the supply of floor space ($\hat{\kappa}_{nit}^{-\epsilon}$, $\hat{H}_{nt}$), which abstracts from changes in amenities and productivity across locations ($\hat{\mathbb{B}}_{nt} = 1$ and $\hat{\mathbb{A}}_{nt}^{T} = 1$). In a third counterfactual, we change commuting costs, the supply of floor space and amenities ($\hat{\kappa}_{nit}^{-\epsilon}$, $\hat{H}_{nt}$, $\hat{\mathbb{B}}_{nt}$), which abstracts from changes in productivity across locations ($\hat{\mathbb{A}}_{nt}^{T} = 1$). In a fourth counterfactual, we change commuting costs and all our structural residuals ($\hat{\kappa}_{nit}^{-\epsilon}$, $\hat{H}_{nt}$, $\hat{\mathbb{B}}_{nt}$, $\hat{\mathbb{A}}_{nt}^{T}$), which replicates the observed data and the results of our baseline quantitative analysis of the model.

In Figure G.1, we show the results of these counterfactuals for net commuting into the City of London for each decade back to before the first railway line in 1831. The solid gray line with no markers reports net commuting in our first counterfactual, in which we change only commuting costs. The solid black line with no markers shows net commuting in our second counterfactual, in which we change both commuting costs and the supply of floor space. The solid gray line with circle markers indicates net commuting in our third counterfactual, in which we change commuting costs, the supply of floor space and amenities. The solid black line with triangle markers gives net commuting in our fourth counterfactual, in which we change commuting costs, the supply of floor space, amenities and productivity (which replicates our baseline quantitative analysis). As apparent from the figure, we find that much of the decline in net commuting into the City of London following the construction of the railway network is driven by the pure change in commuting costs. We also find that each of the other three model components of changes in supply of floor space, productivity and amenities makes quantitatively relevant contributions towards the evolution of net commuting into the City of London over time. Taken together, these results provide further support for the idea that the invention of the steam railway was central to the emergence of the large-scale separation of workplace and residence in Greater London over our sample period.

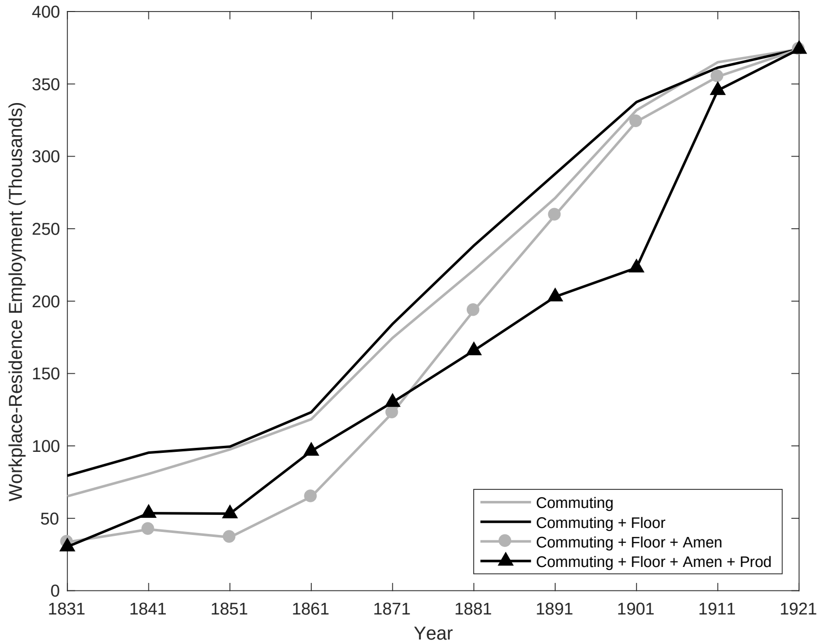

FIGURE G.1:
Model-Based Decomposition of the Evolution of Net Commuting into the City of London 1831-1921

Note: Figure shows counterfactual net commuting into the City of London in our model-based decompositions; (i) "Commuting" changes only commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$); (ii) "Commuting + Floor" changes both commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and the supply of floor space ($\hat{H}_{nt}$); (iii) "Commuting + Floor + Amen" changes commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), the supply of floor space ($\hat{H}_{nt}$) and composite residential amenities ($\hat{\mathbb{B}}_{nt}$); (iv) "Commuting + Floor + Amen + Prod" changes commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), the supply of floor space ($\hat{H}_{nt}$), composite residential amenities ($\hat{\mathbb{B}}_{nt}$) and composite traded productivity ($\hat{\mathbb{A}}_{nt}^{T}$), which reproduces our baseline quantitative analysis. In each case, we start in the initial equilibrium in our baseline year of $t = 1921$, and undertake counterfactuals for the new spatial equilibrium distribution of economic activity in each census decade going back to 1831 before the first railway line.

G4 Agglomeration Forces

While the main focus of our analysis is the implications of the construction of the railway network for the internal organization of economic activity within Greater London, in this section of the online appendix, we provide further details on our estimation of the strength of agglomeration forces in Section VII.C. of the paper.

We assume that traded productivity ($\mathbb{A}_{nt}^{T}$) depends on production fundamentals and production externalities. Production fundamentals (a_{nt}) capture features of physical geography that make a location more or less productive independently of the surrounding density of economic activity (e.g. access to natural water). Production externalities capture spillovers from local economic activity and are assumed to be a constant elasticity function of workplace employment density (L_{nt}/K_n):

$$\mathbb{A}_{nt}^{T} = \left(\frac{L_{nt}}{K_n}\right)^{\eta^L} a_{nt}, \tag{G.37}$$

where we focus on a borough's own workplace employment density, because of the relatively large area of the boroughs in our data, and the typical high rates of spatial decay of spillovers from local economic activity.

Similarly, we assume that amenities ($\mathbb{B}_{nt}$) depend on residential fundamentals and residential externalities. Residential fundamentals (b_{nt}) capture features of physical geography that make a location more or less attractive for residence independently of the surrounding density of economic activity (e.g. green areas). Residential externalities capture spillovers from local economic activity and are assumed to be a constant elasticity function of residence employment density (R_{nt}/K_n):

$$\mathbb{B}_{nt} = \left(\frac{R_{nt}}{K_n}\right)^{\eta^R} b_{nt}, \tag{G.38}$$

where we again focus on a borough's own residence employment density, because of the relatively large area of the boroughs in our data, and the typical high rates of spatial decay of spillovers from local economic activity.

Taking logarithms in equations (G.37) and (G.38), and differencing between an earlier year $\tau < t$ and our baseline year of $t = 1921$, we obtain the following expressions for the log change in composite productivity and amenities from equations (30) and (31) in the paper:

$$\ln \hat{\mathbb{A}}^T_{nt} = \varsigma^L + \eta^L \ln \hat{L}_{nt} + \ln \hat{a}_{nt}, \tag{G.39}$$

$$\ln \hat{\mathbb{B}}_{nt} = \varsigma^R + \eta^R \ln \hat{R}_{nt} + \ln \hat{b}_{nt}, \tag{G.40}$$

where recall $\hat{L}_{nt} = L_{n\tau}/L_{nt}$ for $\tau < t = 1921$; log land area has differenced out on the right-hand side, because it is constant over time; the constants ς^L and ς^R control for any factors that are common across all locations within Greater London, such as common changes in productivity or amenities, or changes in expected utility in the wider economy; and $\ln \hat{a}_{nt}$ and $\ln \hat{b}_{nt}$ capture idiosyncratic shocks to production and residential fundamentals.

A key challenge in estimating the strength of agglomeration forces (η^L, η^R) in equations (G.39) and (G.40) is that workplace and residence employment are endogenous to productivity and amenities. In particular, the workplace choice probabilities (λ^L_{nt}) in equation (7) in the paper depend on the wage (w_{nt}), which in turn depends on traded productivity ($\mathbb{A}^T_{nt}$). Therefore, changes in workplace employment are likely to be positively correlated with idiosyncratic shocks to production fundamentals ($\ln \hat{a}_{nt}$) in the error term, thereby inducing an upwards bias in the estimated production elasticity (η^L). Similarly, the residence choice probabilities (λ^R_{nt}) in equation (7) in the paper are determined by amenities ($\mathbb{B}_{nt}$). Hence, changes in residence employment are likely to be positively correlated with idiosyncratic shocks to residential fundamentals ($\ln \hat{b}_{nt}$), thus imparting an upward bias in the estimated residential elasticity (η^R).

To address this challenge, we use the quasi-experimental variation from the invention of the steam railway. In particular, we estimate the strength of agglomeration forces by requiring that the observed reorganization of workplace and residence employment within Greater London is explained by the model's mechanisms of a change in commuting costs and agglomeration forces, rather than by systematic changes in production and residential fundamentals. In particular, we estimate equations (G.39) and (G.40) using two stage least squares, instrumenting the log changes in workplace and residence employment with indicator variables for 5 km distance grid cells from the Guildhall, where our excluded category is locations more than 20 km away. To allow changes in production and residential fundamentals to depend on the initial density of economic activity, we include as controls initial log employment (workplace and residence respectively) and log land area. Additionally, we include as a control an indicator variable for whether a borough is located in the London County Council (LCC) area to allow for potential differences in the supply of local public goods within and outside the LCC boundaries. Therefore, we estimate the production and residential externalities parameters (η^L and η^R respectively) using the identifying assumption that conditional on our controls the idiosyncratic shocks to production and residential fundamentals ($\ln \hat{a}_{nt}$, $\ln \hat{b}_{nt}$) are unrelated to distance from the Guildhall, and hence have the same mean value across these distance grid cells.

In Table II in Section VII.C. of the paper, we report the second-stage estimation results for this two-stage least squares estimation. In Table G.1 below, we report the corresponding first-stage estimation results. As apparent from the table, we find that our instruments have power in the first-stage regressions, with first-stage F-statistics above the conventional threshold of 10. In Hansen-Sargan overidentification tests, we are unable to reject the model's overidentifying restrictions at conventional levels of significance, as reported in Table II in Section VII.C. of the paper.

TABLE G.1:
First-Stage Regressions for our Instrumental Variables Estimates of the Strength of Agglomeration Forces

	(1)	(2)
	$\ln \widehat{L}_{nt}$	$\ln \widehat{R}_{nt}$
$\mathbb{I}_n^{0-5}$	1.149^{***}	2.094^{***}
	(0.404)	(0.531)
$\mathbb{I}_n^{5-10}$	-0.103	0.160
	(0.238)	(0.380)
$\mathbb{I}_n^{10-15}$	-0.538^{***}	-0.238
	(0.179)	(0.242)
$\mathbb{I}_n^{15-20}$	-0.253	-0.268
	(0.166)	(0.192)
$\ln L_{nt}$	-0.364^{***}	-
	(0.074)	
$\ln R_{nt}$	-	-0.672^{***}
		(0.113)
$\ln K_n$	0.323^{***}	0.318^{***}
	(0.099)	(0.106)
$\mathbb{I}_{nt}^{\text{LCC}}$	0.762^{***}	1.190^{***}
	(0.176)	(0.240)
F-statistic Excluded Variables	11.26	12.76
Estimation	OLS	OLS
Observations	99	99
R-squared	0.527	0.699

Note: $\ln \widehat{L}_{nt} = \ln (L_{n\tau}/L_{nt})$ is the log change in workplace employment; $\ln \widehat{R}_{nt} = \ln (R_{n\tau}/R_{nt})$ is the log change in residence employment; $\{\mathbb{I}_n^{0-5}, \mathbb{I}_n^{5-10}, \mathbb{I}_n^{10-15}, \mathbb{I}_n^{15-20}\}$ are indicator variables for 5 km distance grid cells from the Guildhall in the center of the City of London; the excluded category is $>$ 20 km from the Guildhall; $\ln L_{nt}$ is initial log workplace employment; $\ln R_{nt}$ is initial log residence employment; $\ln K_n$ is log land area; $\mathbb{I}_{nt}^{\text{LCC}}$ is an indicator variable for whether a borough is located within the London County Council (LCC) area; first-stage F-statistic is the F-statistic for the joint significance of the distance grid cell indicators; OLS refers to ordinary least squares; Heteroskedasticity robust standard errors in parentheses: * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

H Counterfactuals

In this section of the online appendix, we report further details on the counterfactuals in Section VIII. of the paper. We undertake these counterfactuals under a range of alternative assumptions about the floor space supply elasticity (μ) and the strength of agglomeration forces (η^L, η^R).

Given a change in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) from the removal of parts of the railway network and the observed values of variables ($\mathbb{Q}_{nt}$, λ_{nt}^L, λ_{nt}^R, $\bar{L}_t$, $\lambda_{nit|n}^R$, λ_{nit}, w_{it}) in our initial equilibrium in year $t = 1921$, we solve for counterfactual changes in the following five endogenous variables: (i) the workplace choice probability ($\hat{\lambda}_{nt}^L$); (ii) the residence choice probability ($\hat{\lambda}_{nt}^R$); (iii) the price of floor space ($\hat{Q}_{nt}$); (iv) the wage ($\hat{w}_{nt}$); (v) total city employment ($\hat{L}_{\mathbb{N}t}$). These five endogenous variables solve the following system of five equations for the general equilibrium of the model: (i) the land market clearing condition (H.1); (ii) the zero-profit condition (H.2); (iii) the workplace choice probability (H.3), (iv) the residence choice probability (H.4); and (v) the population mobility condition (H.5):

$$\hat{Q}_{nt}^{1+\mu}\mathbb{Q}_{nt} = \left\{(1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}_{it}^{\epsilon}\hat{\kappa}_{nit}^{-\epsilon}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}}\hat{w}_{it}w_{it}\right]\hat{\lambda}^R_{nt}\lambda^R_{nt} + \frac{\beta^H}{\beta^L}\hat{w}_{nt}w_{nt}\hat{\lambda}^L_{nt}\lambda^L_{nt}\right\}\hat{\bar{L}}\bar{L}_t, \tag{H.1}$$

$$\hat{Q}_{nt} = \left(\hat{\mathbb{A}}^T_{nt}\right)^{1/\beta^H} \hat{w}_{nt}^{-\beta^L/\beta^H}, \tag{H.2}$$

$$\hat{\lambda}^L_{nt}\lambda^L_{nt} = \frac{\sum_{i\in\mathbb{N}}\lambda_{int}\hat{\mathbb{B}}_{it}^{\epsilon}\hat{w}_{nt}^{\epsilon}\hat{\kappa}_{int}^{-\epsilon}\hat{Q}_{it}^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}}, \tag{H.3}$$

$$\hat{\lambda}^R_{nt}\lambda^R_{nt} = \frac{\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell t}^{-\epsilon}\hat{Q}_{nt}^{-\epsilon(1-\alpha)}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}}, \tag{H.4}$$

$$\left(\frac{\hat{L}_{\mathbb{N}t}L_{\mathbb{N}t}}{L_{\mathbb{M}t}}\right)^{\frac{1}{\epsilon}} = \left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\lambda_{k\ell t}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{k\ell t}^{-\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}\right]^{\frac{1}{\epsilon}}, \tag{H.5}$$

where recall that $\hat{x}_t = x_\tau/x_t$; we have used $\hat{\mathbb{Q}}_{nt} = \hat{Q}_{nt}\hat{H}_{nt} = \hat{Q}_{nt}^{1+\mu}$; and we hold total employment and expected utility in the economy as a whole constant ($\hat{L}_{\mathbb{M}t} = 1$ and $\hat{\bar{U}}_t = 1$). The relative changes in composite traded productivity ($\hat{\mathbb{A}}^T_{nt}$) and composite amenities ($\hat{\mathbb{B}}_{it}$) satisfy:

$$\hat{\mathbb{A}}^T_{nt} = \hat{L}_t^{\eta^L}, \tag{H.6}$$

$$\hat{\mathbb{B}}_{nt} = \hat{R}_t^{\eta^R}, \tag{H.7}$$

where we hold production and residential fundamentals constant ($\hat{a}_{nt} = 1$ and $\hat{b}_{nt} = 1$).

In the special case of the model in which productivity, amenities and the supply of floor space are exogenous $\eta^L = \eta^R = \mu = 0$, there are no agglomeration forces and the supply of land is perfectly inelastic, which ensures the existence of a unique equilibrium, as shown in Proposition H.1 below. Therefore, our counterfactuals yield unique predictions for the impact of the change in commuting costs on the spatial distribution of economic activity. In the presence of agglomeration forces ($\eta^L > 0$ and $\eta^R > 0$) and an elastic supply of land ($\mu > 0$), whether or not the equilibrium is unique depends on the strength of these agglomeration forces relative to the model's congestion forces and the exogenous differences in production and residential fundamentals across locations. For the values of agglomeration forces considered in our counterfactuals, we obtain the same counterfactual equilibrium regardless of our starting values for the counterfactual changes in the model's endogenous variables.

Proposition H.1 *Assume exogenous, finite and strictly positive location characteristics ($P_t \in (0,\infty)$, $A_{nt} \in (0,\infty)$, $B_{nit} \in (0,\infty)\times(0,\infty)$, $\kappa_{nit} \in (0,\infty)\times(0,\infty)$, $H_{nt} = H_n \in (0,\infty)$), which corresponds to $\eta^L = \eta^R = \mu = 0$. Under these assumptions, there exists a unique general equilibrium vector of four variables in each location (λ^L_{nt}, λ^R_{nt}, Q_{nt}, w_{nt}) and the scalar ($L_{\mathbb{N}t}$), given expected utility ($\bar{U}_t$) and total employment ($L_{\mathbb{M}t}$) in the wider economy.*

Proof. Assume exogenous, finite and strictly positive location characteristics ($P_t \in (0,\infty)$, $A_{nt} \in (0,\infty)$, $B_{nit} \in (0,\infty)\times(0,\infty)$, $\kappa_{nit} \in (0,\infty)\times(0,\infty)$, $H_{nt} = H_n = hK_n \in (0,\infty)$), which corresponds to $\eta^L = \eta^R = \mu = 0$. Under these assumptions, all locations are incompletely specialized as both workplaces and residences, because the support of the Fréchet distribution for idiosyncratic amenities is unbounded from above. Using the probability of residing in a location (equation (7) in the paper for λ^R_{nt}), the probability of working in a location (equation (7) in the paper for λ^L_{nt}), the zero-profit condition in equation (28) in the paper, and the population mobility condition between the city and the larger economy in equation (9) in the paper, the fraction of workers residing in location n can be written as:

$$\lambda_{nt}^{R}=\frac{R_{nt}}{L_{\mathbb{N}t}}=\left(\frac{\vartheta}{\bar{U}_t}\right)^{\epsilon}\left(\frac{L_{\mathbb{M}t}}{L_{\mathbb{N}t}}\right)\sum_{\ell\in\mathbb{N}}B_{n\ell t}^{\epsilon}P_t^{\epsilon/\beta^L-\epsilon\alpha}A_{\ell t}^{\epsilon/\beta^L}\kappa_{n\ell t}^{-\epsilon}Q_{\ell t}^{-\epsilon\beta^H/\beta^L}Q_{nt}^{-\epsilon(1-\alpha)},$$

while the fraction of workers employed in location n can be written as:

$$\lambda_{nt}^{L}=\frac{L_{nt}}{L_{\mathbb{N}t}}=\left(\frac{\vartheta}{\bar{U}_t}\right)^{\epsilon}\left(\frac{L_{\mathbb{M}t}}{L_{\mathbb{N}t}}\right)\sum_{k\in\mathbb{N}}B_{knt}^{\epsilon}P_t^{\epsilon/\beta^L-\epsilon\alpha}A_{nt}^{\epsilon/\beta^L}\kappa_{knt}^{-\epsilon}Q_{nt}^{-\epsilon\beta^H/\beta^L}Q_{kt}^{-\epsilon(1-\alpha)},$$

and expected worker income conditional on residing in block i from equation (15) in the paper can be written as:

$$v_{nt}=\sum_{i\in\mathbb{N}}\frac{B_{nit}^{\epsilon}A_{it}^{\epsilon/\beta^L}\kappa_{nit}^{-\epsilon}Q_{it}^{-\epsilon\beta^H/\beta^L}}{\sum_{\ell\in\mathbb{N}}B_{n\ell t}^{\epsilon}A_{\ell t}^{\epsilon/\beta^L}\kappa_{n\ell t}^{-\epsilon}Q_{\ell t}^{-\epsilon\beta^H/\beta^L}}\left[P_t^{1/\beta^L}A_{it}^{1/\beta^L}Q_{it}^{-\beta^H/\beta^L}\right],$$

and the land market clearing condition from equation (16) in the paper can be written as:

$$\left(\frac{\beta^H}{\beta^L}\right)\frac{w_{nt}\lambda_{nt}^{L}}{Q_{nt}}+(1-\alpha)\frac{v_{nt}\lambda_{nt}^{R}}{Q_{nt}}=\frac{hK_n}{L_{\mathbb{N}t}}.$$

Combining the above relationships, this land market clearing condition can be re-expressed as:

$$\begin{aligned}D_{nt}(\boldsymbol{Q_t})&=\tfrac{\beta^H}{\beta^L}\left[\frac{P_t^{1/\beta^L}A_{nt}^{1/\beta^L}}{Q_{nt}^{1+\beta^H/\beta^L}}\right]\left[\sum_{k\in\mathbb{N}}\frac{B_{knt}^{\epsilon}P_t^{\epsilon/\beta^L-\epsilon\alpha}A_{nt}^{\epsilon/\beta^L}\kappa_{knt}^{-\epsilon}}{Q_{nt}^{\epsilon\beta^H/\beta^L}Q_{kt}^{\epsilon(1-\alpha)}}\right]\\&+\tfrac{1-\alpha}{Q_{nt}}\left[\sum_{i\in\mathbb{N}}\left(\frac{B_{nit}^{\epsilon}A_{it}^{\epsilon/\beta^L}\kappa_{nit}^{-\epsilon}Q_{it}^{-\epsilon\beta^H/\beta^L}}{\sum_{\ell\in\mathbb{N}}B_{n\ell t}^{\epsilon}A_{\ell t}^{\epsilon/\beta^L}\kappa_{n\ell t}^{-\epsilon}Q_{\ell t}^{-\epsilon\beta^H/\beta^L}}\right)\frac{P_t^{1/\beta^L}A_{it}^{1/\beta^L}}{Q_{it}^{\beta^H/\beta^L}}\right]\left[\sum_{\ell\in\mathbb{N}}\frac{B_{n\ell t}^{\epsilon}P_t^{\epsilon/\beta^L-\epsilon\alpha}A_{\ell t}^{\epsilon/\beta^L}\kappa_{n\ell t}^{-\epsilon}}{Q_{\ell t}^{\epsilon\beta^H/\beta^L}Q_{nt}^{\epsilon(1-\alpha)}}\right]=hK_n,\end{aligned}$$

for all $n\in\mathbb{N}$, where we have chosen units in which to measure utility such that $(\bar{U}_t/\vartheta)^{\epsilon}/L_{\mathbb{M}t}=1$ for a given year t. The above land market clearing condition provides a system of equations for the N boroughs in terms of the N unknown floor space prices Q_{nt}, which has the following properties:

$$\lim_{Q_{nt}\to 0}D_{nt}(\boldsymbol{Q_t})=\infty>hK_n,\qquad\lim_{Q_{nt}\to\infty}D_{nt}(\boldsymbol{Q_t})=0<hK_n,$$

$$\frac{dD_{nt}(\boldsymbol{Q_t})}{dQ_{nt}}<0,\qquad\frac{dD_{nt}(\boldsymbol{Q_t})}{dQ_{it}}<0,\qquad\left|\frac{dD_{nt}(\boldsymbol{Q_t})}{dQ_{nt}}\right|>\left|\frac{dD_{nt}(\boldsymbol{Q_t})}{dQ_{it}}\right|.$$

It follows that there exists a unique vector of floor space prices $\boldsymbol{Q_t}$ that solves this system of land market clearing conditions. Having solved for the vector of floor space prices ($\boldsymbol{Q_t}$), the vector of wages $\boldsymbol{w_t}$ follows immediately from the zero-profit condition for production in equation (28) in the paper. Given floor space prices ($\boldsymbol{Q_t}$) and wages ($\boldsymbol{w_t}$), the probability of residing in a location ($\boldsymbol{\lambda_t^R}$) follows immediately from equation (7) in the paper, and the probability of working in a location ($\boldsymbol{\lambda_t^L}$) follows immediately from equation (7) in the paper. Having solved for ($\boldsymbol{\lambda_t^L}$, $\boldsymbol{\lambda_t^R}$, $\boldsymbol{Q_t}$, $\boldsymbol{w_t}$), the total measure of workers residing in the city can be recovered from our choice of units in which to measure utility $((\bar{U}_t/\vartheta)^{\epsilon}/L_{\mathbb{M}t}=1)$ for our given year t, which together with the population mobility condition in equation (9) in the paper implies:

$$L_{\mathbb{N}t}=\left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}B_{k\ell t}^{\epsilon}P_t^{\epsilon/\beta^L-\epsilon\alpha}A_{\ell t}^{\epsilon/\beta^L}\kappa_{k\ell t}^{-\epsilon}Q_{\ell t}^{-\epsilon\beta^H/\beta^L}Q_{kt}^{-\epsilon(1-\alpha)}\right].$$

We therefore obtain $\boldsymbol{L_t}=\boldsymbol{\lambda_t^L}L_{\mathbb{N}t}$ and $\boldsymbol{R_t}=\boldsymbol{\lambda_t^R}L_{\mathbb{N}t}$. This completes the determination of the equilibrium vector ($\boldsymbol{\lambda_t^L}$, $\boldsymbol{\lambda_t^R}$, $L_{\mathbb{N}t}$, $\boldsymbol{Q_t}$, $\boldsymbol{w_t}$). ■

Under our assumptions of population mobility and a constant expected utility in the wider economy, the total population of Greater London adjusts such that the expected utility of workers is unaffected by the construction of the railway network. Therefore, as in the classical approach to valuing public goods using land values following George (1879), the welfare gains from the new transport technology are experienced by landlords through changes in the value of land and buildings. In Section VIII. of the paper, we assess the magnitude of these welfare gains by comparing the counterfactual changes in the net present value of rateable values from the removal of the railway network to its construction costs. To measure construction costs, we distinguish between shallow "cut-and-cover" underground railways, deep "bored-tube" underground railways, and overground railways. We measure the length of each of these types of railways in Greater London, classifying the parts of an underground railway company's lines that run above ground as overground railways. We measure total construction costs as the length of each type of line times the construction cost per mile for that type of line. As discussed further in Section J6 of the online appendix, we measure construction cost per mile using historical estimates of authorized capital per mile for the private-sector companies that built these lines, which yields estimates of $\pounds 555,000$ per mile for bored-tube underground railways, $\pounds 355,000$ per mile for cut-and-cover underground railways, and $\pounds 60,000$ per mile for overground railways (all in 1921 prices).

In Table III in the paper, we report the results of these comparisons of the economic impact of the railway network to its construction costs. In Table H.1 below, we report the results of an additional robustness test, in which we address the concern that the floor space supply elasticity in 19th-century London could be larger than in other settings with more building regulations. Table H.1 has a similar structure to Table III in the paper. In the top panel, we remove the entire railway network. In the middle panel, we eliminate only the underground railway network. In the bottom panel, we examine the extent to which there were diminishing returns to the construction of the railway network, by only removing overground and underground railway lines constructed in the final decade of our sample from 1911-1921.

In Column (1) of Table H.1, we reproduce the results from Column (1) of Table III in the paper, with an inelastic supply of floor space and constant productivity and amenities ($\mu = \eta^L = \eta^R = 0$). In Column (2) of Table H.1, we set the floor space supply elasticity equal to half our calibrated value for 19th-century London and continue to assume constant productivity and amenities ($\mu = 1.83/2 = 0.92$ and $\eta^L = \eta^R = 0$). In Column (3) of Table H.1, we set the floor space supply elasticity equal to our calibrated value for 19th-century London and continue to assume constant productivity and amenities ($\mu = 1.83$ and $\eta^L = \eta^R = 0$), which corresponds to Column (2) of Table III in the paper. Across all panels and columns, we find substantial impacts of the construction of the railway network on the total value of land and buildings in Greater London. Consistent with diminishing returns to the expansion of the railway network, we find lower ratios of reductions in the net present value of rateable values to savings in construction costs for overground and underground lines added in the final decade of our sample (bottom panel) than for the entire railway network (top panel). Nevertheless, in all cases, the change in the net present value of rateable values exceeds construction costs. Therefore, in both Table III in the paper and Table H.1 below, we find a similar pattern of results for the economic impact of the railway network across a range of alternative parameter values.

TABLE H.1:
Counterfactuals for Removing the Entire Railway Network, the Entire Underground Railway Network, and Railway Lines Constructed from 1911-1921

	(1)	(2)	(3)
Floor Space Supply Elasticity	$\mu = 0$	$\mu = 0.92$	$\mu = 1.83$
Production Agglomeration Force	$\eta^L = 0$	$\eta^L = 0$	$\eta^L = 0$
Residence Agglomeration Force	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0$
Removing the Entire Overground and Underground Railway Network			
Economic Impact			
Rateable Value	$-£8.24$m	$-£12.26$m	$-£15.55$m
NPV Rateable Value (3 percent)	$-£274.55$m	$-£408.72$m	$-£518.26$m
NPV Rateable Value (5 percent)	$-£164.73$m	$-£245.23$m	$-£310.96$m
Construction Costs			
Cut-and-Cover Underground		$-£9.96$m	
Bored-tube Underground		$-£22.90$m	
Overground Railway		$-£33.19$m	
Total All Railways		$-£66.05$m	
Ratio Economic Impact / Construction Cost			
$\frac{\text{NPV Rateable Value (3 percent)}}{\text{Construction Cost}}$	4.16	6.19	7.85
$\frac{\text{NPV Rateable Value (5 percent)}}{\text{Construction Cost}}$	2.49	3.71	4.71
Removing the Entire Underground Railway Network			
Economic Impact			
Rateable Value	$-£2.65$m	$-£4.52$m	$-£6.21$m
NPV Rateable Value (3 percent)	$-£88.46$m	$-£150.58$m	$-£206.87$m
NPV Rateable Value (5 percent)	$-£53.08$m	$-£90.35$m	$-£124.12$m
Construction Costs			
Cut-and-Cover Underground		$-£9.96$m	
Bored-tube Underground		$-£22.90$m	
Total All Underground		$-£32.86$m	
Ratio Economic Impact / Construction Cost			
$\frac{\text{NPV Rateable Value (3 percent)}}{\text{Construction Cost}}$	2.69	4.58	6.30
$\frac{\text{NPV Rateable Value (5 percent)}}{\text{Construction Cost}}$	1.62	2.75	3.78
Removing Overground and Underground Railway Lines Constructed from 1911-21			
Economic Impact			
Rateable Value	$-£0.17$m	$-£0.22$m	$-£0.24$m
NPV Rateable Value (3 percent)	$-£5.63$m	$-£7.21$m	$-£8.09$m
NPV Rateable Value (5 percent)	$-£3.38$m	$-£4.33$m	$-£4, 86$m
Construction Costs			
Cut-and-Cover Underground		$-£0.00$m	
Bored-tube Underground		$-£2.35$m	
Overground Railway		$-£0.34$m	
Total All Railways		$-£2.69$m	
Ratio Economic Impact / Construction Cost			
$\frac{\text{NPV Rateable Value (3 percent)}}{\text{Construction Cost}}$	2.09	2.68	3.01
$\frac{\text{NPV Rateable Value (5 percent)}}{\text{Construction Cost}}$	1.26	1.61	1.81

Note: Counterfactuals start in our baseline year of 1921 and remove either the entire railway network or parts thereof; we hold the omnibus and tram network constant at its 1921 structure; all values reported in the table are expressed in millions of 1921 pounds sterling; $\mu = 0$ corresponds to an inelastic supply of floor space; $\mu = 0.92$ equals half our calibrated floor space supply elasticity; $\mu = 1.83$ is our calibrated floor space supply elasticity; all specifications assume exogenous productivity and amenities ($\eta^L = \eta^R = 0$); all specifications assume population mobility between Greater London and the wider economy, with the elasticity of population supply determined by our calibrated Fréchet shape parameter of $\epsilon = 5.25$; net present values are evaluated over an infinite lifetime, assuming either 3 or 5 percent discount rate; construction costs are based on capital issued per mile for cut-and-cover, bored-tube and surface railway lines and the length of lines of each type of railway in Greater London in 1921, as discussed further in Section J6 of this online appendix.

I Additional Empirical Results

In this section of the online appendix, we report additional empirical results that are discussed in the paper. In Section I1, we present evidence on passenger journeys by public transport in Greater London over time, as discussed in Section IV. of the paper. In Section I2, we include additional results for the difference-in-differences event-study specifications from Section IV.B. of the paper. In Section I3, we report the results of an alternative non-parametric reduced-form empirical specification, in which we estimate a separate railway treatment effect for each parish, as discussed in Section IV. of the paper. In Section I4, we present additional evidence that commuting is pervasive across workers in all occupations and industries, as discussed in Section V. of the paper. In Section I5, we provide evidence that the same change in the organization of economic activity as observed in London is found for other metropolitan areas following the transport improvements of the 19th century, as discussed in Section IV. of the paper. In Section I6, we report additional gravity equation estimation results, as discussed in Section VI. of the paper. In Section I7, we present an overidentification check on our calibrated floor space supply elasticity, as discussed in Section VII. of the paper.

I1 Further Reduced-Form Evidence for London

In Figure I.1, we provide further evidence on a change in transport use by graphing passenger journeys using public transport per head of population in the County of London in each year (see also Barker 1980). Public transport includes underground and overground railways, horse and electric trams, short-stage coaches, and horse and motor omnibuses. As discussed in Section IV.A. of the paper, the increasing specialization of locations as workplace or residence from the mid-19th century onwards is reflected in a sharp increase in the intensity of public transport use, with journeys per head of population increasing from around 7 in 1834 to just under 400 in 1921.

I2 Additional Event-Study Difference-in-Differences Results

In this section of the online appendix, we report additional empirical results and robustness tests for the event-study difference-in-differences specifications that are reported in Section IV.B. of the paper. In Table I.1, we report the regression results for our baseline specification, which corresponds to the estimated treatment effects shown in Figure V in the paper. Column (1) corresponds to the specification displayed in the top panel of Figure V in the paper. Column (2) corresponds to the first specification (gray circle markers) presented in the middle and bottom panels of Figure V in the paper; Column (3) corresponds to the second specification (black triangle markers) shown in the middle and bottom panels of Figure V in the paper; and Column (4) corresponds to the third specification (gray triangle markers) displayed in the middle and bottom panels of Figure V in the paper.

Our baseline specification in Figure V in the paper and Table I.1 reports standard errors clustered by borough, which allows the error term to be serially correlated within parishes over time and to be spatially-correlated across parishes within boroughs. In Table I.2, we report a robustness test using standard errors clustered by parish, which only allows the error term to be serially correlated within parishes over time. By construction, the coefficients in Table I.2 are identical to those in Table I.1. In principle, the standard errors clustered by parish can be either larger or smaller than the standard errors clustering by borough, depending on the properties of the variance-covariance matrix. In practice, comparing the results in Tables I.1 and I.2, we find that the standard errors clustered by parish are typically smaller than the standard errors clustered by borough. To be conservative, we focus in our baseline specification on the standard

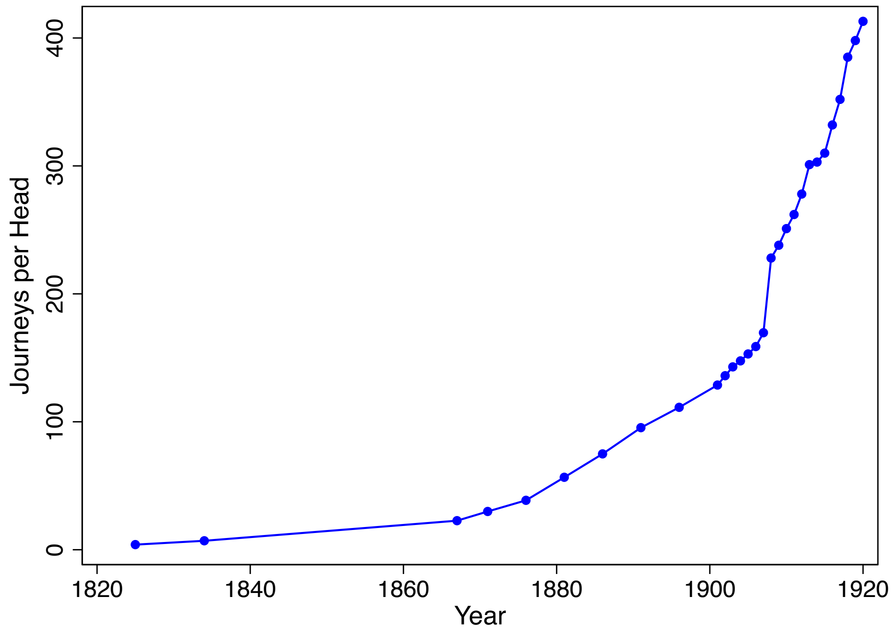

FIGURE I.1:
Public Transport Passenger Journeys per Head in the County of London

Note: Journeys per head measured as millions of passengers carried per year on public transport divided by the population of the County of London. Public transport includes underground and overground railways, horse and electric trams, short-stage coaches, and horse and motor omnibuses. Sources: Barker (1980) and London County Council (1921).

errors clustered by borough, as reported in Table I.1.

In our baseline specification in Figure V in the paper and Table I.1, we include six interaction terms for decades from 10 to 60 years before and after a parish receives a railway station. We aggregate treatment years more than 60 years before and more than 60 years after the arrival of a railway station to ensure that these initial and final categories have a sufficient number of observations. As the railway arrives in some parishes in a different calendar years from others, parishes differ in terms of the number of treatment years before and after the arrival of a railway station. Therefore, as we vary the number of years before and after the treatment, we change the composition of the treatment group of parishes. In Table I.3, we report a robustness test in which we hold the composition of the treated parishes constant, by restricting attention to treatment windows 30 years before and after the arrival of a railway station, and dropping any treated parishes with less than 30 years before and after the arrival of a railway station. In this specification, the coefficients for -30 years before the treatment are perfectly colinear with the parish-specific time trends, and hence are excluded. As apparent from the table, we find a similar pattern of results as in our baseline specification. We find substantial and significant deviations in log population from the parish-specific time trends immediately after the arrival of the railway and small and often insignificant deviations from these trends before the arrival of the railway.

TABLE I.1:
Event-Study Treatment Effects for the Arrival of an Overground or Underground Railway Station in Greater London Parishes from 1801-1901 (Baseline Specification)

	(1)	(2)	(3)	(4)
	$\log R_{jt}$	$\log R_{jt}$	$\log R_{jt}$	$\log R_{jt}$
$\beta_{\tau=60}$	1.438***	1.654**	1.650**	0.568
	(0.524)	(0.689)	(0.671)	(0.351)
$\beta_{\tau=50}$	1.154***	1.335**	1.361**	0.524*
	(0.424)	(0.561)	(0.552)	(0.304)
$\beta_{\tau=40}$	0.758**	0.928**	0.989**	0.362
	(0.289)	(0.417)	(0.411)	(0.236)
$\beta_{\tau=30}$	0.511**	0.676**	0.755**	0.300
	(0.210)	(0.338)	(0.333)	(0.194)
$\beta_{\tau=20}$	0.273**	0.378*	0.380*	0.136
	(0.120)	(0.195)	(0.193)	(0.119)
$\beta_{\tau=10}$	0.104*	0.149	0.152*	0.052
	(0.060)	(0.091)	(0.090)	(0.059)
$\beta_{\tau=0}$	–	–	–	–
$\beta_{\tau=-10}$	-0.048	-0.065	-0.039	0.051
	(0.066)	(0.073)	(0.075)	(0.056)
$\beta_{\tau=-20}$	-0.085	-0.109	-0.057	0.128
	(0.125)	(0.132)	(0.137)	(0.097)
$\beta_{\tau=-30}$	-0.092	-0.129	-0.028	0.201
	(0.166)	(0.181)	(0.193)	(0.142)
$\beta_{\tau=-40}$	-0.109	-0.150	-0.038	0.211
	(0.224)	(0.238)	(0.250)	(0.185)
$\beta_{\tau=-50}$	-0.084	-0.113	0.010	0.259
	(0.268)	(0.273)	(0.286)	(0.221)
$\beta_{\tau=-60}$	-0.093	-0.100	0.037	0.248
	(0.322)	(0.313)	(0.328)	(0.261)
$\gamma_{\tau=30}$	–	-1.501***	-1.002***	-0.465**
		(0.118)	(0.303)	(0.188)
$\gamma_{\tau=20}$	–	-0.833***	-0.431**	-0.126
		(0.075)	(0.195)	(0.133)
$\gamma_{\tau=10}$	–	-0.299***	-0.173**	-0.012
		(0.040)	(0.071)	(0.052)
$\gamma_{\tau=0}$	–	–	–	–
$\gamma_{\tau=-10}$	–	0.045	-0.025	-0.085*
		(0.039)	(0.046)	(0.045)
$\gamma_{\tau=-20}$	–	-0.003	-0.088	-0.246***
		(0.068)	(0.079)	(0.085)
$\gamma_{\tau=-30}$	–	0.033	-0.203	-0.406***
		(0.085)	(0.134)	(0.111)
Travel Time Reduction	–	–	–	-2.257***
				(0.208)
Travel Time Reduction $\times$ Distance to Guildhall	–	–	–	0.720***
				(0.066)
Central London definition		City of London	< 5 km	< 5 km
Observations	3113	3113	3113	3113
R-squared	0.981	0.981	0.981	0.983

Note: Observations are parishes and census years; all regressions include parish fixed effects, census-year dummies, and parish-specific census-year trends; railway treatment is defined based on whether a parish has an overground or underground railway station; τ is a treatment year indicator, which equals census year minus the last census year in which a parish had no railway, so that positive values of τ correspond to post-treatment years. For example, if the railway arrives in a parish in 1836, census year 1831 corresponds to $\tau = 0$, census year 1841 corresponds to $\tau = 10$ and census year 1821 corresponds to $\tau = -10$; $\tau = 0$ is the excluded category; β_τ is the rail treatment for treatment year τ; γ_τ allows the rail treatment for treatment year τ to differ between the central and outlying parts of Greater London; travel time reduction is the log population-weighted average travel time reduction to other parishes from the railway network; distance to Guildhall is the log straight-line distance from the centroid of each parish to the Guildhall in the center of the City of London; in Column (2), Central London is defined as the City of London; in Columns (3)-(4), Central London is defined as parishes with centroids within 5 kilometers of the Guildhall; standard errors are clustered on boroughs; * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

TABLE I.2:
Event-Study Treatment Effects for the Arrival of an Overground or Underground Railway Station in Greater London Parishes from 1801-1901 (Standard Errors Clustered on Parish)

	(1)	(2)	(3)	(4)
	$\log R_{jt}$	$\log R_{jt}$	$\log R_{jt}$	$\log R_{jt}$
$\beta_{\tau=60}$	1.438***	1.654***	1.650***	0.568*
	(0.297)	(0.295)	(0.314)	(0.328)
$\beta_{\tau=50}$	1.154***	1.335***	1.361***	0.524*
	(0.247)	(0.248)	(0.264)	(0.277)
$\beta_{\tau=40}$	0.758***	0.928***	0.989***	0.362
	(0.195)	(0.196)	(0.209)	(0.219)
$\beta_{\tau=30}$	0.511***	0.676***	0.755***	0.300*
	(0.145)	(0.147)	(0.160)	(0.165)
$\beta_{\tau=20}$	0.273***	0.378***	0.380***	0.136
	(0.096)	(0.096)	(0.106)	(0.108)
$\beta_{\tau=10}$	0.104**	0.149***	0.152***	0.052
	(0.047)	(0.047)	(0.053)	(0.055)
$\beta_{\tau=0}$	–	–	–	–
$\beta_{\tau=-10}$	-0.048	-0.065	-0.039	0.051
	(0.041)	(0.042)	(0.047)	(0.048)
$\beta_{\tau=-20}$	-0.085	-0.109	-0.057	0.128
	(0.082)	(0.080)	(0.091)	(0.091)
$\beta_{\tau=-30}$	-0.092	-0.129	-0.028	0.201
	(0.119)	(0.117)	(0.136)	(0.134)
$\beta_{\tau=-40}$	-0.109	-0.150	-0.038	0.211
	(0.152)	(0.147)	(0.169)	(0.169)
$\beta_{\tau=-50}$	-0.084	-0.113	0.010	0.259
	(0.182)	(0.178)	(0.198)	(0.198)
$\beta_{\tau=-60}$	-0.093	-0.100	0.037	0.248
	(0.217)	(0.217)	(0.233)	(0.234)
$\gamma_{\tau=30}$	–	-1.501***	-1.002***	-0.465***
		(0.227)	(0.184)	(0.163)
$\gamma_{\tau=20}$	–	-0.833***	-0.431***	-0.126
		(0.238)	(0.137)	(0.126)
$\gamma_{\tau=10}$	–	-0.299***	-0.173***	-0.012
		(0.098)	(0.060)	(0.059)
$\gamma_{\tau=0}$	–	–	–	–
$\gamma_{\tau=-10}$	–	0.045	-0.025	-0.085*
		(0.063)	(0.047)	(0.044)
$\gamma_{\tau=-20}$	–	-0.003	-0.088	-0.246***
		(0.153)	(0.093)	(0.090)
$\gamma_{\tau=-30}$	–	0.033	-0.203	-0.406***
		(0.259)	(0.159)	(0.139)
Travel Time Reduction	–	–	–	-2.257***
				(0.251)
Travel Time Reduction $\times$ Distance to Guildhall	–	–	–	0.720***
				(0.086)
Central London definition		City of London	< 5 km	< 5 km
Observations	3113	3113	3113	3113
R-squared	0.981	0.981	0.981	0.983

Note: Observations are parishes and census years; all regressions include parish fixed effects, census-year dummies, and parish-specific census-year trends; railway treatment is defined based on whether a parish has an overground or underground railway station; τ is a treatment year indicator, which equals census year minus the last census year in which a parish had no railway, so that positive values of τ correspond to post-treatment years. For example, if the railway arrives in a parish in 1836, census year 1831 corresponds to $\tau = 0$, census year 1841 corresponds to $\tau = 10$ and census year 1821 corresponds to $\tau = -10$; $\tau = 0$ is the excluded category; β_τ is the rail treatment for treatment year τ; γ_τ allows the rail treatment for treatment year τ to differ between the central and outlying parts of Greater London; travel time reduction is the log population-weighted average travel time reduction to other parishes from the railway network; distance to Guildhall is the log straight-line distance from the centroid of each parish to the Guildhall in the center of the City of London; in Column (2), Central London is defined as the City of London; in Columns (3)-(4), Central London is defined as parishes with centroids within 5 kilometers of the Guildhall; standard errors are clustered on parishes; * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

TABLE I.3:
Event-Study Treatment Effects for the Arrival of an Overground or Underground Railway Station in Greater London Parishes from 1801-1901 (Constant Composition of Treated Parishes)

	(1)	(2)	(3)	(4)
	$\log R_{jt}$	$\log R_{jt}$	$\log R_{jt}$	$\log R_{jt}$
$\beta_{\tau=30}$	0.483***	0.592***	0.733***	0.492***
	(0.104)	(0.186)	(0.175)	(0.096)
$\beta_{\tau=20}$	0.265***	0.347***	0.419***	0.294***
	(0.062)	(0.114)	(0.109)	(0.071)
$\beta_{\tau=10}$	0.085***	0.125**	0.153***	0.120***
	(0.031)	(0.053)	(0.050)	(0.038)
$\beta_{\tau=0}$	–	–	–	–
$\beta_{\tau=-10}$	-0.022*	-0.030*	-0.031*	-0.004
	(0.013)	(0.016)	(0.016)	(0.016)
$\beta_{\tau=-20}$	-0.018*	-0.025**	-0.024*	-0.001
	(0.010)	(0.011)	(0.012)	(0.014)
$\gamma_{\tau=30}$	–	-1.433***	-1.030***	-0.716***
		(0.111)	(0.237)	(0.167)
$\gamma_{\tau=20}$	–	-1.046***	-0.617***	-0.402***
		(0.082)	(0.182)	(0.145)
$\gamma_{\tau=10}$	–	-0.495***	-0.266***	-0.155**
		(0.057)	(0.077)	(0.067)
$\gamma_{\tau=0}$	–	–	–	–
$\gamma_{\tau=-10}$	–	0.085**	0.028	0.018
		(0.039)	(0.022)	(0.025)
$\gamma_{\tau=-20}$	–	0.059*	0.017	0.006
		(0.032)	(0.017)	(0.019)
Travel Time Reduction	–	–	–	-2.168***
				(0.221)
Travel Time Reduction $\times$ Distance to Guildhall	–	–	–	0.660***
				(0.070)
Central London definition		City of London	<5 km	<5 km
Observations	2433	2433	2433	2433
R-squared	0.980	0.980	0.980	0.982

Note: Observations are parishes and census years; constant composition sample includes (i) untreated parishes that never receive a railway station during our sample period; (ii) treated parishes that have thirty or more years of data before the arrival of a railway and thirty or more years of data after the arrival of the railway; we drop any observations for these treated parishes with more than thirty years before the arrival of a railway and more than thirty years after the arrival of a railway, in order to ensure that all treated parishes have the same set of years before and after treatment (10-30 years); all regressions include parish fixed effects, census-year dummies, and parish-specific census-year trends; railway treatment is defined based on whether a parish has an overground or underground railway station; τ is a treatment year indicator, which equals census year minus the last census year in which a parish had no railway, so that positive values of τ correspond to post-treatment years. For example, if the railway arrives in a parish in 1836, census year 1831 corresponds to $\tau = 0$, census year 1841 corresponds to $\tau = 10$ and census year 1821 corresponds to $\tau = -10$; the excluded category is $\tau = 0$; β_τ is the rail treatment for treatment year τ; γ_τ allows the rail treatment for treatment year τ to differ between the central and outlying parts of Greater London; in this constant composition sample, $\beta_{\tau=-30}$ and $\gamma_{\tau=-30}$ are also excluded, because they are perfectly colinear with the parish-specific time trends; this colinearity does not arise in Tables I.1 and I.2, because the number of census years before and after the treatment varies across treated parishes in those specifications; travel time reduction is the log population-weighted average travel time reduction to other parishes from the railway network; distance to Guildhall is the log straight-line distance from the centroid of each parish to the Guildhall in the center of the City of London; in Column (2), Central London is defined as the City of London; in Columns (3)-(4), Central London is defined as parishes with centroids within 5 kilometers of the Guildhall; standard errors are clustered on boroughs; * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

I3 Alternative Non-parametric Specification

In this section of the online appendix, we provide further evidence on the heterogeneity in railway treatment effects within Greater London using an alternative non-parametric specification. In contrast to Section IV.B. in the paper, we no longer impose that these estimated treatment effects are related to distance from central London. Instead, we first estimate a separate railway treatment effect for each parish, and then later examine whether these estimated treatments are in fact related to distance from central London. As in Section IV.B. of the paper, we consider a "difference-in-differences" specification, in which the first difference is across parishes, and the second difference is across time.

In a first step, we compute the relative population of parishes, by differencing the log population for each parish in each year from the mean across parishes in that year. This differencing from mean log population in each year removes any secular trend in population across all Greater London parishes over time, which allows us to control for the fact that different parishes are treated with the railway in different census years. In a second step, we compute the growth in the relative population of each parish over the thirty-year period before the arrival of the railway, where this difference over time differences out any fixed effect in the level of log relative parish population. We focus on a narrow thirty-year window to ensure a similar time interval over which population growth is computed for all parishes. We cannot compute this difference for parishes that are never treated with the railway, and hence we drop these parishes. All other parishes have at least thirty years before the arrival of the railway, because our sample begins in 1801, and the first railway in Greater London is built in 1836.

In a third step, we compute the growth in the relative population of each parish over the thirty-year period after the arrival of the railway, where this difference over time again differences out any fixed effect in the level of log relative parish population. We again focus on a narrow thirty-year window and drop any parish with less than thirty years between its treatment year and the end of our parish-level sample in 1901. In a fourth and final step, we compute the "difference-in-difference," namely the change in each parish's growth in relative population between the thirty-year periods before and after the arrival of the railway. By taking this difference between the growth rates before and after the arrival of the railway, we difference out any linear parish time trend that is common to these two periods. Therefore, we again focus on deviations from parish time trends, as in Section IV.B. in the paper.

In Figure I.2, we display these double differences in relative population growth for each parish against the straight-line distance in kilometers from its centroid to the Guildhall in the center of the City of London. We indicate parishes in the City of London by hollow red circles, while parishes in the other parts of Greater London are denoted by solid blue circles. We also show the locally-weighted linear least squares regression relationship between the two variables as the solid black line. We find a sharp non-linear relationship between the railway treatment and distance from the Guildhall. For parishes within five kilometers of the Guildhall, we find negative average estimated treatment effects (an average of -0.64 log points), particularly for those parishes inside the City of London. In contrast, for parishes beyond five kilometers from the Guildhall, we find positive average estimated treatment effects (an average of 0.16 log points), with these treatment effects becoming smaller for more peripheral parishes. These substantial differences between the two groups are statistically significant at conventional critical values.

Therefore, in this non-parametric specification that allows for heterogeneous treatment effects across parishes, we again find evidence of a systematic reorganization of economic activity. Following the arrival of the railway, we find a reduction in population growth relative to trend in parishes close to the center of Greater London, and an increase in population growth relative to trend in parishes further from the center of Greater London.

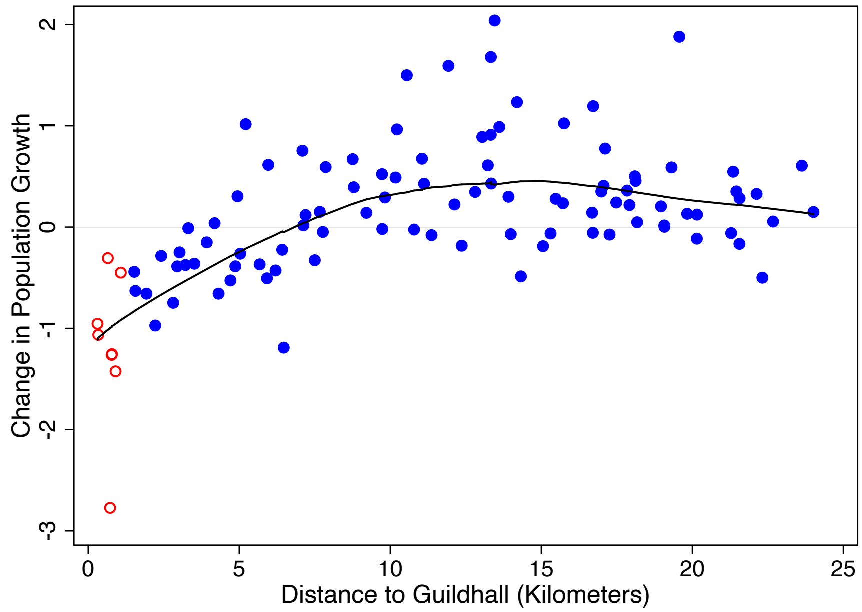

FIGURE I.2:
Non-parametric Railway Treatment Estimates for each Parish in Greater London

Note: Difference in mean population growth between the thirty-year periods before and after the arrival of a railway station in a parish; hollow red circles denote parishes in the City of London; solid blue circles denote parishes in other parts of Greater London.

I4 Worker Heterogeneity Across Occupations

In our class of quantitative urban models, we allow workers to be heterogeneous *ex post* in terms of idiosyncratic amenities (in the paper) or effective units of labor (in Section E of this online appendix), but we abstract from *ex ante* heterogeneity across workers in different occupations. We do so because the dramatic increase in travel speed from the invention of the steam railway permitted the first large-scale separation of workplace and residence for workers in all occupations. Therefore, we focus in our analysis on the overall effect of this new transport technology across workers in all occupations. Furthermore, commuting data by occupation are not available in the 1921 population census of England and Wales, which implies that we do not have any commuting data to discipline the ways in which *ex ante* heterogeneity across occupations could be introduced into the model. In this section of the online appendix, we provide two further pieces of evidence in support of the view that the invention of the steam railway enabled longer commuting distances for workers across a broad range of occupations.

In particular, we combine two different types of data from the population censuses of England and Wales. For 1921, we have borough-level data on bilateral commuting flows in Greater London. For 1851, 1881 and 1911, we have individual-level data that report residence by parish and occupation according the HISCO occupation classification from Minnesota Population Center (2018), England and Wales (1851, 1881) and Great Britain (1911). Although the individual-level data for 1921 are not yet available (because of the 100-year census confidentiality rule) and bilateral commuting data are not reported before 1921, we combine these different types of data to provide evidence that the invention of the steam railway enabled pervasive commuting.

I4.1 Commuting Intensity and Occupational Employment Specialization by Residence

In our first exercise, we examine whether commuting inflow and outflow rates across boroughs in Greater London are related to occupational employment specialization by residence. Using our 1921 bilateral commuting data, we construct a commuting outflow rate for each borough equal to the number of commuters from residence divided by total employment by residence, and a commuting inflow rate for each borough equal to the number of commuters to workplace divided by total employment by workplace. Using the 1911 individual-level census data, we construct the share of employment by residence in each occupation in the total employment by residence of each borough. We consider the following five occupation categories from the HISCO classification: (a) Administrative and Managerial, (b) Clerical and Related, (c) Production Workers, (d) Professional, (e) Sales Workers, and (f) Service Workers.

In Figure I.3, we display the commuting outflow rate against the occupational employment share by residence for each of these five occupations, where each observation corresponds to a borough, and each panel reports results for a given occupation. We find some evidence of an upward-sloping relationship for Clerical and Related occupations, and some evidence of a downward-sloping relationship for Service occupations, but both relationships are relatively noisy. Furthermore, for Professional occupations, Sales occupations and Production occupations, there is almost no relationship between commuting outflows and occupational employment specialization by residence.

In Figure I.4, we display the corresponding commuting inflow rate against the occupational employment share by residence, where each observation again corresponds to a borough, and each panel reports results for a given occupation. We find some evidence of a downward-sloping relationship for Clerical and Related occupations and Sales occupations, but again both relationships are noisy. Moreover, for Professional occupations, Administrative and Managerial occupations, Service occupations and Production occupations, there is again almost no relationship between commuting inflows and occupational employment specialization by residence.

Taking the results in Figures I.3 and I.4 together, we find that the intensity with which workers commute into or out of a borough is not strongly related to the occupation of employment of the residents of that borough. This pattern of results is in line with pervasive commuting between boroughs for workers across a broad range of occupations. These findings of extensive commuting even in relatively low-wage occupations are also consistent with the provision in Acts of Parliament from 1860 onwards of "workmen's trains," with cheap fares for working-class passengers, as ultimately reflected in the 1883 Cheap Trains Act.

I4.2 Distance of Occupational Employment by Residence from the City of London

In our second exercise, we examine the evolution over time of the distance of the residences of workers in each occupation from the City of London. Using the 1851, 1881 and 1911 individual-level data, we compute the average distance of each occupation by residence from the City of London. In particular, we first calculate the employment share of each parish in an occupation by dividing employment by residence in that occupation in that parish by total employment by residence in that occupation in Greater London. We next calculate the straight-line distance from the centroid of each parish to the Guildhall in the center of the City of London. Multiplying each parish's employment share for an occupation by its distance from the Guildhall, and summing across parishes, we obtain a measure of the residence-employment-share-weighted-average distance of that occupation from the City of London.

In Figure I.5, we display these average distances of each of the HISCO occupational categories from the City of

London in 1851, 1881 and 1911. Two features of the figure are noteworthy. First, looking across occupations for a given year, although there are differences across occupations in average distance from the City of London, they are relatively small. Second, looking across time for a given occupation, there is a broad-based increase in the average distance of all occupations from the City of London. These results come with a number of caveats, because such an increase in the average distance of residences from the City of London is what one would expect from an expansion in the geographical boundaries of Greater London, and we do not know the workplace location to which workers commuted from these residences. Nevertheless, these results highlight that the decentralization of population from Central London is not localized to workers in a few occupations. In Section I5 of this online appendix, we provide direct evidence on an increase in the average distance travelled to work in Central Philadelphia for workers across a broad range of occupations following the late-19th century improvements in transport technology.

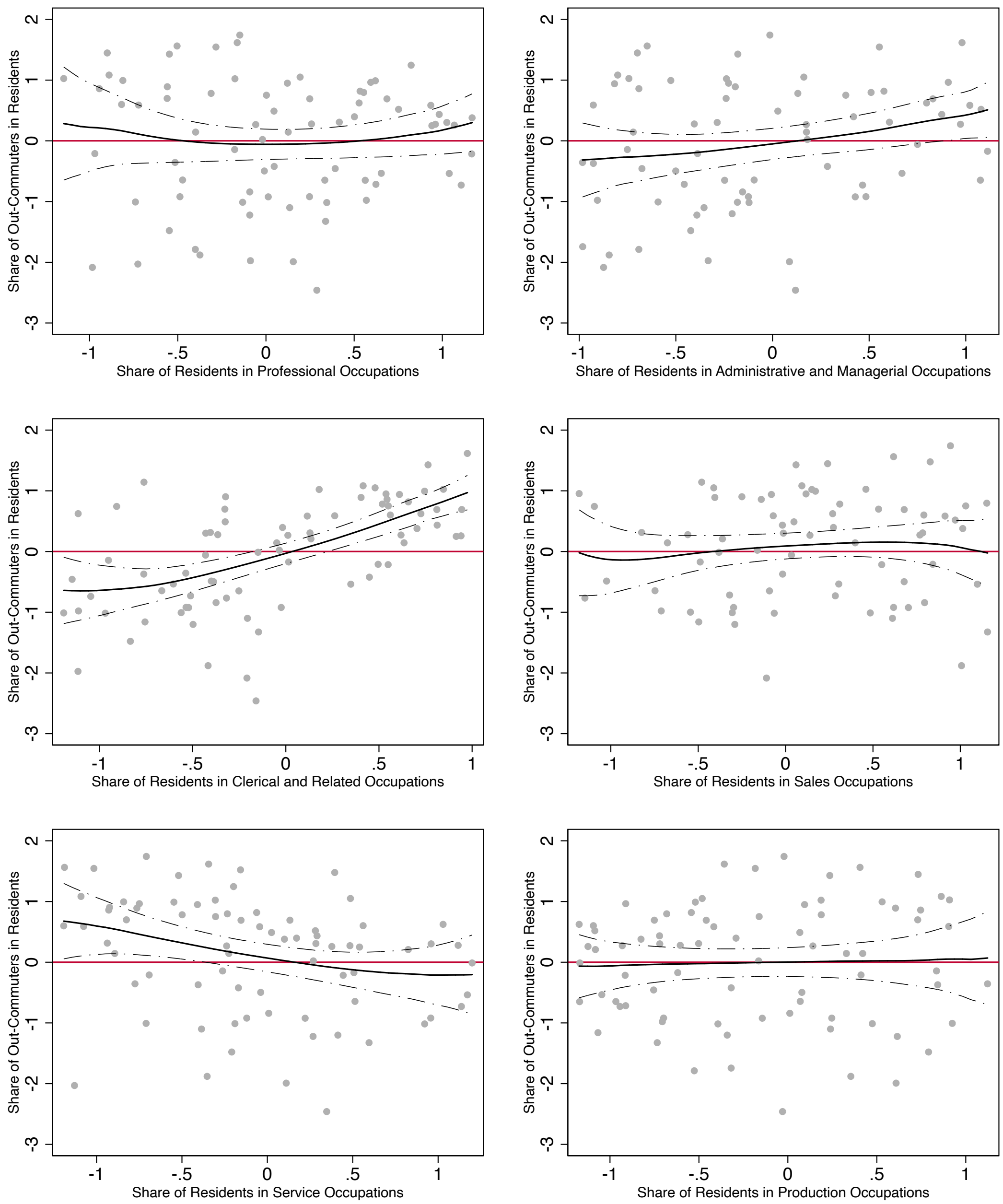

FIGURE I.3:
Commuting Outflow Rates and Occupational Employment Specialization by Residence

Note: Commuting outflow rate for each borough is defined as the number of outward commuters from residence divided by total employment by residence in 1921 for that borough. Share of residents in each occupation for each borough equals employment by residence in that occupation and borough in 1911 divided by total employment by residence in that borough in 1911. Each observation corresponds to a borough in Greater London. Solid black lines show the locally-weighted linear least squares relationship between the two variables. Black dashed lines show the 95 percent confidence intervals. Occupation definitions are from the HISCO Occupation Classification. Data source: Minnesota Population Center (2018), Great Britain (1911) and England and Wales (1921).

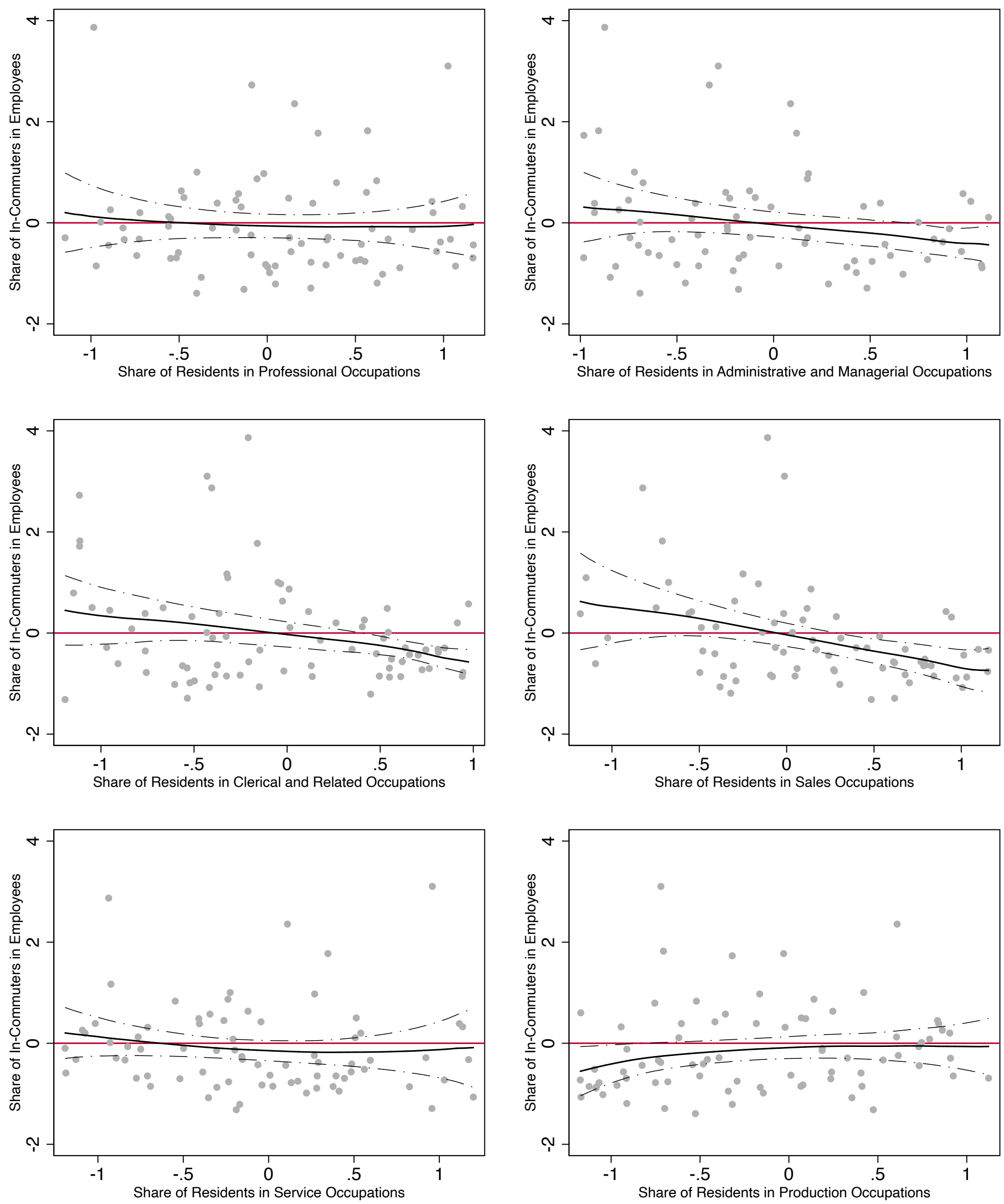

FIGURE I.4:
Commuting Inflow Rates and Occupational Employment Specialization by Residence

Note: Commuting inflow rate for each borough is defined as the number of inward commuters to workplace divided by total employment by workplace in 1921 for that borough. Share of residents in each occupation for each borough equals employment by residence in that occupation and borough in 1911 divided by total employment by residence in that borough in 1911. Each observation corresponds to a borough in Greater London. Solid black lines show the locally-weighted linear least squares relationship between the two variables. Black dashed lines show the 95 percent confidence intervals. Occupation definitions are from the HISCO Occupation Classification. Data source: Minnesota Population Center (2018), Great Britain (1911) and England and Wales (1921).

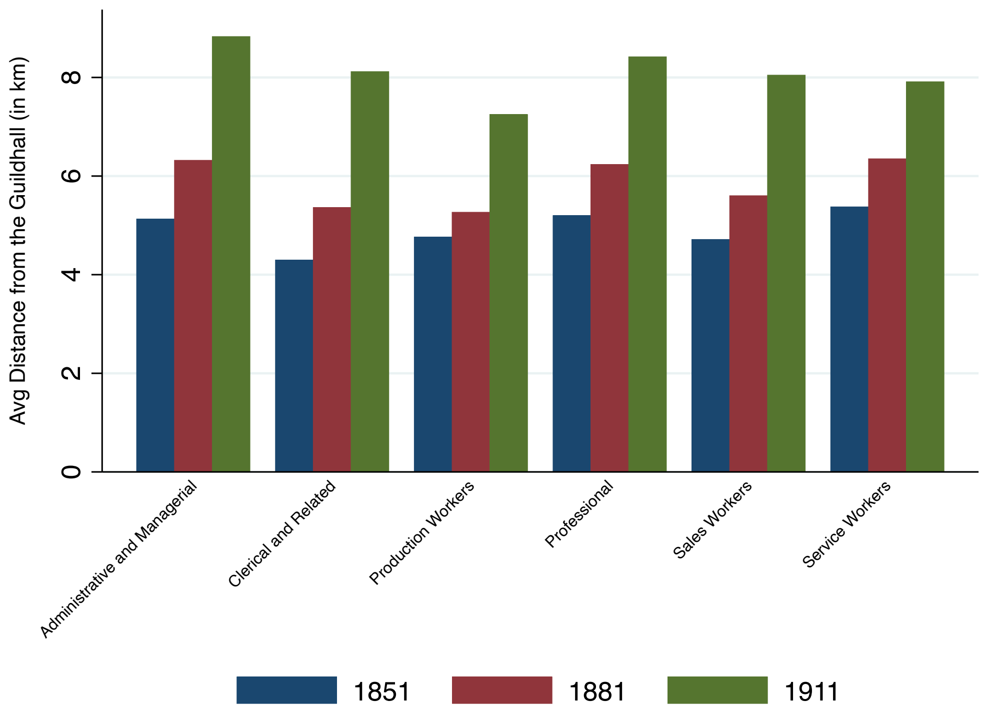

FIGURE I.5:
Residence-Employment-Share-Weighted-Average Distances from the Guildhall for each Occupation

Note: For each occupation, residence-employment-share-weighted-average distance from the Guildhall equals the sum across parishes of the share of each parish in employment by residence for an occupation in Greater London times the straight-line distance between the centroid of that parish and the Guildhall in the center of the City of London. Occupation definitions are from the HISCO Occupation Classification. Data source: Minnesota Population Center (2018), England and Wales (1851,1881) and Great Britain (1911).

I5 Reduced-Form Evidence for Other Cities

In the paper, we focus on London in our empirical analysis, because of the existence of rich historical data before and after the arrival of the railway, and the availability of bilateral commuting data for 1921. In this section of the online appendix, we show that our findings for London are representative of other large metropolitan areas following the improvement in transport technology from the invention of the steam railway during the 19th century. We undertook a major archival data collection exercise to construct data on the organization of economic activity in Paris, Berlin, Boston, Chicago, New York, and Philadelphia. In the remainder of this section, we show that we find the same reduced-form patterns for these other metropolitan areas as we find for London.

For each of the metropolitan areas, we managed to obtain population data back to the 19th century at a fine level of spatial disaggregation, corresponding to census tracts for U.S. cities and their equivalent for Berlin and Paris. Using these spatially disaggregated data, we construct total population for both the downtown area and the metropolitan area as a whole for each of these other cities. Although we have population data for all of these cities, the information on employment, commuting distances and the transport network varies somewhat across cities, depending on data availability back to the early-19th century. Further information about the data sources and definitions for each metropolitan area is contained in the data appendix in Section J11 of this online appendix. After presenting the evidence for each city, we briefly review the broader existing historical literature, which provides further evidence that our findings for London are representative of those for other metropolitan areas.

I5.1 Paris

We begin by discussing the construction of our data on population and the transport network for Paris, as discussed further in Section J11 of this online appendix. We define downtown Paris as the post-1860 *arrondissements* 1-4 (Louvre, Bourse, Temple and Hôtel-de-Ville) and the metropolitan area as the *petite courande* (which contains the *departements* of Paris, Hauts de Seine, Seine-Saint-Denis, and Val-de-Marne). To overcome the change in 1860 in the boundaries of arrondissements in central Paris, we collected population data before 1860 at the *quartier* level, where each pre-1860 arrondissement consisted of four *quartiers*. We allocated data on population for pre-1860 *quartiers* to post-1860 arrondissements using area weights. We thus obtained a consistent time-series on the population of both the downtown and metropolitan area of Paris for the period 1800-1926.

The history of the development of the public transport network in Paris has many similarities with that for London. The first railway line between Paris and Saint German-en-Laye opened in 1837. A second line to Versailles was opened in 1840 and a syndicate of five firms constructed a belt line named *Petite Ceinture Rive Droite* between 1852 and 1854. The creation of the Paris metro system lagged somewhat behind the construction of the London Underground network, with the first Parisian metro line only opening in 1900. To measure the development of the railway network over time, we collected data on the number of overground and underground railway stations in the Paris metropolitan area (*petite courande*) over the same time period of 1800-1926 as our population data.

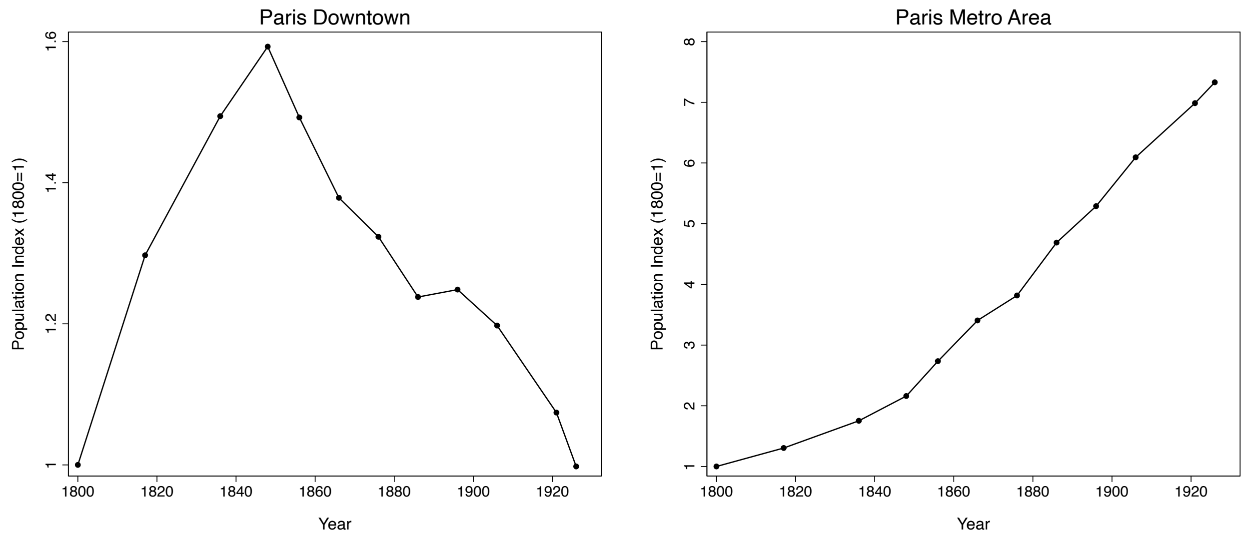

FIGURE I.6:
Population Indexes over Time for Paris (Downtown and Metro Area)

Note: Paris downtown comprises post-1860 *arrondissements* 1-4 (Louvre, Bourse, Temple and Hôtel-de-Ville). Paris metro area defined as the *petite courande* (the *departements* of Paris, Hauts de Seine, Seine-Saint-Denis, and Val-de-Marne).

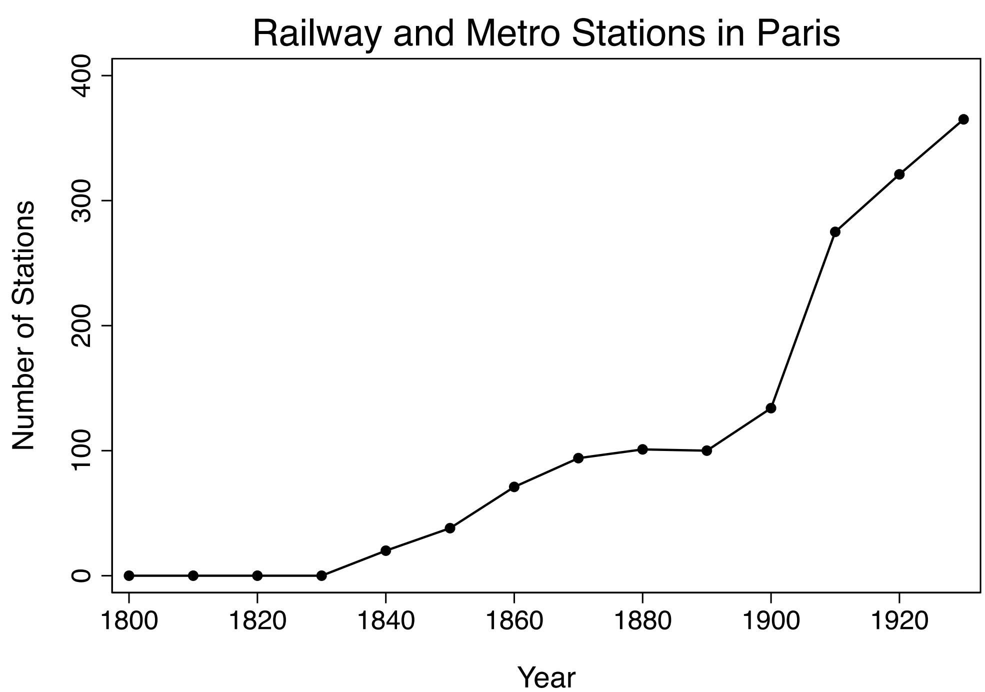

FIGURE I.7:
Overground and Underground Railway Stations in the Paris Metropolitan Area over Time

Note: Paris metro area defined as the *petite courande* (the *departements* of Paris, Hauts de Seine, Seine-Saint-Denis, and Val-de-Marne).

In Figure I.6, we display total population over time for downtown Paris (left panel) and the metropolitan area of Paris (right panel). In each case, population is expressed as an index relative to its value in 1800 (such that 1800=1). Consistent with our results for London in Figure II in the paper, we find rapid population growth for the metropolitan area of Paris throughout the 19th century. Total population increases more than sevenfold from 631,585 in 1800 to 4,628,700 in 1926. In contrast, the total population of downtown Paris first rises from 255,985 in 1800 to 407,750 in 1848, before declining sharply from that point onwards to 255,411 in 1926. As shown in Figure I.7, this decline in population in downtown Paris in the second half of the 19th century coincides closely with the expansion of the overground and underground railway network.

Therefore, our findings for Paris confirm those for London in the paper. Following the expansion of the railway network in the late-19th century, we find a population decline in the downtown combined with a population expansion for the metropolitan area as a whole.

I5.2 Berlin

An advantage of considering Berlin is that we have data on both employment by workplace and population by residence, as we have for the City of London in the paper. As discussed further in Section J11 of this online appendix, we define downtown Berlin as the six most central wards ("Stadtteile") of Berlin, which are Berlin, Köln, Friedrichswerder, Neukölln, Dorotheenstadt and Friedrichstadt, and define the metropolitan area as Greater Berlin ("Gross Berlin" as defined in the "Gross Berlin Gesetz" of 1920). Data on employment by workplace are not available at as fine a level of spatial disaggregation as our definition of downtown Berlin, but are reported from 1875 for the administrative city of Berlin in its pre-1920 boundaries ("Stadt Berlin"). The pre-1920 administrative city includes downtown Berlin and its surrounding inner suburbs and corresponds to the post-1920 districts Mitte, Tiergarten, Wedding, Prenzlauer Berg, Friedrichshain and Kreuzberg. We thus obtained two consistent time-series, one on the population of both the downtown and metropolitan area of Berlin for 1803-1925, and the other on population and employment by workplace for the pre-1920 administrative city of Berlin for 1875-1925.

The development of the public transport network in Berlin follows a similar trajectory to those for London and Paris. The first railway line between Berlin and Potsdam opened in 1838. Six radial lines were subsequently developed connecting Berlin with Koethen (1841), Stettin (1842-43), Frankfurt (Oder)-Breslau (1842-46), Hamburg (1846) and Magdeburg (1846). A major improvement in the railway network came with the completion of the circular Ringbahn in 1877, which followed the route of the former city wall, and which subsequently developed into the S-Bahn (the network of suburban commuter railways in Berlin). A further improvement occurred with the construction of the first elevated railway in Berlin between Stralauer Tor and Potsdamer Platz in 1902, which subsequently developed into the U-Bahn (the network of elevated and underground railways in Berlin). To measure the development of the railway network over time, we collected data on the number of overground, underground and elevated railway stations in the Berlin metropolitan area from the opening of the first railway line in 1838 until 1925.

In Figure I.8, we display total population over time for downtown Berlin (left panel) and the metropolitan area of Berlin (right panel). In each case, population is expressed as an index relative to its value in 1803 (such that 1803=1). Consistent with our results for London in Figure II in the paper, we find rapid population growth for the metropolitan area of Berlin throughout the 19th century. Total population increases by around a factor of 20 from 202,848 in 1803 to 4,024,165 in 1926. In contrast, the total population of downtown Berlin rises from 94,811 in 1803 to 178,021 in 1861, before declining sharply thereafter to 56,346 in 1925.

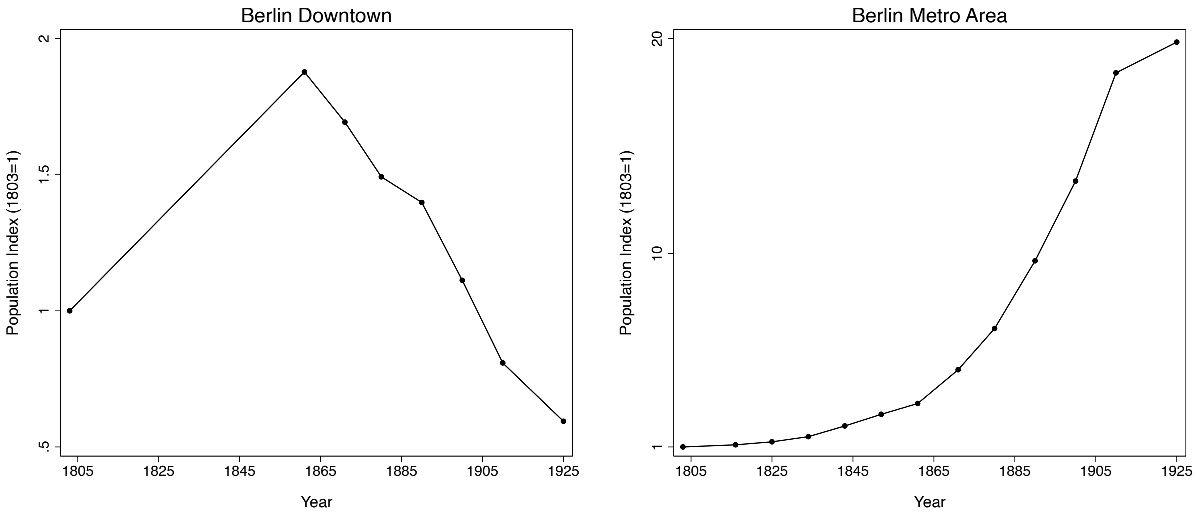

FIGURE I.8:
Population Indexes over Time for Berlin (Downtown and Metro Area)

Note: Berlin Downtown includes the six most central wards of the district "Mitte" (Berlin, Köln, Friedrichswerder, Neukölln, Dorotheenstadt and Friedrichstadt). The Berlin metro area is defined as "Gross Berlin."

FIGURE I.9:
Population by Residence and Employment by Workplace over Time (Pre-1920 Administrative City of Berlin)

Note: Data is for the pre-1920 administrative city of Berlin ("Stadt Berlin"), which includes the most central parts of the metropolitan area including downtown Berlin and its surrounding inner suburbs. It encompasses the post-1920 districts of Mitte, Tiergarten, Wedding, Prenzlauer Berg, Friedrichshain and Kreuzberg.

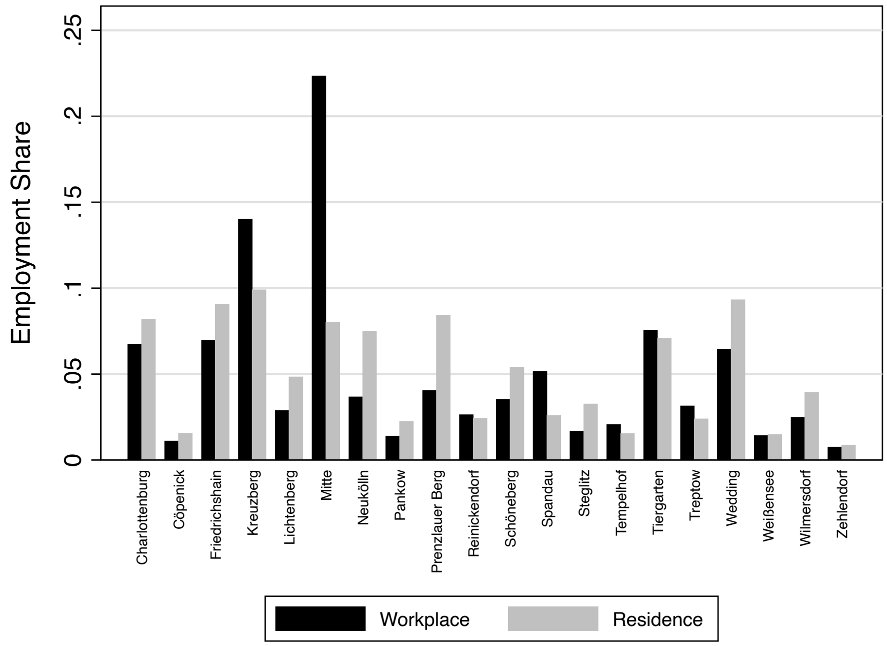

FIGURE I.10:
Private Employment by Workplace and Residence for Each District in the Berlin Metro Area in 1925

Note: The Berlin metro area is defined as Greater Berlin ("Gross Berlin"). Our earlier definition of downtown Berlin includes the six most central wards of the district "Mitte." Our earlier definition of the pre-1920 administrative city of ("Stadt Berlin") includes the central districts of Mitte, Tiergarten, Wedding, Prenzlauer Berg, Friedrichshain and Kreuzberg.

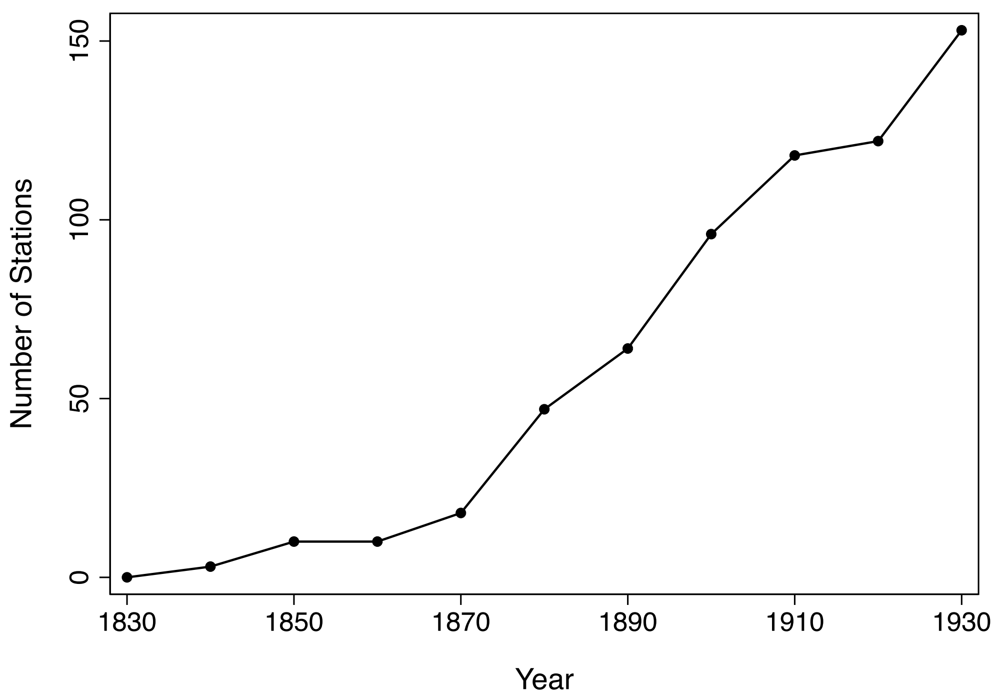

FIGURE I.11:
Overground, Underground and Elevated Railway Stations in the Berlin Metro Area

Note: The Berlin metro area is defined as Greater Berlin ("Gross Berlin").

In Figure I.9, we display total population by residence and total employment by workplace for the pre-1920 administrative city of Berlin, where each variable is expressed as an an index relative to its value in 1875 (such that 1875=1). Although the pre-1920 administrative city of Berlin encompasses a much larger area than the City of London in the paper, and includes much of the inner residential suburbs of Berlin, we again find evidence of a change in patterns of specialization from a residence to a workplace in the central city.[8] Throughout the late-19th century, employment by workplace grows substantially more rapidly than population by residence. While employment by workplace more than quadruples between 1875 and 1925, population by residence approximately doubles over the same period. From 1907 onwards, population by residence declines by around 10,000, while employment by workplace continues to grow rapidly.

In Figure I.10, we show private employment by workplace and employment by residence in 1925 for each of the districts of the Berlin metropolitan area. We find a similar pattern for Berlin in 1925 as for London in 1921 in the paper. Employment by workplace is more geographically concentrated than employment by residence. In the same way that the City of London and the City of Westminster have much higher ratios of employment to residents in 1921 than any other borough in the London metropolitan area, Mitte and Kreuzberg have much higher ratios of employment to residents in 1925 than any other district in the Berlin metropolitan area.

As shown in Figure I.11, the decline in population in downtown Berlin in Figure I.8, and the change in patterns of specialization from a residence to a workplace in the pre-1920 administrative city of Berlin in Figure I.9, coincide closely with the expansion of the railway network in the second half of the 19th century. Therefore, our findings for population, employment and the transport network for Berlin again confirm those for London in the paper.

I5.3 Boston

We begin by discussing the construction of our population data for Boston. Our definition of downtown Boston follows the U.S. Census definition, which includes Chinatown and the Leather District, and includes the four 2010 census tracts listed in Section J11 of this online appendix. We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. Our definition of the metropolitan area of Boston follows the U.S. Census definition of the Boston-Cambridge-Newton, MA-NH Metro Area, and includes the counties listed in Section J11 of this online appendix. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the Boston metro area.

In Figure I.12, we display total population over time for downtown Boston (left panel) and the metropolitan area of Boston (right panel). Consistent with our results for London in Figure II in the paper, we find rapid population growth for the metropolitan area as a whole is combined with a decline in the downtown population. In Figure I.13, we show the average distance travelled to work (in miles) for attorneys employed in Boston, using data reported in Jackson (1987). Although attorneys are only one occupation, we find an increase in average distance travelled to work by attorneys in the opening decades of the 20th century, which is consistent with the idea that the decline in downtown population from 1880-1940 reflects an increase in commuting.

[8] As a point of comparison, in 1880 our definition of downtown Berlin has a population of 141,493, while the pre-1920 administrative city of Berlin has a population of 1,123,749 and Greater Berlin a population of 1,321,000.

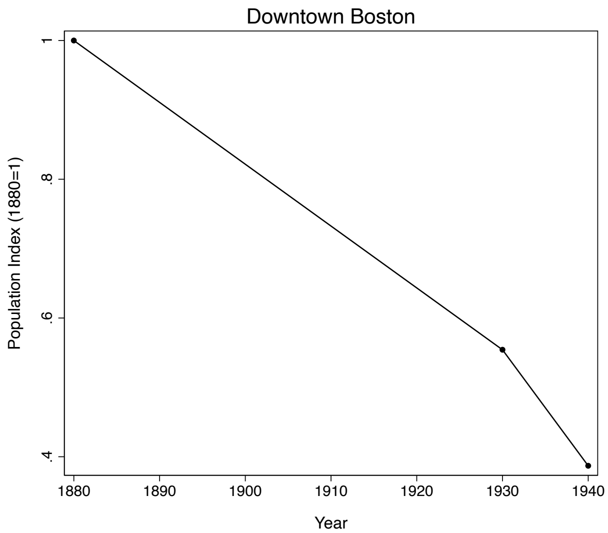

FIGURE I.12:
Population Indexes over Time for Boston (Downtown and Metro Area)

Note: Boston Downtown includes the four central 2010 census tracts in Downtown, Chinatown and the Leather District. Boston metro area is the Boston-Cambridge-Newton, MA-NH Metro Area.

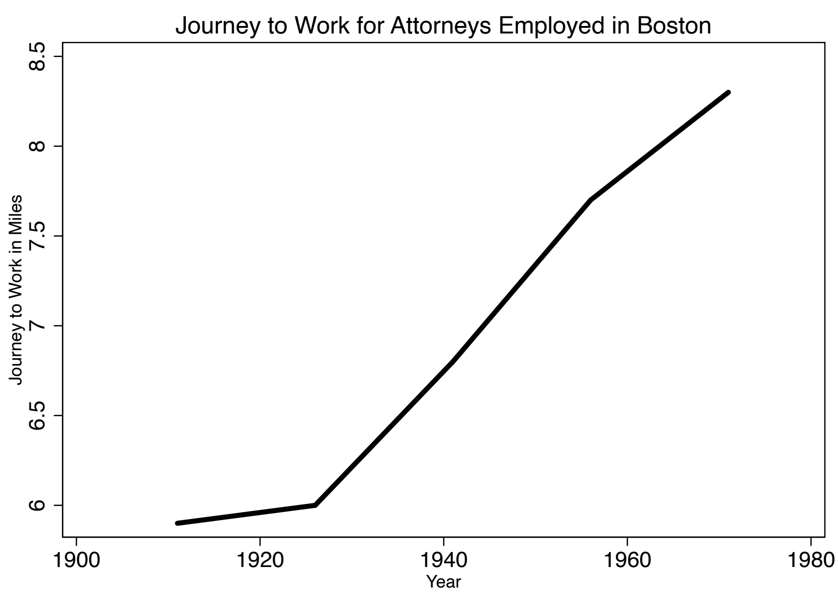

FIGURE I.13:
Journey to Work (in Miles) for Attorneys Employed in Boston

Note: The figure shows the average length of the journey-to-work for attorneys with offices in Boston at five-year intervals between 1911 and 1971, as reported in Jackson (1987). The sample of attorneys for each year was chosen by Jackson (1987) by taking every tenth attorney in the Boston City Directory for that year until a total of 76-78 was reached.

Therefore, our findings for Boston again confirm those for London in the paper. In the late-19th and early-20th centuries, we find a population decline in the downtown combined with a population expansion for the metropolitan area as a whole, and an increase in commuting distances.

I5.4 Chicago

We begin by discussing the construction of our population data for Chicago. We define downtown Chicago as "Chicago Loop," which includes the five central 2010 census tracts listed in Section J11 of this online appendix. We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. Our definition of the metropolitan area of Chicago follows the U.S. Census definition of the Chicago-Naperville-Elgin, IL-IN-WI Metro Area, and includes the counties listed in Section J11 of this online appendix. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the Chicago metro area. In Figure I.14, we display total population over time for downtown Chicago (left panel) and the metropolitan area of Chicago (right panel). Consistent with our results for London in Figure II in the paper, we find rapid population growth for the metropolitan area as a whole is combined with a decline in the downtown population.

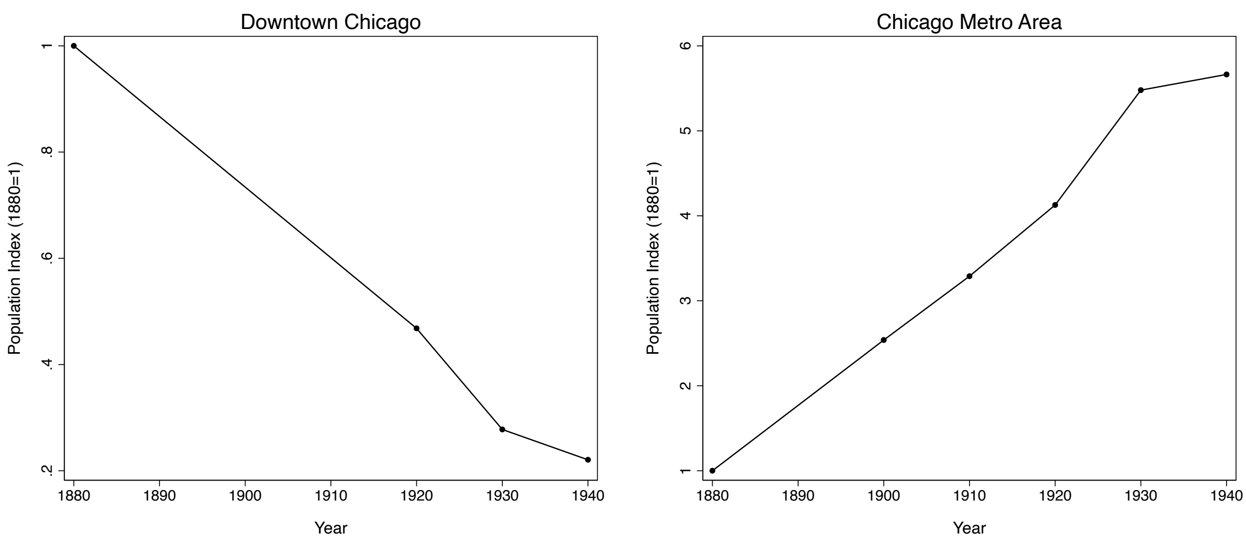

FIGURE I.14:
Population Indexes over Time for Chicago (Downtown and Metro Area)

Note: Chicago Downtown includes the five 2010 census tracts in Chicago Loop. Chicago metro area is the Chicago-Naperville-Elgin, IL-IN-WI Metro Area.

I5.5 New York

We begin by discussing the construction of our population data for New York. We define downtown New York as the parts of Lower Manhattan with the longest histories of European settlement, including all 2010 census tracts South of a line following Canal Street, as listed in Section J11 of this online appendix. We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. Our definition of the New York metropolitan area follows the U.S. Census definition of the New York-Newark-Jersey City, NY-NJ-PA Metro Area, and includes the counties listed in Section J11 in the data appendix below. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the New York metro area.

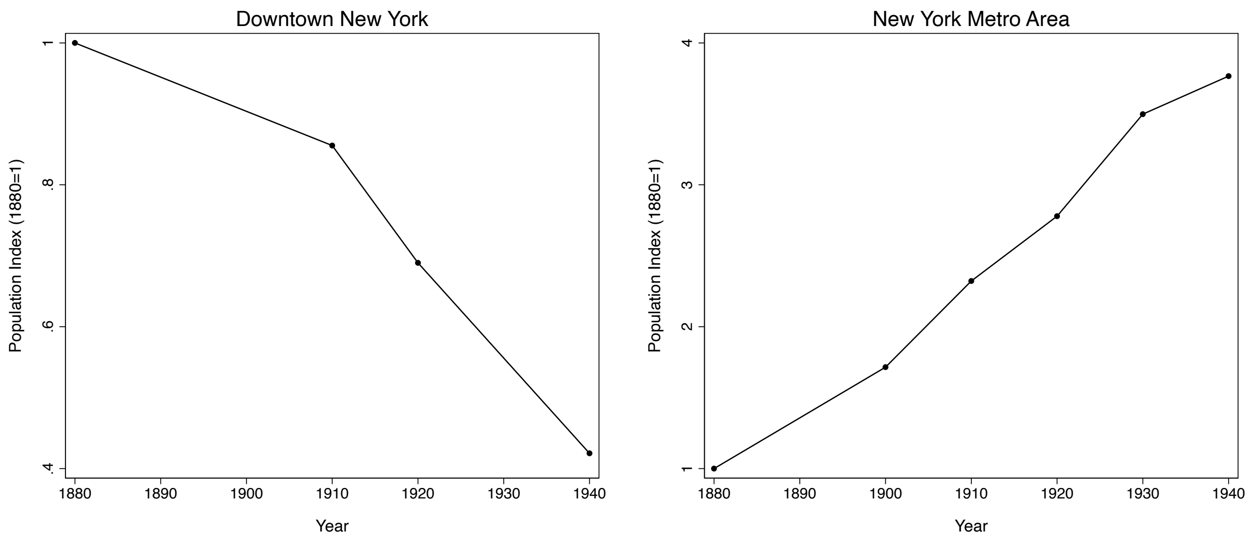

FIGURE I.15:
Population Indexes over Time for New York (Downtown and Metro Area)

Note: New York downtown corresponds to all 2010 census tracts south of the line of Canal Street. New York metro area is the New York-Newark-Jersey City, NY-NJ-PA Metro Area.

In Figure I.15, we display total population over time for downtown New York (left panel) and the metropolitan area of New York (right panel). Consistent with our results for London in Figure II in the paper, we again find rapid population growth for the metropolitan area as a whole is combined with a decline in the downtown population. In Figure I.16, we show the average distance travelled to work (in miles) for attorneys employed in Manhattan, using data reported in Jackson (1987). Although attorneys are only one occupation, we find an increase in average distance travelled to work in the late-19th century and opening decades of the 20th century, which is consistent with the idea that the decline in downtown population from 1880-1940 reflects an increase in commuting.

Therefore, our findings for New York again confirm those for London in the paper. In the late-19th and early-20th centuries, we find a population decline in the downtown combined with a population expansion for the metropolitan area as a whole, and an increase in commuting distances.

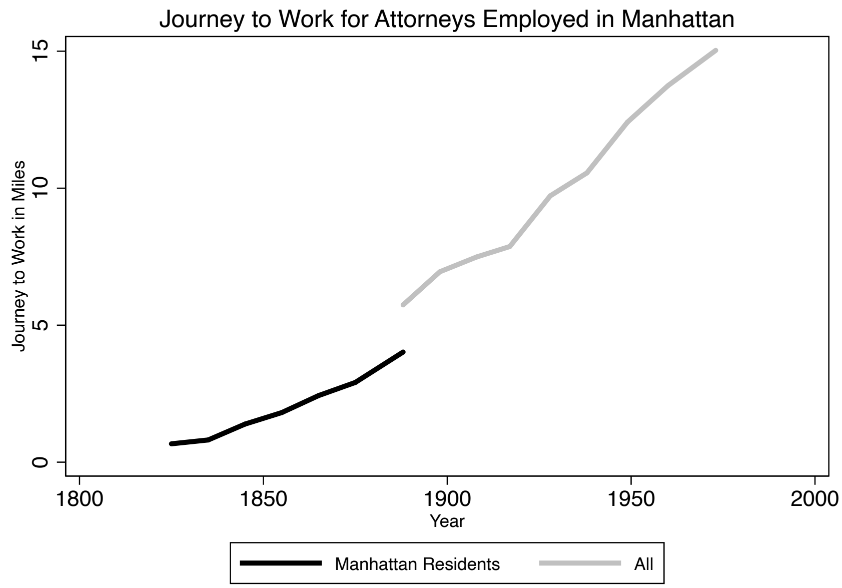

FIGURE I.16: Journey to Work (in Miles) for Attorneys Employed in Manhattan

Note: The figure shows the average length of the journey-to-work for attorneys with offices in Manhattan by decades between 1825 and 1973, as reported in Jackson (1987). The samples for each year were drawn by Jackson (1987) from City Directories for that year and the sample size varies from 88 in 1825 to 120 in 1973.

I5.6 Philadelphia

We begin by discussing the construction of our population data for Philadelphia. Our definition of downtown Philadelphia follows that of the Center City District and Central Philadelphia Development Corporation, including the 2010 census tracts listed in Section J11 of this online appendix. We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. Our definition of the metropolitan area of Philadelphia follows the U.S. Census definition of the Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area, and includes the counties listed in Section J11 of this online appendix. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the Philadelphia metro area.

In Figure I.17, we display total population over time for downtown Philadelphia (left panel) and the metropolitan area of Philadelphia (right panel). Consistent with our results for London in Figure II in the paper, we find that rapid population growth for the metropolitan area as a whole is combined with a decline in the downtown population. In Figure I.18, we show the average distance travelled to work (in miles) for workers in a number of different occupations in central Philadelphia, as reported in Hershberg (1981). Consistent with the idea that the decline in downtown population from 1880-1940 reflects an increase in commuting, we find an increase in average distance travelled to work across this broad range of occupations in the immediately preceding period from 1850-1880.

Therefore, our findings for Philadelphia again confirm those for London in the paper. In the late-19th and early-20th centuries, we find a population decline in the downtown combined with a population expansion for the metropolitan area as a whole, and an increase in commuting distances.

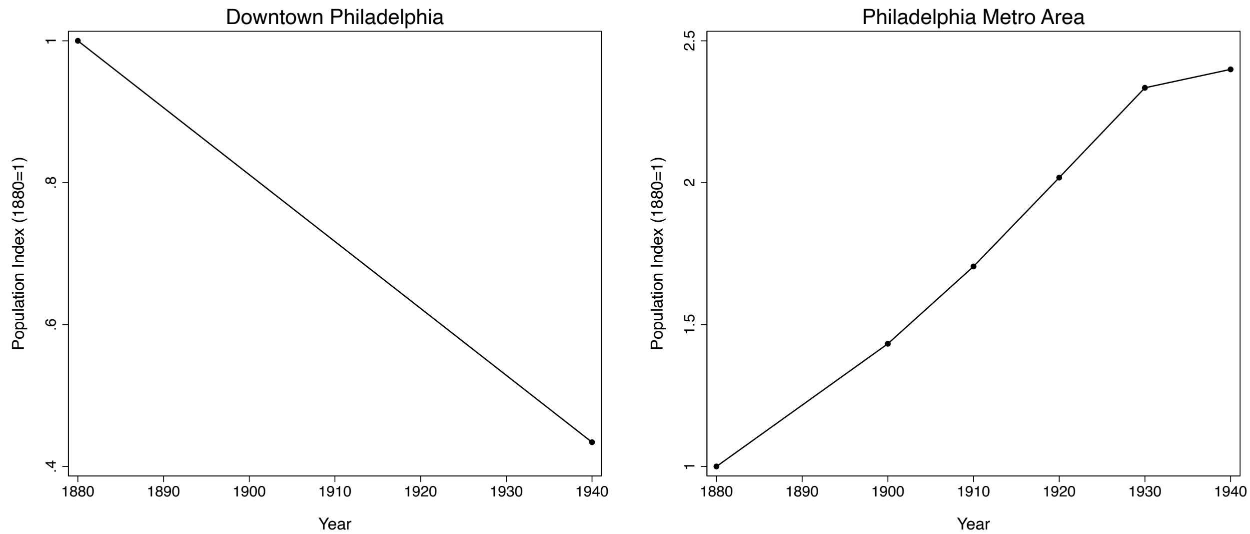

FIGURE I.17: Population Indexes over Time for Philadelphia (Downtown and Metro Area)

Note: Philadelphia downtown based on the central 2010 census tracts reported by the Center City District and Central Philadelphia Development Corporation. Philadelphia metro area is the Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area.

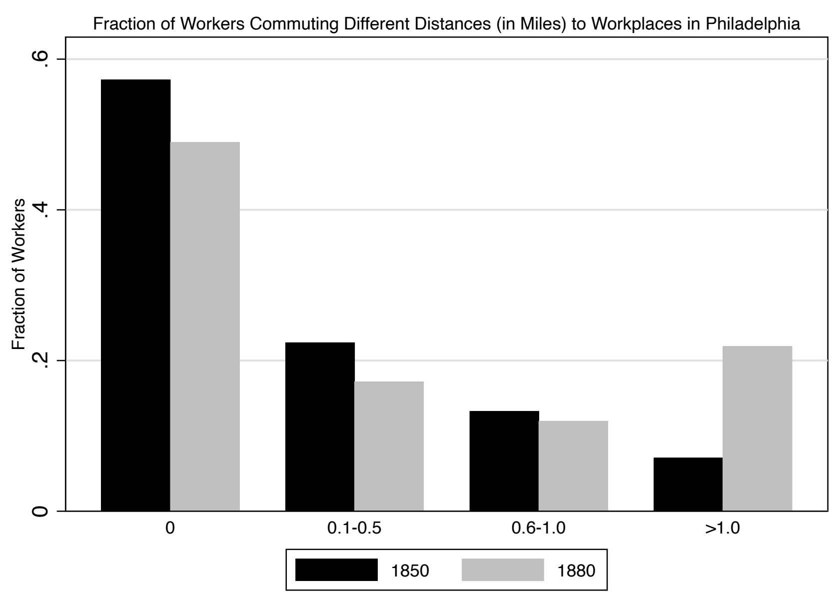

FIGURE I.18: Journey (in Miles) to Workplaces in Philadelphia

Note: Sample of occupations includes Blacksmiths, Bookbinders, Cabinetmakers, Carpenters, Confectioners, Lawyers, and Physicians. The data were compiled by the Philadelphia Social History Project (PSHP) from the manuscript schedules of the U.S. Census of Population, the manuscript schedules of the U.S. Census of Manufacturing, City Business Directories, and City Street Directories. Source: Hershberg (1981).

I5.7 Historical Literature

The reduced-form evidence for London in the paper and for the other metropolitan areas in the preceding sections of this online appendix is consistent with a broader historical literature that has discussed the decline in central city population following the transport improvements during the late-19th century, including Leyden (1933), Warner (1978), Hershberg (1981), Jackson (1987) and Fogelson (2003). This historical literature emphasizes the central mechanism in our model of a change in specialization of the central city from a residence to a workplace, with the resulting competition for floor space leading to a higher price for land. An example is the following quotation for New York: “Already the nation’s largest city in 1830, New York grew phenomenally over the next forty years. Its population soared from under 250,000 to nearly 1.5 million, and its economy expanded at a rate that amazed contemporaries. Together with the huge influx of immigrants, what a special New York State Senate commission called ‘the inexorable demands of business’ transformed the structure of the city, turning lower Manhattan mainly into stores, offices, workshops, and warehouses and upper Manhattan largely into residences. As early as 1836 Hone, who then lived on lower Broadway, feared he would be forced to move uptown. ‘Almost everybody downtown in the same predicament’ he wrote ’for all the dwelling houses are to be converted into stores. We are tempted with prices so exorbitantly high that none can resist.’ . . . ” (Fogelson 2003, page 10).

I6 Additional Gravity Equation Results

In this section of the online appendix, we report further robustness tests for the commuting gravity equation estimation in Section VI.D. of the paper. In Column (1) of Table I.4, we begin by reproducing our baseline instrumental variables specification from Column (2) of Table I in the paper. This baseline instrumental variables specification has a local average treatment effect (LATE) interpretation, in which the elasticity of commuting flows with respect to travel time in the second-stage regression is identified from the variation in travel time induced by straight-line distance after conditioning on the workplace and residence fixed effects. To examine potential heterogeneity in average treatment effects over short versus long straight-line distances, we re-estimated this baseline specification for subsamples of observations with below and above median straight-line distances in Columns (2) and (3) of Table I.4 respectively. Comparing the results for the full sample and these two subsamples, we find a somewhat larger local average treatment effect (LATE) for below-median straight-line distances than for above-median straight-line distances. Nonetheless, these differences are relatively small compared to the estimated elasticity of commuting flows with respect to travel times of around five.

In Column (4) of Table I.4, we report another robustness test on our baseline instrumental variables specification that examines the idea that the first-stage relationship between travel times and straight-line distance could be non-linear in logs. In particular, conditional on there existing a railway connection between a pair of locations, railways reduce travel times by more over longer straight-line distances, as reflected in the elasticity of travel times with respect to straight-line distance as less than one in our first-stage regression. However, there are likely to be fewer railway connections over longer straight-line distances, which implies that this relationship in the first-stage regression could be convex, with higher elasticities of travel times with respect to straight-line distance over longer straight-line distances.[9] To explore this idea, Column (4) augments our baseline specification in Column (1) by including both log straight-line distance and the square of log straight-line distance as instruments. As reported in the bottom panel of the table, both instruments are individually statistically significant at conventional critical values, and the F-statistic for their joint significance is well above the conventional threshold of 10. As also shown in the bottom panel, we find a positive estimated coefficient on the square of log straight-line distance, consistent with a convex relationship. As this specification is now overidentified, we also report the results of the Hansen-Sargan overidentification test in the top panel, and are unable to reject the null hypothesis of the model’s overidentifying restrictions.

[9]This elasticity of travel times to straight-line distances converges to one as the number of transport connections approaches zero.

In the remaining columns of the table, we further probe our identifying assumption. We begin by exploring the concern that there could be unobserved factors that affect bilateral commuting flows in Central London relative to other parts of Greater London. To address this concern, we define nine bilateral types of flows within Greater London based on whether the origin or destination is in the following three areas: (i) Central London (the City of London, the City of Westminster and Holborn); (ii) The rest of the County of London; (iii) the rest of Greater London. In Column (5), we augment our baseline specification in Column (1) by including fixed effects for these nine bilateral types of flows. Even relying solely on variation in straight-line distance within each type of flow to identify the estimated coefficient on bilateral travel times, we continue to find a similar pattern of results. Both instruments remain highly significant and we again pass the Hansen-Sargan overidentification test. Finally, we examine the concern that there could be unobserved factors that affect bilateral commuting flows over very long distances relative to those over very short distances. To address this concern, Column (6) further augments the specification in Column (5) with fixed effects for quintiles of straight-line distance. Therefore, in this final specification, we identify the estimated coefficient on bilateral travel times solely from variation in straight-line distance within these quintiles and within our nine bilateral types of flows. Using this more limited variation, we find that the instruments are individually less significant, although the first-stage F-statistic for their joint significance remains well above the conventional threshold of 10, and we continue to pass the Hansen-Sargan overidentification test. Finally, instead of assuming prohibitive commuting costs ($\kappa_{nit} \rightarrow \infty$) for all pairs of boroughs with zero commuting flows, we estimate the commuting gravity equation including the zeros and using the Poisson Pseudo Maximum Likelihood estimator of Santos Silva and Tenreyro (2006). Again we find a similar pattern of results with a somewhat large estimated elasticity of commuting flows to travel times of -6.298 (standard error 0.079), which if anything would imply an even larger impact of the removal of the railway network on the organization of economic activity in our baseline quantitative analysis.

As a final specification check on our estimates of commuting costs, Figure I.19 shows the conditional correlation between the log commuting probabilities to other boroughs $n \neq i$ and our estimates of bilateral commuting costs, after removing workplace and residence fixed effects. Consistent with our model's predictions, we find an approximately log linear relationship between bilateral commuting flows and our estimates of bilateral commuting costs, with a conditional correlation of over 0.7. While this relationship in part reflects the strong correlation between commuting flows and straight-line distance, it suggests that our parameterization of commuting costs provides a reasonable approximation to the data. We report further specification checks on our commuting cost estimates in Sections VI.E. and VI.F. of the paper, where we compare our model's predictions for the impact of the removal of the railway network on workplace employment and commuting patterns with the available historical data.

TABLE I.4:
Gravity Estimation Using 1921 Bilateral Commuting Data

	(1)	(2)	(3)	(4)	(5)	(6)
Second-stage Regression						
	$\log \lambda_{nit}$	$\log \lambda_{nit}$	$\log \lambda_{nit}$	$\log \lambda_{nit}$	$\log \lambda_{nit}$	$\log \lambda_{nit}$
$\log d_{nit}^{W}$	-5.203^{***}	-5.753^{***}	-4.379^{***}	-5.201^{***}	-5.281^{***}	-5.514^{***}
	(0.069)	(0.126)	(0.146)	(0.068)	(0.072)	(0.127)
Workplace fixed effects	yes	yes	yes	yes	yes	yes
Residence fixed effects	yes	yes	yes	yes	yes	yes
Zone-pair fixed effects					yes	yes
Distance-cell fixed effects						yes
Kleibergen-Paap (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Hansen-Sargen (p-value)	–	–	–	0.42	0.11	0.13
Estimation	IV	IV	IV	IV	IV	IV
Sample	Full	Below-Median Straight-line Distance	Above-Median Straight-line Distance	Full	Full	Full
Observations	3023	1510	1502	3023	3023	3023
R-squared	–	–	–	–	–	–

First-stage Regression						
	$\log d_{nit}^{W}$	$\log d_{nit}^{W}$	$\log d_{nit}^{W}$	$\log d_{nit}^{W}$	$\log d_{nit}^{W}$	$\log d_{nit}^{W}$
$\log d_{ni}^{S}$	0.429^{***}	0.412^{***}	0.482^{***}	0.256^{***}	0.242^{***}	0.366^{***}
	(0.003)	(0.005)	(0.009)	(0.034)	(0.033)	(0.059)
$\left(\log d_{ni}^{S}\right)^2$				0.018^{***}	0.020^{***}	0.005
				(0.004)	(0.003)	(0.007)
Workplace fixed effects	yes	yes	yes	yes	yes	yes
Residence fixed effects	yes	yes	yes	yes	yes	yes
Zone-pair fixed effects					yes	yes
Distance-cell fixed effects						yes
First-stage F-statistic	22,235	6,155	2,620	11,209	8,224	3,162
Sample	Full	Below-Median Straight-line Distance	Above-Median Straight-line Distance	Full	Full	Full
Observations	3023	1510	1502	3023	3023	3023
R-squared	0.949	0.937	0.903	0.950	0.951	0.951

Note: λ_{nit} is the commuting probability from equation (6); d_{nit}^{W} is our least-cost-path travel time measure with the following weights: overground railways 1 (base category 21 mph); underground railways 1.4 (21/15 mph); omnibus and tram 3.5 (21/6 mph); and walking 7 (21/3 mph); d_{ni}^{S} is straight-line distance; Kleibergen-Paap is the p-value for the Kleibergen-Paap underidentification test; Hansen-Sargen is the p-value for the Hansen-Sargan overidentification test; OLS refers to ordinary least squares; the second-stage R-squared is omitted from the instrumental variables (IV) specifications, because it does not have a meaningful interpretation; First-stage F-statistic is the F-statistic for the joint significance of the excluded exogenous variables in the first-stage regression; 11 singleton observations are dropped from the subsample of above-median straight-line distances in Column (3), which explains why the sum of the observations in the subsamples in Columns (2) and (3) is less than the number of observations for the full sample in the other columns; Heteroskedasticity robust standard errors in parentheses: * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

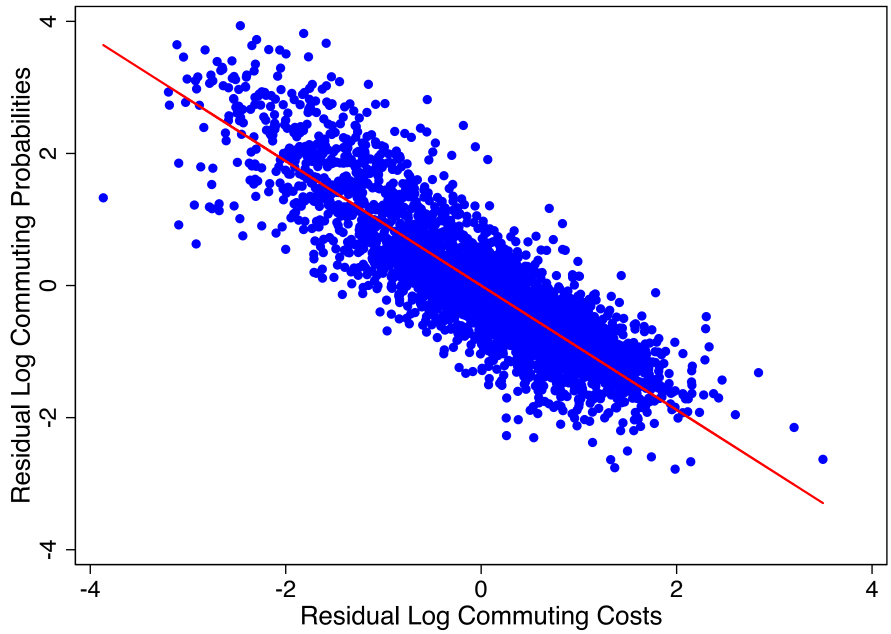

FIGURE I.19:
Conditional Correlation Between Bilateral Commuting Probabilities and Commuting Costs in 1921

Note: Conditional correlation after removing workplace and residence fixed effects between log bilateral commuting probabilities (equation (6) in the paper) and log estimated bilateral commuting costs to other boroughs in 1921.

I7 Floor Space Supply Elasticity

In this section, we discuss our calibration of the floor space supply elasticity (μ) using separate data on the construction of new buildings that were compiled by London County Council for part of our time period from 1871-1921 for the subset of our boroughs in the County of London. As a check on our assumption of a common floor space supply elasticity, we solve for the model's predictions for changes in the supply of floor space for these boroughs over this time period, and compare these predictions to the changes in the supply of floor space reported in the data.

I7.1 Poor Law Union Data

Following the Metropolis Act of 1869, London County Council reports a consistent time-series for Poor Law Unions in the County of London that separates outs changes in rateable values into the contributions of new buildings and the revaluation of existing buildings, where these revaluations occur every five years. Each borough in our data typically consists of several of these Poor Law Unions. Therefore we begin by matching Poor Law Unions to boroughs and aggregating the data to the level of the borough. In a few cases, Poor Law Union boundaries span borough boundaries, in which case we aggregate these boroughs. We thus obtain data for the following 22 boroughs and aggregations of boroughs in the County of London: (1) Bermondsey, (2) Bethnal Green, (3) Camberwell, (4) Chelsea, (5) City of London, (6) Hampstead, (7) Islington, (8) Kensington, (9) Lambeth, (10) Paddington, (11) Poplar, (12) Shoreditch, (13) Southwark, (14) St. Marylebone, (15) St. Pancras, (16) Stepney, (17) Aggregation of Battersea and Wandsworth, (18) Aggregation of Deptford, Greenwich, Lewisham and Woolwich, (19) Aggregation of Fulham and Hammersmith, (20) Aggregation of Hackney and Stoke Newington, (21) Aggregation of Holborn and Finsbury, and (22) City of

Westminster.

I7.2 Observed Rateable Values

In the data, we observe rateable values in each year and know the years in which revaluations occurred. In particular, we observe a time-series on rateable values for each pair of valuation years, $\{t, t-5\}$ and the intervening years $\{t-1, t-2, t-3, t-4\}$ for each location n:

$$\{\mathbb{Q}_{nt-5}, \mathbb{Q}_{nt-4}, \mathbb{Q}_{nt-3}, \mathbb{Q}_{nt-2}, \mathbb{Q}_{nt-1}, \mathbb{Q}_{nt}\}. \tag{I.1}$$

For intra-valuation years, changes in rateable value correspond to changes in quantities (H_{nt}) only (new buildings) valued at prices (Q_{nt-5}) in the previous valuation year $t-5$:

$$\begin{aligned} \mathbb{Q}_{nt-4} - \mathbb{Q}_{nt-5} &= (H_{nt-4} - H_{nt-5})\, Q_{nt-5}, \\ \mathbb{Q}_{nt-3} - \mathbb{Q}_{nt-4} &= (H_{nt-3} - H_{nt-4})\, Q_{nt-5}, \\ \mathbb{Q}_{nt-2} - \mathbb{Q}_{nt-3} &= (H_{nt-2} - H_{nt-3})\, Q_{nt-5}, \\ \mathbb{Q}_{nt-1} - \mathbb{Q}_{nt-2} &= (H_{nt-1} - H_{nt-2})\, Q_{nt-5}. \end{aligned} \tag{I.2}$$

Combining equations (I.1) and (I.2), we obtain the following recursive representation of rateable values in year $t-1$ immediately before revaluation as a function of rateable values in the last valuation year $t-5$ and new buildings during intra-valuation years:

$$\mathbb{Q}_{nt-1} = H_{nt-5} Q_{nt-5} + \sum_{x=1}^{4} (H_{nt-x} - H_{nt-x-1})\, Q_{nt-5}. \tag{I.3}$$

For the valuation year t, the overall change in rateable value equals the sum of the revaluation of existing buildings plus new buildings constructed in that year:

$$\mathbb{Q}_{nt} - \mathbb{Q}_{nt-1} = \underbrace{(Q_{nt} - Q_{nt-5})\, H_{nt-1}}_{\text{Revaluation Term}} + \underbrace{(H_{nt} - H_{nt-1})\, Q_{nt}}_{\text{New Buildings Term}}. \tag{I.4}$$

Therefore, the percentage change in rateable values in the valuation year t can be decomposed into a revaluation term and a new buildings term:

$$\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-1}} - 1 = \underbrace{\frac{(Q_{nt} - Q_{nt-5})\, H_{nt-1}}{Q_{nt-5} H_{nt-1}}}_{\text{Revaluation Term}} + \underbrace{\frac{(H_{nt} - H_{nt-1})\, Q_{nt}}{Q_{nt-5} H_{nt-1}}}_{\text{New Buildings Term}}. \tag{I.5}$$

I7.3 Estimating Revaluation and New Buildings

We now use this decomposition in equation (I.5) to recover changes in the price and supply of floor space for each five-year period between valuation years. We start by noting that the revaluation term on the right-hand side of equation (I.5) is equal to the percentage change in floor space between the valuation years t and $t-5$:

$$\left(\frac{Q_{nt}}{Q_{nt-5}} - 1\right) = \frac{(Q_{nt} - Q_{nt-5})\, H_{nt-1}}{Q_{nt-5} H_{nt-1}}. \tag{I.6}$$

London County Council estimates this revaluation term by subtracting an estimate of the new buildings term from the overall percentage change in rateable values in the valuation year t in equation (I.5). This estimate of the new buildings term is constructed as the average of the changes in rateable values in the years immediately before and after the valuation year t, which are entirely the result of new buildings in those years:

$$\frac{\left(H_{nt}-H_{nt-1}\right)Q_{nt}}{Q_{nt-5}H_{nt-1}}=\frac{\left(0.5\times\left(\mathbb{Q}_{nt-1}-\mathbb{Q}_{nt-2}\right)\right)+\left(0.5\times\left(\mathbb{Q}_{nt+1}-\mathbb{Q}_{nt}\right)\right)}{\mathbb{Q}_{nt-1}}. \tag{I.7}$$

Subtracting this estimate of the new buildings term in equation (I.7) from the overall percentage change in rateable values in the valuation year t in equation (I.5), we obtain the revaluation term, which equals the percentage change in the price of floor space between valuation years t and $t-5$:

$$\begin{aligned}\left(\frac{Q_{nt}}{Q_{nt-5}}-1\right)&=\frac{\left(Q_{nt}-Q_{nt-5}\right)H_{nt-1}}{Q_{nt-5}H_{nt-1}},\\&=\left(\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-1}}-1\right)-\frac{\left(H_{nt}-H_{nt-1}\right)Q_{nt}}{Q_{nt-5}H_{nt-1}},\\&=\left(\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-1}}-1\right)-\frac{\left(0.5\times\left(\mathbb{Q}_{nt-1}-\mathbb{Q}_{nt-2}\right)\right)+\left(0.5\times\left(\mathbb{Q}_{nt+1}-\mathbb{Q}_{nt}\right)\right)}{\mathbb{Q}_{nt-1}}.\end{aligned} \tag{I.8}$$

I7.4 Changes in the Price and Supply of Floor Space

Using equation (I.8), we recover the relative change in the price of floor space between valuation years t and $t-5$:

$$G^{Q}_{n,t,t-5}=\frac{Q_{nt}}{Q_{nt-5}}=\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-1}}-\frac{\left(0.5\times\left(\mathbb{Q}_{nt-1}-\mathbb{Q}_{nt-2}\right)\right)+\left(0.5\times\left(\mathbb{Q}_{nt+1}-\mathbb{Q}_{nt}\right)\right)}{\mathbb{Q}_{nt-1}}. \tag{I.9}$$

We directly observe the relative change in rateable values between valuation years t and $t-5$:

$$G^{\mathbb{Q}}_{n,t,t-5}=\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-5}}. \tag{I.10}$$

Noting that rateable values are the product of the price and supply of floor space ($\mathbb{Q}_{nt}=Q_{nt}H_{nt}$), we have:

$$G^{\mathbb{Q}}_{n,t,t-5}=G^{Q}_{n,t,t-5}G^{H}_{n,t,t-5}=\frac{Q_{nt}}{Q_{nt-5}}\frac{H_{nt}}{H_{nt-5}}. \tag{I.11}$$

Using equation (I.11), we solve for the relative change in the supply of floor space between valuation years t and $t-5$ from the observed relative change in rateable values and the relative change in the price of floor space from equation (I.9) above:

$$G^{H}_{n,t,t-5}=\frac{H_{nt}}{H_{nt-5}}=\frac{G^{\mathbb{Q}}_{n,t,t-5}}{G^{Q}_{n,t,t-5}}=\frac{\mathbb{Q}_{nt}/\mathbb{Q}_{nt-5}}{Q_{nt}/Q_{nt-5}}. \tag{I.12}$$

Cumulating these five-year relative changes from 1871 to 1921, we obtain the cumulative changes in rateable values, the price of floor space and the supply of floor space over this period:

$$G^{\mathbb{Q}}_{n,1871,1921}=G^{Q}_{n,1871,1921}\,G^{H}_{n,1871,1921}. \tag{I.13}$$

I7.5 Specification Check

We now compare our model's predictions for the changes in the supply of floor space in each borough under our assumption of a constant floor space supply elasticity (μ) to these separate estimates of the change in the supply of floor space from London County Council. Using the floor space supply function from equation (25) in the paper, we recover the model's predictions for the relative change in the price and supply of floor space as a function of the observed relative change in rateable values and our calibrated floor space supply elasticity (μ):

$$\frac{Q_{nt}}{Q_{nt-5}}=\left(\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-5}}\right)^{\frac{1}{1+\mu}},\qquad\qquad\frac{H_{nt}}{H_{nt-5}}=\left(\frac{\mathbb{Q}_{nt}}{\mathbb{Q}_{nt-5}}\right)^{\frac{\mu}{1+\mu}}. \tag{I.14}$$

Using these predictions of the model from equation (I.14), we compute relative changes in the supply of floor space in the model from 1871 to 1921 for each of the 22 boroughs and aggregations of boroughs discussed above. In Figure I.20, we display these model predictions for the change in the supply of floor space against the estimates computed by London County Council according to equation (I.12). As apparent from the figure, we find a strong, positive and approximately log linear relationship between the two variables, with a statistically significant correlation of 0.77. Therefore, while the assumption of a constant floor space supply elasticity (μ) is necessarily an abstraction, these results suggest that this assumption provides a reasonable approximation to the observed data.

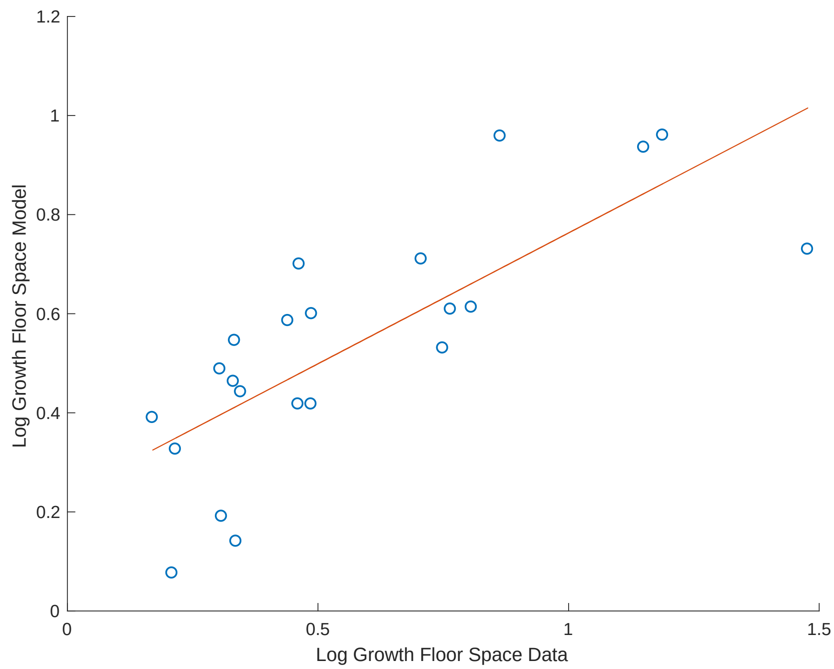

FIGURE I.20:
Model's Predictions for the Log Relative Change in the Supply of Floor Space from 1871-1921 Compared to Separate Estimates from London County Council

Note: Model's predictions for the log relative change in the supply of floor space under the assumption of a constant floor space supply elasticity (μ) from equation (I.14). Separate estimates of the log relative change in the supply of floor space from London County Council computed according to equation (I.12). Observations are the 22 boroughs and aggregations of boroughs in the County of London discussed above.

J Data Appendix

This section of the online appendix reports additional information about the data sources and definitions, supplementing the discussion in Section III. of the paper. Section J1 discusses the population data from the population censuses of England and Wales from 1801-1921. Section J2 summarizes the rateable value data from 1815-1921. Section J3 introduces our bilateral commuting data for 1921 from the population census of England and Wales. Section J4 outlines the historical data on day population from the City of London Day Censuses. Section J5 explains the construction of our geographical information systems (GIS) shapefiles of the overground and underground railway networks over time. Section J6 discusses our estimates of historical construction costs of the overground and underground railway networks. Section J7 explains the construction of our geographical information systems (GIS) shapefiles of the omnibus and tram network over time. Section J8 reports the historical data on average travel speeds for alternative transport models in

London during our sample period. Section J9 contains further details on our historical data on commuting patterns from the personnel ledgers of Henry Poole Bespoke Tailors. Sections J10 discusses our calibration of model parameters based on historical data for our sample period. Section J11 summarizes the historical data on population, employment and commuting distances for Berlin, Paris, Boston, Chicago, New York, and Philadelphia that we use to show that our findings for London are representative of those for other large metropolitan areas following the improvements in transport technology during the 19th-century.

J1 Population Data

Population data from the population censuses of England and Wales from 1801-1891 was provided by the *Cambridge Group for the History of Population and Social Structure* (Cambridge Group), as documented in Wrigley (2011). The original sources for the population data are as follows:

- 1801 Census Report, Abstract of answers and returns, PP 1801, VI
- 1811 Census Report, Abstract of answers and returns, PP 1812, XI
- 1821 Census Report, Abstract of answers and returns, PP 1822, XV
- 1831 Census Report, Abstract of the Population Returns of Great Britain, PP 1833, XXXVI to XXXVII
- 1841 Census Report, Enumeration Abstract, PP 1843, XXII
- 1851 Census Report: Population Tables, part II, vols. I to II, PP 1852-3, LXXXVIII, parts I to II
- 1861 Census Report: Population tables, vol. II, PP 1863, LIII, parts I to II
- 1871 Census Report: vol. III, Population abstracts: ages, civil condition, occupations and birthplaces of people, PP 1873, LXXI, part I
- 1891 Census Report: vol. II, Area, Houses and Population: registration areas and sanitary districts, PP 1893-4, CV [which also includes the 1881 data, as used in our analysis]

The smallest unit of observation and the lowest tier of local government are civil parishes, which we refer to simply as *parishes*. The boundaries of these parishes can change across the population censuses. To create a consistent panel of mappable spatial units over time, the Cambridge Group has developed a two-stage procedure. First, they spatially match parish level polygons and geographical units from each census to derive all spatial units that existed in any period between 1801–1891. They refer to this as *CGKO* (Cambridge Group Kain Oliver) map. This dataset includes 456 polygons for the Greater London Authority (GLA). Next, they employ a *Transitive Closure Algorithm* from graph theory (see for example Cormen, Leiserson, Rivest and Stein 2009) to determine the lowest common unit between parish polygons in different years, which defines the *mappable units*. This procedure implies that these mappable units do not necessarily represent real parishes, but for simplicity we continue to refer to them as *parishes*. After applying the transitive closure algorithm, we obtain 283 mappable units for the GLA. The average mappable unit has a size of 5.60 square kilometers, with 4,042 inhabitants in 1801 and 19,686 inhabitants in 1891. Within the GLA, we further distinguish London County Council (LCC) which encloses 183 mappable units with an average size of 1.64 square kilometers and an average number of inhabitants of 5,432 (22,890) in 1801 (1891) respectively; and the City of London (COL) with 111 mappable units of an average size of 0.02 square kilometers and an average number of inhabitants of 1,219 (1801) and 348 (1891) respectively.

Population data from 1901-1921 stem from the Integrated Census Microdata Project (I-CeM). The majority of parishes did not experience any change in boundaries from 1891-1901. Therefore, we can simply extend the parish panel for 1801-1891 discussed above to 1901. However, from 1911 onwards, there are a number of major changes in parish boundaries. Most notably, the City of London (COL) consisted of more than 100 parishes in the censuses for 1801-1901, which were amalgamated into a single parish in 1907. To avoid having to make assumptions in order to disaggregate the 1911 and 1921 population data for the COL, we end our parish-level panel dataset in 1901.

The next smallest unit of observation is referred to as either metropolitan borough, urban district or rural district in the population census, depending on the level of urbanization of that location. For simplicity, we refer to these units as *boroughs*. We use the boundaries of these boroughs from the 1921 population census to construct consistent panel data on the population of boroughs from 1801-1921. There are 99 of these boroughs in the GLA, 29 in the LCC, and the COL is its own borough. The average borough has an area of 16 square kilometers in the GLA (10.81 in the LCC and 2.98 in the COL). For 1921, we obtain borough population data directly from the population census for that year. For the years before 1921, we overlay the 1921 boroughs and the mappable units discussed above, and allocate the population of the mappable units to the 1921 boroughs, by weighting the values for each mappable unit by its share of the geographical area of the 1921 boroughs. Given that mappable units have a much smaller geographical area than boroughs, most of them lie within a single borough.

J2 Rateable Value Data

We measure the value of floor space using rateable values, which correspond to the annual flow of rent for the use of land and buildings, and equal the price times the quantity of floor space in the model. In particular, these rateable values correspond to "The annual rent which a tenant might reasonably be expected, taking one year with one another, to pay for a hereditament, if the tenant undertook to pay all usual tenant's rates and taxes ... after deducting the probable annual average cost of the repairs, insurance and other expenses" (see London County Council 1907).

These rateable values cover all categories of property, including public services (such as tramways, electricity works etc), government property (such as courts, parliaments etc), private property (including factories, warehouses, wharves, offices, shops, theaters, music halls, clubs, and all residential dwellings), and other property (including colleges and halls in universities, hospitals and other charity properties, public schools, and almshouses). As discussed in Stamp (1922), there are three categories of exemptions: (1) Crown property occupied by the Crown (Crown properties leased to other tenants are included); (2) Places for divine worship (church properties leased to other tenants are included); (3) Concerns listed under No. III Schedule A, namely: (i) Mines of coal, tin, lead, copper, mundic, iron, and other mines; (ii) Quarries of stone, slate, limestone, or chalk; ironworks, gasworks, salt springs or works, alum mines or works, waterworks, streams of water, canals, inland navigations, docks, drains and levels, fishings, rights of markets and fairs, tolls, railways and other ways, bridges, ferries, and cemeteries. Rateable values were assessed at the parish level approximately every five years during our sample period. All of the above categories of properties are included, regardless of whether or not their owners are liable for income tax.

These rateable values have a long history in England and Wales, dating back to the 1601 Poor Relief Act, and were originally used to raise revenue for local public goods. Different types of rateable values can be distinguished, depending on the use of the revenue raised: Schedule A Income Taxation, Local Authority Rates, and Poor Law Rates. Where available, we use the Schedule A rateable values, since Schedule A is the section of the national income tax concerned with income from property and land, and these rateable values are widely regarded as corresponding most closely to market valuations. For example, Stamp (1922) argues that "It is generally acknowledged that the income tax, Schedule A, assessments are the best approach to the true values" (page 25). After the Metropolis Act of 1869, all rateable values for the County of London are computed on the basis of Schedule A Income Taxation. Where these Schedule A rateable values are not available, we use the Local Authority rateable values, Poor Law rateable values, or property valuations for income tax. For years for which more than one of these measures is available, we find that they are highly correlated with one another across parishes.

The original sources for the rateable values data used for each year are as follows:

- **1815**: Property valuations for income tax. Return to an address of the Honourable the House of Commons, dated 21 February 1854; House of Commons Papers, vol. LVI.1, paper no: 509.

- **1843**: Property valuations for income tax. Return to an address of the Honourable the House of Commons, dated 21 February 1854; House of Commons Papers, vol. LVI.1, paper no: 509.

- **1847**: Poor Law Rateable Values. Return to an order of the Honourable the House of Commons, dated 31 August 1848; House of Commons Papers, vol. LIII.11, paper no: 735.

- **1852**: Property valuations for income tax. Return to an address of the Honourable the House of Commons, dated 21 February 1854; House of Commons Papers, vol. LVI.1, paper no: 509.

- **1860**: Property valuations for income tax. Return to an order of the Honourable the House of Commons, dated 13 August 1860; House of Commons Papers, vol. XXXIX, paper no: 546.

- **1881**: Poor Law Rateable Values. A Statement of the Names of the Several Unions And Poor Law Parishes In England And Wales; And of the Population, Area, And Rateable Value Thereof in 1881. London: Her Majesty's Stationery Office, 1887.

- **1896**: Schedule A Rateable Values. Agricultural Rates Act, 1896. Reports separate data on the rateable value of agricultural land and the rateable value of other land and buildings. Return to an order of the Honourable the House of Commons, dated 27 July 1897; House of Commons Papers, paper no: 368; 1897.

- **1905**: Schedule A Rateable Values. Local taxation returns (England and Wales). The annual local taxation returns. Year 1904-05. Part I. House of Commons Papers, vol. CI.1, paper no: 311, 387; 1906.

- **1911**: Schedule A Rateable Values. Local taxation returns (England and Wales). The annual local taxation returns. Year 1910-11. Part I. House of Commons Papers, vol. LXXII.1, Paper no: 264, 268, 364, 282; 1912.

- **1921**: Local Authority Rateable Values, Ministry of Health. Statement showing, for each borough and other urban district in England and Wales, and for 100 typical rural parishes, the amount of the local rates per pound of assessable value, for the financial years 1920-21 and 1921-22, and the assessable values in force at the commencement of the year 1921-22. House of Commons Papers, vol. XVII. 625, Paper no: 1633; 1922.

To create consistent spatial units over time, we manually match parishes with the spatial units provided by the *CGKO* (Cambridge Group Kain Oliver) map, as discussed for the population data in Section J1 of this online appendix. We then use area weights to create the same *mappable units* as for the population data. This procedure gives us a parish-level panel for the years 1815, 1843, 1848, 1852, 1860, 1881 and 1896. Finally, we aggregate these parish data to the 1921 boroughs using area weights, as discussed for the population data in Section J1 above.

For the years 1905, 1911 and 1921, we use rateable values at the borough level. For the year 1921, we observe rateable values for all 1921 boroughs, including metropolitan boroughs, urban districts and rural districts. However, some boroughs were created for the first time in 1921, when a previously-existing borough was sub-divided into separate urban and rural districts. Therefore, for the years 1905 and 1911, we are missing data for these newly-created sub-divisions: Croydon (created in 1912), Orset (created in 1912), Watford (created in 1906), Hitchin (created in 1919), Dartford (created in 1920), Bexley (created in 1920), Uxbridge (created in 1920), Chertsey (created in 1909) and Hambeldon (not separately reported in 1905 and 1911). To deal with these sub-divisions, we allocate the data for the larger 1905 and 1911 spatial units across their 1921 sub-divisions using area weights.

As discussed above, Schedule A rateable values, local authority rateable values, poor law rateable values and property valuations for income tax are highly correlated across parishes. Nonetheless, the level of the property valuations for income tax is somewhat lower than the rateable values, which is consistent with the fact that rateable values include all properties, regardless of whether their owners are liable for income tax, whereas the property valuations for income tax are based on income tax liability. To address this difference, we use a consistent time-series on Schedule A rateable values for the County of London. This consistent time-series was constructed for an aggregate of 28 boroughs in the County of London by London County Council for the years 1830, 1835, 1840, 1845, 1850, 1855, 1860, 1865, 1871, 1876, 1881 and 1891. We extend this time-series forward to 1921, using the reported Schedule A rateable values for these 28 boroughs reported in *London Statistics*. We also extend this time-series back to 1815, using the annual population growth rate for these 28 boroughs from 1815-1830. In Figure J.1, we display the resulting time-series for this aggregate of 28 boroughs in the County of London.

For each year, we first construct an adjustment factor, which is equal to the ratio of our property valuation to the consistent Schedule A rateable value for the aggregate of 28 boroughs in the County of London. We next adjust our property valuation upwards or downwards for all boroughs using this adjustment factor. The adjustment factors for each year are 0.72 (1815), 0.58 (1843), 0.97 (1847), 0.96 (1852), 0.57 (1860), 1.01 (1881), 1.03 (1896), 1.00 (1905), 0.99 (1911) and 1 (1921). These adjustment factors are all close to one in years for which we use rateable values, which is consistent with the idea that Schedule A rateable values, Local Authority rateable values and Poor Law rateable values are all highly correlated with one another. These adjustment factors are less than one for years in which we use property valuations for income tax, which is consistent with the fact that the Schedule A rateable values include all properties, regardless of whether the owners of those properties are liable for income tax, whereas the property valuations for income tax are based on income tax liability.

Finally, we use linear interpolation in between the above years to construct a time-series on rateable values for each borough and for each census decade, from 1815 before the arrival of the railway in 1831, to the end of our sample period in 1921.

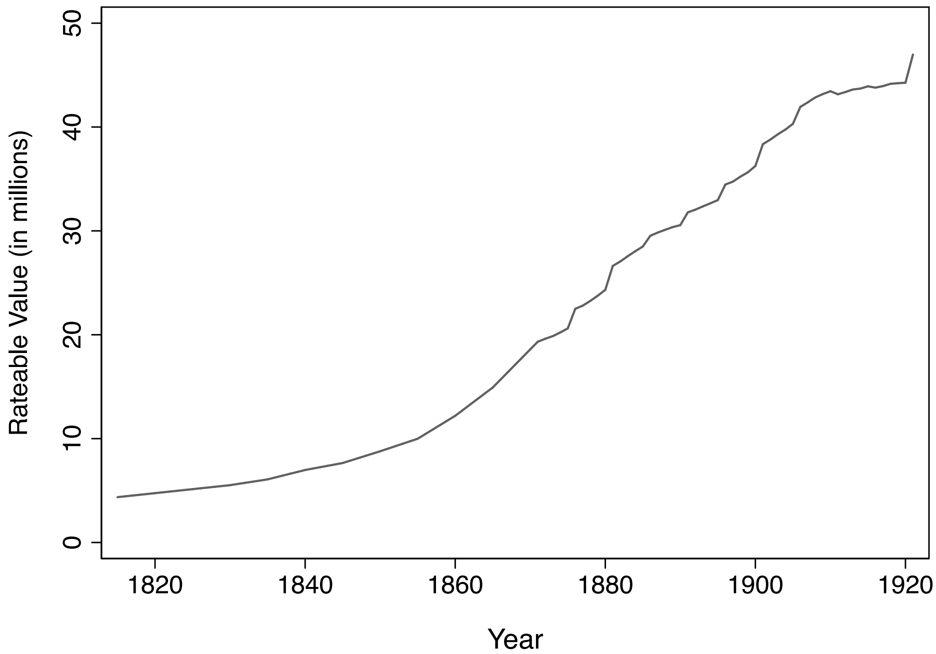

FIGURE J.1:
Schedule A Rateable Value for an Aggregate of 28 Boroughs in the County of London from 1815-1921

Notes: Current price millions of pounds, London County Council (LCC) and authors' calculations.

J3 Bilateral Commuting Data for 1921

The 1921 population census of England and Wales reports bilateral flows of commuters between their residence and workplace boroughs. This data was first published for London and the five Home Counties (Essex, Hertfordshire, Kent, Middlesex and Surry) in Census of England & Wales 1921 (1923) "Workplaces in London and the Five Home Counties, Tables Part III (Supplementary)" and then for all boroughs in England and Wales in Census of England & Wales 1921 (1925) "Workplaces." We used the publication for London and the Home Counties as our main data source and used the later publication for all of England and Wales to obtain information on inflows of workers to London and the Home Counties from other parts of England and Wales. The residence of a worker is the borough in which the person was located on census evening whether as a permanent resident or as a temporary visitor. The workplace of a worker is reported in four categories: (a) workers who work in the borough in which they were located on census evening; (b) workers who have no fixed workplace; (c) workers whose workplace is not known; (d) workers who work in another borough than the borough in which they were on census evening. Groups (b) and (c) are typically very small, and we assume that these workers work in their borough of residence. For workers with a workplace outside their residence borough, the census reports flows to each destination borough. Bilateral flows of less than 20 people are not reported for confidentiality reasons and are omitted. Summing these reported bilateral flows, the resulting sums of workplace employment and residence employment are close to the totals for workplace employment and residence employment (including flows of less than 20 people) that are separately reported in the population census.

J4 Employment and Day Population Data

For 1921, we use the bilateral matrix of flows of commuters from each residence borough (rows) to each workplace borough (columns) discussed above to measure employment by residence and employment by workplace. Summing across columns in this matrix, we obtain employment by residence for 1921 (which we refer to as "residence employment") for each borough. Summing across rows in this matrix, we obtain employment by workplace for 1921 (which we refer to as "workplace employment"). We also construct an employment participation rate for each borough in 1921 by dividing residence employment by population.

For years prior to 1921, we construct residence employment using our population data from the population censuses for England and Wales. Assuming that the ratio of residence employment to population is stable for a given borough over time, we use the 1921 value of this ratio and the historical population data to construct residence employment for each borough for each decade from 1801-1911. Consistent with a stable employment participation rate, we find relatively little variation in the ratio of residence employment to population across boroughs in 1921.

Data on workplace employment are not available prior to 1921. Therefore, in our structural estimation of the model, we use our bilateral commuting data for 1921, together with our data on residence employment and rateable values for earlier years, to generate model predictions for workplace employment for earlier years. We calibrate the Fréchet shape parameter (ϵ) in the model by comparing these model predictions to the data on the day population from the City of London Day Censuses. In the face of the increased commuting from the mid-19th century onwards, the City of London Corporation recognized that the measure of population from the population census of England and Wales, which is based on where one slept on census night ("night population"), could be a misleading indicator of the population present during the daytime ("day population"). Therefore, the City of London Corporation undertook Day Censuses in 1866, 1881, 1891 and 1911 to record "... every person, male or female, of all ages, residing, engaged, occupied, or employed in each and every house, warehouse, shop, manufactory, workshop, counting house, office, chambers, stable, wharf, etc, and to include all persons, of both sexes and all ages, on the premises during the working hours of the day, whether they sleep or do not sleep there ..." (Salmon 1891, page 97).

The original sources for the City of London day census data are as follows:

- Day Census, City of London, Report, Local Government and Taxation Committee, 13th December, 1866.
- Report on the City Day Census, 1881, By the Local Government and Taxation Committee of the Corporation of London, Second Edition, London: Longmans, Green and Company.
- Ten Years' Growth of the City of London, Report, Local Government and Taxation Committee of the Corporation, by James Salmon, London: Simpkin, Marshall, Hamilton, Kent and Company, 1891.
- City of London Day Census, 1911, Report, County Purposes Committee of the Corporation, by Henry Percival Monckton, London: Simpkin, Marshall, Hamilton, Kent and Company.

J5 Overground and Underground Railway Network

We have geographical information systems (GIS) information on the location of all railway lines and stations opened for the public carriage of passengers and/or goods and the year in which they opened. This GIS dataset was provided by the *Cambridge Group for the History of Population and Social Structure*, which based its digitization on Cobb (2003).

Using these data, we construct separate networks for overground and underground railways in each census year. In Table J.1 below, we summarize the opening years and respective lengths of the London Underground lines. In Figures J.2 to J.10 below, we show the decennial evolution of the underground and overground railway network across the Greater London Authority (GLA), where 1841 is the first census year in which an overground railway exists, and 1871 in the first census year in which an underground railway exists.

In Table J.1, we use the modern names of underground lines, which do not always correspond to their names in 1921. We further exclude parts of the London Underground that did not exist in 1921. The *Victoria Line* and the *Jubilee Line* did not open until 1968 and 1979 respectively. We do not list the *Circle Line* and the *Hammersmith & City Line*, because they were both part of the network of the District and Metropolitan line in 1921. We also exclude the *Waterloo and City Line*, because it was not classified as part of the London Underground when it opened in 1898, even though its tracks run underground from Waterloo station underneath the River Thames to Bank station in the City of London with no intermediate stops. We follow this convention and classify this line, which is only 1.59 miles long, as an overground railway. Although this line was formally owned by the Waterloo and City Railway Company, it was operated from the start by an overground railway company: the London and South Western Railway (LSWR). In 1907, the LSWR formally absorbed the Waterloo and City Railway Company. From that year onwards, the Waterloo and City Line continued to be operated by overground railway companies, and was only officially taken over by the London Underground system in 1994. Finally, we classify the *East London Line* as part of the London Underground, because it was initially operated by a consortium that included the District and Metropolitan lines. After 1933, this line became known as the East London Part of the Metropolitan Line. Today, it is part of the London overground railway network. This line is mostly above ground, but it uses the Thames tunnel built by Isambard Kingdom Brunel between 1825 and 1843 for horse-drawn carriages.

In measuring construction costs for underground railways, we distinguish shallow lines built using "cut-and-cover" techniques and deep lines built using "bored tubes," as discussed in Section J6 of this online appendix. In central London, the *District Line* and the *Metropolitan Line* are cut-and-cover lines, while the remaining lines use bored tubes. Outside central London, parts of both types of line are above ground. We also measure the length of each type of line that is below and above ground, and take this into account in our measures of construction costs.

TABLE J.1:
Opening Years and Lengths of London Underground Lines up to 1921

	All tracks		Tracks Underneath Ground	
Line	Opening Date	Length (in km)	Opening Date	Length (in km)
Bakerloo Line	1906	6.04	1906	6.04
Bakerloo Line	1907	1.04	1907	1.04
Bakerloo Line	1913	0.76	1913	0.76
Bakerloo Line	1915	3.32	1915	3.32
Bakerloo Line	1916	1.25		
Bakerloo Line	1917	11.29		
		23.70		11.15
Central Line	1900	9.30	1900	9.30
Central Line	1908	0.76	1908	0.76
Central Line	1912	0.63	1912	0.63
Central Line	1920	6.92		
		17.62		10.69
District Line	1868	4.49	1868	4.48
District Line	1869	2.71	1869	1.03
District Line	1871	2.84	1871	2.84
District Line	1874	2.34		
District Line	1877	7.65		
District Line	1879	3.79		
District Line	1880	2.17	1880	0.75
District Line	1883	8.85		
District Line	1884	3.66	1884	3.01
District Line	1902	2.59	1902	2.59
District Line	1905	0.69		
		41.79		14.70
East London Line	1869	3.45	1869	0.40
East London Line	1871	2.23		
East London Line	1876	2.67		
East London Line	1880	1.77		
		10.12		0.40
Metropolitan Line	1863	6.46	1863	6.46
Metropolitan Line	1865	0.71	1865	0.71
Metropolitan Line	1868	7.16	1868	7.16
Metropolitan Line	1875	0.52	1875	0.52
Metropolitan Line	1876	0.44	1876	0.44
Metropolitan Line	1879	3.48	1879	0.60
Metropolitan Line	1880	8.87		
Metropolitan Line	1884	0.35		
Metropolitan Line	1885	3.48		
Metropolitan Line	1887	5.93		
Metropolitan	1904	14.76	1904	4.70
		52.16		20.60
Northern Line	1890	3.87	1890	3.87
Northern Line	1900	3.77	1900	3.77
Northern Line	1901	2.30	1901	2.30
Northern Line	1907	14.70	1907	14.70
		24.63		24.63
Piccadilly Line	1903	8.12		
Piccadilly Line	1906	14.59	1906	14.59
Piccadilly Line	1907	0.71	1907	0.71
Piccadilly Line	1910	1.70		
		25.11		15.30

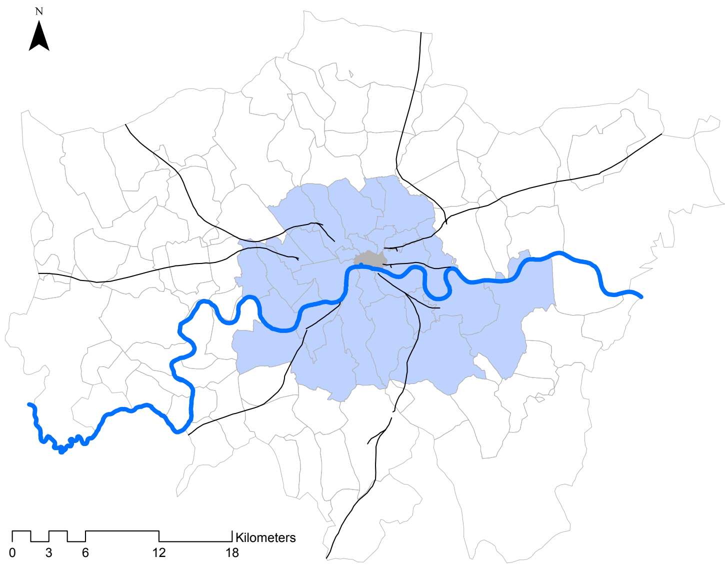

FIGURE J.2:
Overground Railway Network in Greater London 1841

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black.

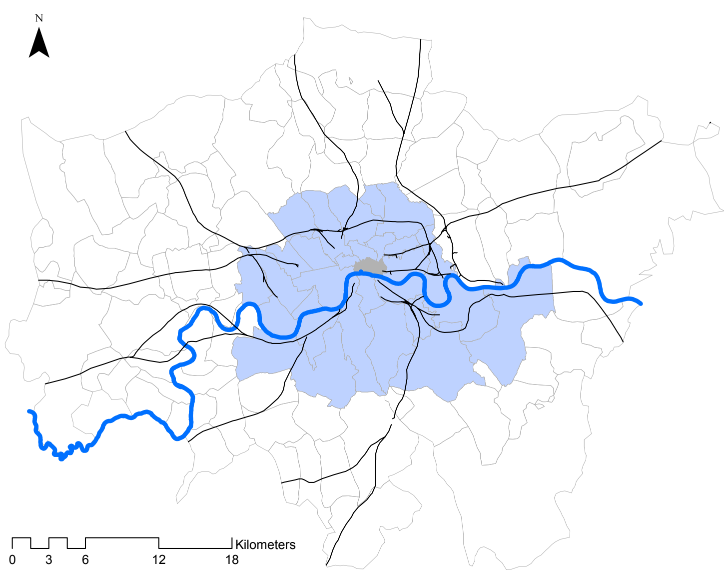

FIGURE J.3:
Overground Railway Network in Greater London 1851

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black.

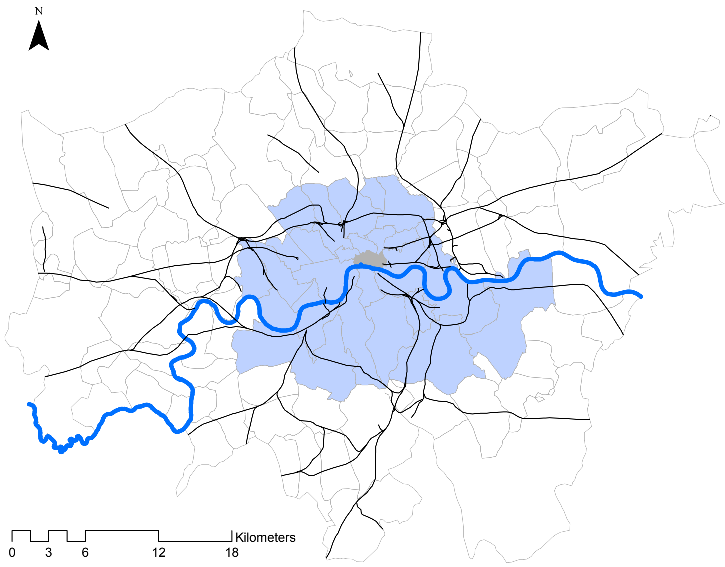

FIGURE J.4:
Overground Railway Network in Greater London 1861

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black.

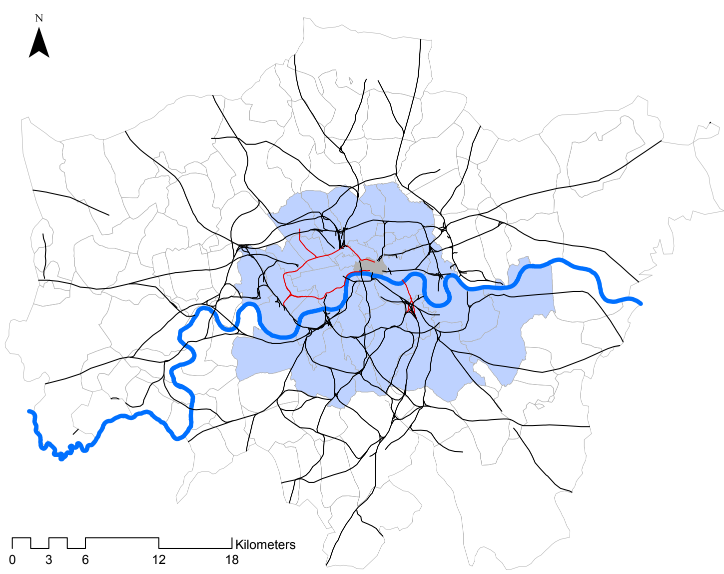

FIGURE J.5:
Overground and Underground Railway Network in Greater London 1871

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

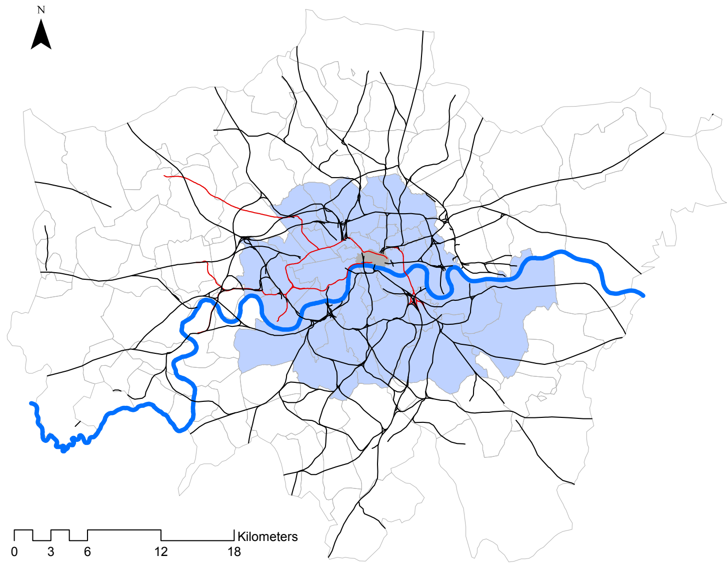

FIGURE J.6:
Overground and Underground Railway Network in Greater London 1881

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

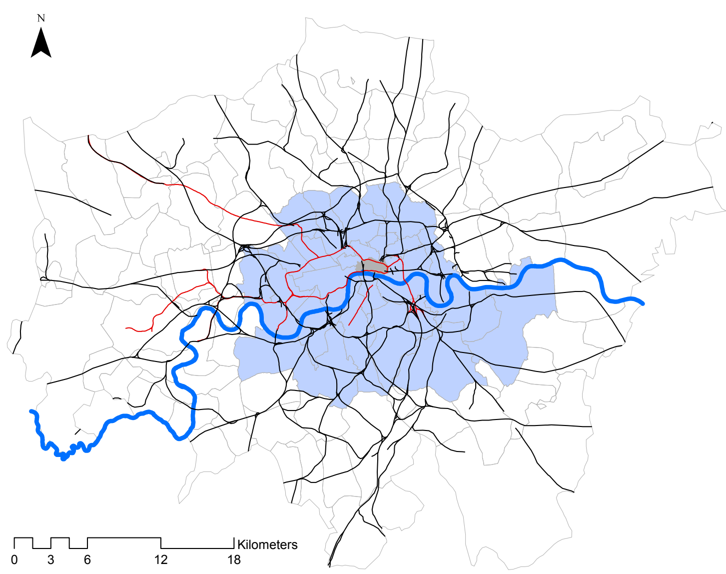

FIGURE J.7:
Overground and Underground Railway Network in Greater London 1891

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

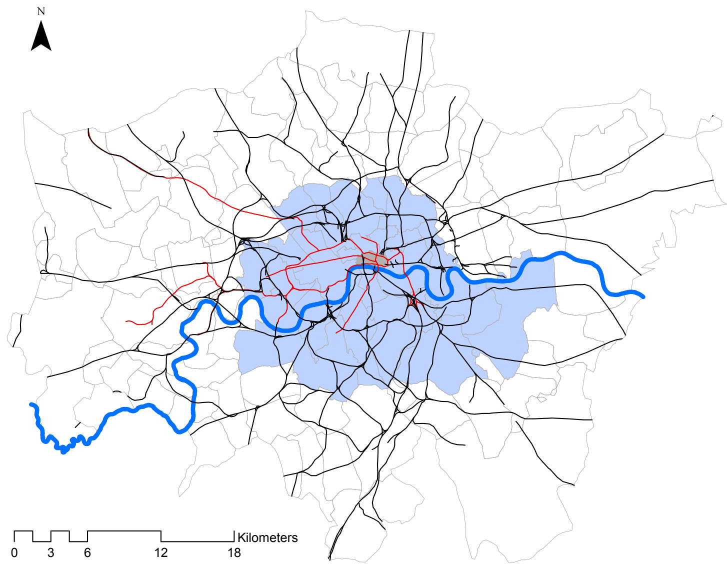

FIGURE J.8:
Overground and Underground Railway Network in Greater London 1901

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

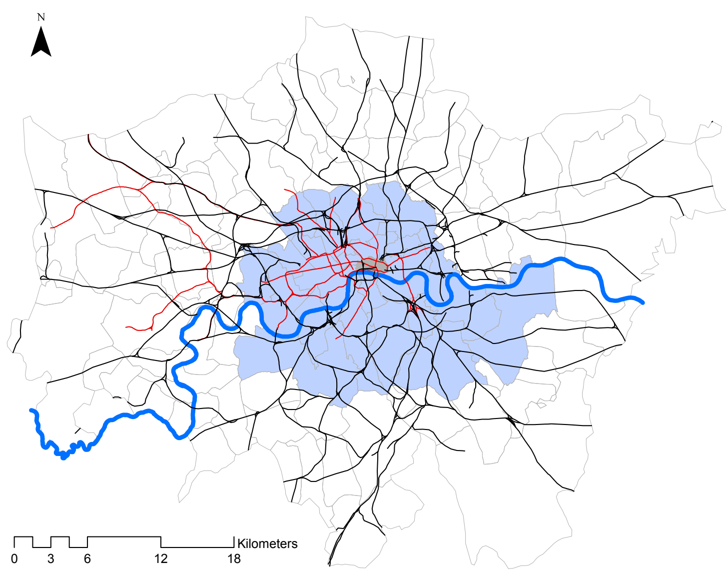

FIGURE J.9:
Overground and Underground Railway Network in Greater London 1911

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

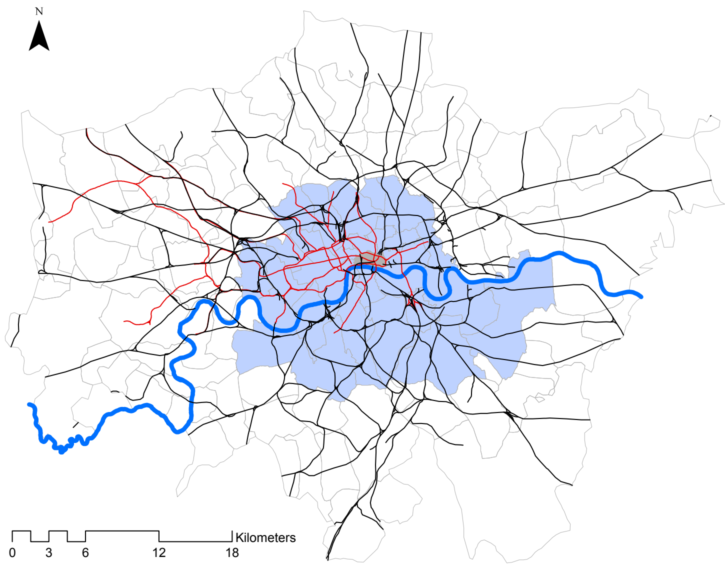

FIGURE J.10:
Overground and Underground Railway Network in Greater London 1921

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

J6 Construction Costs Estimates of the Railway and Underground Lines

In this section of the online appendix, we estimate separate construction costs per mile of line for overground and underground railways. For underground railways, we further distinguish between the different construction costs of shallow "cut-and-cover" and deep "bored-tube" underground railways.

J6.1 Underground Railways

The earliest London Underground lines (such as the Metropolitan District Railway that was opened in 1863) used shallow "cut-and-cover" construction methods, with the lines frequently running underneath existing streets. These "cut-and-cover" methods involve excavating a trench for the underground line and constructing a roof overhead to bear the load of whatever is above. Following improvements to existing tunneling shields by James Greathead in the 1870s and 1880s, the first London Underground railway that was built using deep "bored-tube" techniques was the City and South London railway, which opened in 1890. Greathead's shield consisted of an iron cylinder, which was inched forward as the working face was excavated, while behind it a permanent tunnel lining of cast iron segments was fitted into place.[10]

Before the Second World War, both overground and underground lines in Great Britain were constructed by private companies, who raised capital to finance construction and paid dividends on this capital from their ticket revenue. Construction of each railway line or extension of an existing line had to be approved by Parliament. A simple measure of the construction costs of overground and underground railway lines is therefore the amount of capital that these private companies raised per mile of line constructed. This measure should capture the full costs of constructing railway lines, including not only the cost of construction work (tunnels, rails, stations and other buildings and structures) but also the cost of purchasing land, rolling stock and fees for operating permissions.[11]

For a 1901 parliamentary report ("Report From The Joint Select Committee of The House of Lords on London Underground Railways") Henry L. Cripps compiled an overview of extensions constructed and proposed by London Underground companies. His data is contained in Appendix B of the report. Cripps reports the length of the extensions and the amount of capital authorized by parliament per mile of underground line. From Cripps's data we extract, for each London Underground company, the years in which the extensions were authorized, the total length in miles of line authorized, and the average authorized capital per mile as summarized in Table J.2.[12]

We classify each company in Table J.2 as either a "bored-tube" or "cut-and-cover" tube operator, based on the classification provided by Croome and Jackson (1993), in order to generate separate cost estimates for these two construction types. During the period surveyed by Cripps, the extensions to the underground network were predominantly for "bored-tube" lines. With the exception of the Metropolitan District extension authorized in 1897, all other projects listed in Table J.2 are "bored-tube" line extensions. For our baseline measure, we use the unweighted average of the authorized capital per mile for these two groups of companies, which yields £330,000 and £500,000 for "cut-and-cover" and "bored-tube" lines, respectively. These averages are reported to the nearest £10,000. We convert these figures in

[10] See the discussion in chapter 4 of Barker and Robbins (1963) for a description of the construction of the Metropolitan line tunnels by "cut-and-cover". Croome and Jackson (1993) provide a detailed history of the construction of the London Underground system from the early "cut-and-cover" construction to the later "bored-tube" lines that could be constructed under all parts of Central London.

[11] Kellet (1969) examines the capital accounts of 26 railway companies in the UK and finds that on average across these companies, as a share of the total cost of construction, the cost of land represented 25 percent and Parliamentary expenses represented a further 6 percent.

[12] Cripps's survey also includes a few projects that were authorized but were never actually built. We exclude such extensions that were proposed but never built from Table J.2.

TABLE J.2:
Authorized Capital per Mile for London Underground Companies

Underground Railway Company	Years Extensions Authorized	Miles of Line Authorized	Authorized Capital per mile in pounds (current prices)	Authorized Capital per mile in pounds (1921 prices)
City and South London	1884-1898	6.88	319,709	374,060
Central London	1891-1892	6.79	559,852	649,428
Great Northern and City	1892-1897	3.48	598,561	682,360
Baker Street and Waterloo	1893-1900	5.25	605,523	653,965
Charing Cross, Euston and Hampstead	1893-1899	6.10	388,196	438,661
Waterloo and City	1893	1.59	453,543	512,504
Brompton and Picadilly Circus	1897-1899	2.41	552,538	574,640
Metropolitan District	1897	4.87	328,205	354,461

Note: Taken from Appendix B of the "Report From The Joint Select Committee of The House of Lords on London Underground Railways" (1901), compiled by Henry L. Cripps. Current year prices converted to constant 1921 prices using a price deflator based on the ratio of overground construction costs in those years from the Railway Returns.

current year prices into common 1921 prices using a price deflator based on the ratio of overground railway construction costs in those years from the "Railway Returns: Returns of the Capital, Traffic, Receipts, and Working Expenditure of the Railway Companies of Great Britain." We thus obtain authorized capital per mile in 1921 prices for "cut-and-cover" and "bored-tube" lines of £355,000 and £555,000. Using a weighted average of authorized capital per mile for of the "bored-tube" underground companies, where the weights are the miles of line authorized, yields a similar figure in 1921 prices of £544,000.

By 1921 many of the London underground lines had added sections to their network that ran overground rather than in tunnels. Typically, as tube lines reached the boundary of the densely built-up area of London the line continued above ground to avoid the heavy costs of tunneling. Based on the work of the *Cambridge Group for the History of Population and Social Structure*, discussed further in Section J5 of this online appendix, we are able to classify the parts of each underground line that run in tunnels and those that are above ground. From this data, it is clear that the extensions to the underground network considered by Cripps were all parts of the network running in tunnels, and hence his authorized capital per mile is for the construction of lines in tunnels. In estimating the overall costs of constructing the underground network, we make the natural assumption that the construction costs per mile for the parts of each underground line that are above ground are the same as those for overground railways, as discussed in the next subsection.

J6.2 Overground Railways

We estimate the construction costs for overground railway lines using the 1921 edition of the "Railway Returns: Returns of the Capital, Traffic, Receipts, and Working Expenditure of the Railway Companies of Great Britain", which compiled large volumes of data on the railways of Great Britain. In particular, the Railway Returns report both the total line length of the UK railway network, as well as both the authorized capital and paid-up capital of all railway companies. For our basic estimate of construction costs, we divide the capital values by the length of line to calculate the average capital per mile.[13] For our baseline measure, we use the authorized capital per mile of railway line, which rounded to the nearest one thousand pounds is £60,000 in 1921 prices. Using instead the average capital paid-up per mile would result in a

[13]Line length is defined treating each line as a single-track line, regardless of whether or not there are multiple tracks that run in parallel.

similar figure of £57,000 in 1921 prices.[14]

J6.3 Robustness

A potential concern with our estimate of overground construction costs is that UK averages may not be representative of construction costs in London and its surroundings. To examine this possibility, we use data from the 1921 edition of the Railway Returns for individual railway companies. This data reports the total authorized capital and the length of the operated network for each major railway company in the UK. From this data, we selected the railway companies that operate services to and from London and computed their average authorized capital per mile of line, which results in a figure of £60,000 to the nearest one thousand pounds.[15] Therefore, London-based companies do not appear to have faced substantially different construction costs from the UK average, and we assume a construction cost of £60,000 per mile for overground railways in Greater London.

As another robustness check, we compared our historical estimates of construction costs for overground railways with those from other sources. The 1911 edition of the "New Dictionary of Statistics" provides a comparison of railway construction costs per mile across 27 countries over the years 1905-1908. Figure J.11 displays the data from the Dictionary of Statistics. The notes provided in the Dictionary of Statistics show that their estimate of the construction costs for the UK in 1907 of £56,000 is also based on the data from the Railway Returns and simply divides total authorized capital by the length of the UK railway network. This estimate for 1907 therefore uses the same data and approach as our estimate of overground construction costs in 1921. The figure shows that construction costs in the UK are, if anything, at the upper bound of construction costs in other countries.

There could be a number of reasons why construction costs in the UK were higher than that of other countries. First, the cost of purchasing land for railway construction in the UK might have been higher than those in other countries. Sir Josiah Stamp, who was the chairman of the London, Midland and Scottish Railway Company, for example, argued that "in international comparisons of costs Great Britain notoriously suffers through the heavy initial outlay in lands and parliamentary expenses to acquire them against bitter opposition." Second, the Dictionary of Statistics notes that the UK probably had a higher share of multiple track railway lines compared to other countries, in particular the US. As a single track line is less expensive to build than multiple track lines, this could contribute to the higher costs per mile of railway line built in the UK. Finally, the UK was one of the leading industrial nations before the First World War. The resulting higher wages compared to other countries are likely to have made construction work, which was labor intensive at this time, more expensive.

As a final robustness check, we examined the evolution of overground railway construction costs over time using the Railway Returns. For each year between 1871 and 1912 and also 1920 and 1921, we can compute the authorized capital per mile of line. A similar time series can be generated for paid-up capital per mile. We find that both capital per mile measures steadily increase over time. In 1871, authorized capital per mile was approximately £40,000. By 1900, this number reaches £60,000, after which it plateaus and remains relatively constant until 1921. These reported changes in authorized capital per mile reflect both inflation and changes in the real costs of railway construction. Nevertheless, they

[14]The total length of railway lines in the UK and also the aggregate capital stocks in the summary table of the Railway Returns implicitly include the London Underground system. As underground railways are more costly to construct than overground railways, this will bias our estimate of construction costs for overground railways upwards, and implies that our estimates of these construction costs for overground railways are upper bounds. In practice, this bias is likely to be small, because the London Underground network is a tiny fraction of the 23,724 miles of railway line that were open in Great Britain in 1921.

[15]The main railway companies that have a terminus in London in 1912 are the Great Eastern, Great Northern, Great Western, London and North Western, London and South Western, London, Brighton and South Coast, North Eastern and South Eastern and Chatham company.

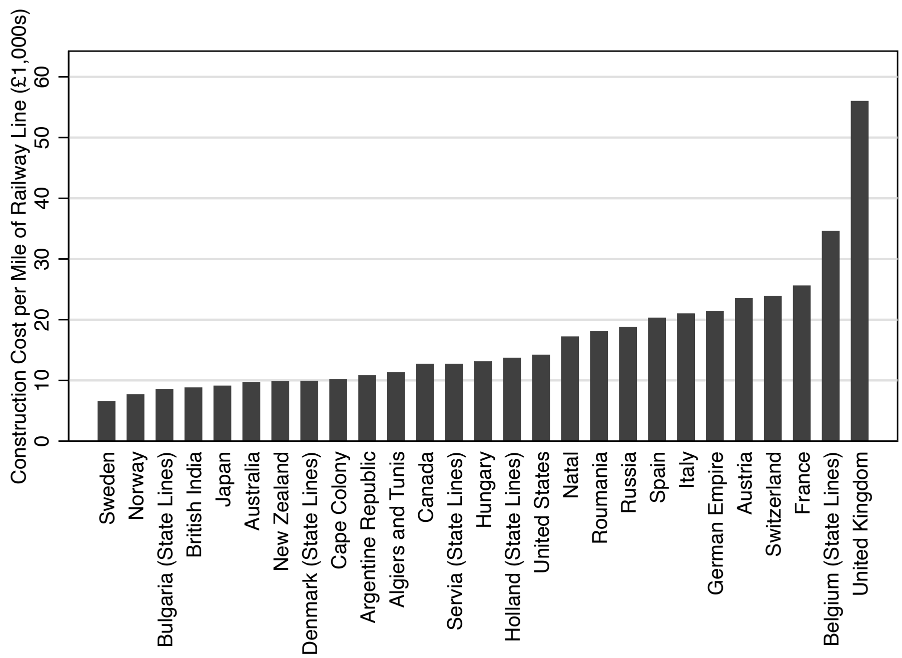

FIGURE J.11:
Construction Costs of Railway Lines Across Countries

Note: Taken from pp. 512-513 of the 1911 edition of the "The New Dictionary of Statistics" compiled by Augustus D. Webb. The data complied by Webb comes from the years 1905 to 1908 depending on data availability for different countries.

suggest that the early parts of the UK network could have been constructed at lower costs, implying that our estimate of £60,000 based on 1921 is again an upper bound on construction costs.

J7 Omnibus and Tram Network

In addition to our geographic information systems (GIS) shapefiles of the railway network for each year, we also constructed an analogous shapefile for each year for the combined horse omnibus, horse tram, motor omnibus, and electric tram networks in each year. Horse omnibuses were first introduced from Paris to London in the 1820s; the first horse trams in London appeared in 1860; the first motor omnibus ran on the streets of London in 1898; and the first fully-operational electric tram service started in 1901. As discussed in Section J8 of this online appendix, the average reported travel speeds for horse omnibuses and horse trams are around 6 miles per hour (mph), which are close to those reported for motor omnibuses and electric trams of around 7-8 mph. Therefore, we construct a single shapefile for each year that contains the combined networks of all forms of omnibus and tram in that year, assuming a single average travel speed of 6 mph.

To construct our GIS shapefiles for all years except 1831, 1841 and 1851, we start with georeferenced original maps of horse omnibus, horse tram, motor omnibus, and electric tram networks. Using these georeferenced maps, we construct line shapefiles for the combined omnibus and tram network in each year. For the years 1831, 1841 and 1851, we use reported route information (origin, intermediate stops and destination) to construct line shapefiles for these networks in each year. The original sources for the data used for each of our census decades are as follows:

- **1831 and 1841**: Appendix 2 of Barker and Robbins (1963) reports bilateral routes for the Board of Stamps list of omnibuses and short-stage coaches licensed to operate in the London area in 1838-9.
- **1851 and 1861**: The Illustrated Omnibus Guide, with a Complete Guide to London, 1851, Simpkin and Company, Stationers' Court, W. H. Smith and Son, 136 Strand, Reprinted for Railwayana Ltd. by Oxford Publishing Company, 1971.
- **1871**: London Horse Omnibus Routes in 1871, map compiled by J. C. Gillham, chiefly from John Murray's Guide to London of 1871 and Adam and Charles Black's Guide.
- **1881**: London Horse Bus and Tram Routes in 1879, map compiled by J. C. Gillham, chiefly from John Murray's, Herbert Fry's and Lambert's Golden Guide Books to London 1879.
- **1891**: London Omnibus Routes early in 1895, map compiled by J. C. Gillham from a list published in 1895 by the London County Council statistical department.
- **1901**: London Omnibus Routes at the end of 1902, map compiled by J. C. Gillham from a list published in 1902 by the London County Council statistical department.
- **1911**: London Omnibus Routes in July 1911, map compiled by J. C. Gillham, chiefly from the London Traffic Report of the Board of Trade, 1911.
- **1921**: London County Council Tramway Map, 1916, London General Omnibus Company Motor Bus Map, 1921.

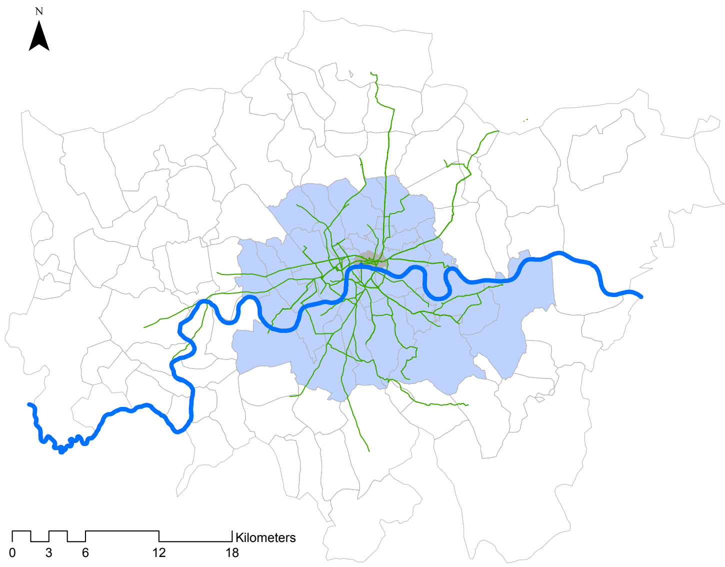

FIGURE J.12:
Omnibus Network in Greater London in 1839

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus lines are shown in green; we use this omnibus network in 1839 for census years 1831 and 1841.

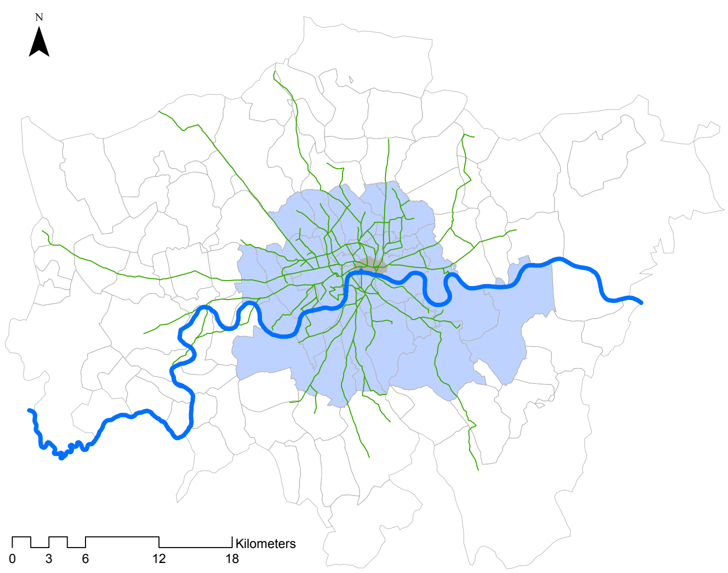

FIGURE J.13:
Omnibus Network in Greater London in 1851

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus lines are shown in green; we use this omnibus network in 1851 for census years 1851 and 1861.

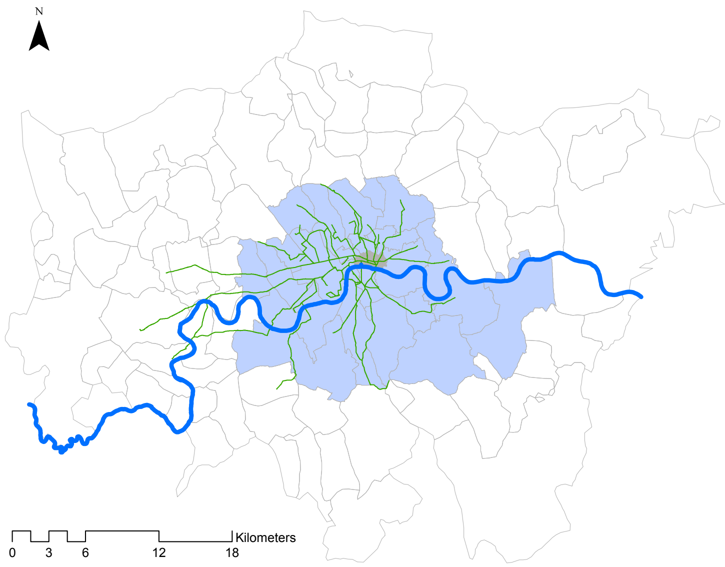

FIGURE J.14:
Omnibus and Tram Network in Greater London in 1871

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus and tram lines are shown in green.

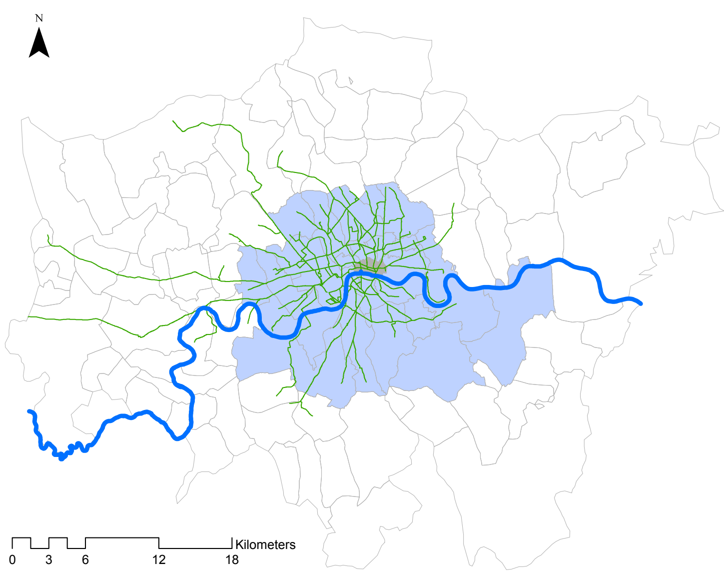

FIGURE J.15:
Omnibus and Tram Network in Greater London in 1881

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus and tram lines are shown in green.

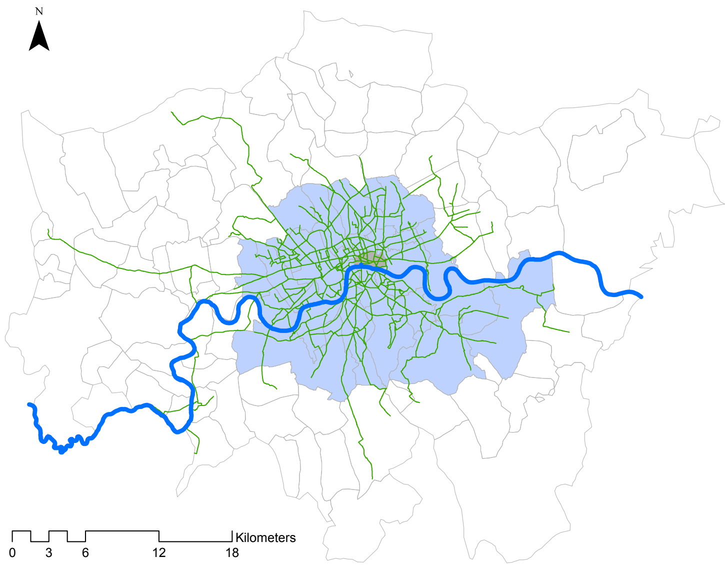

FIGURE J.16:
Omnibus and Tram Network in Greater London in 1891

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus and tram lines are shown in green.

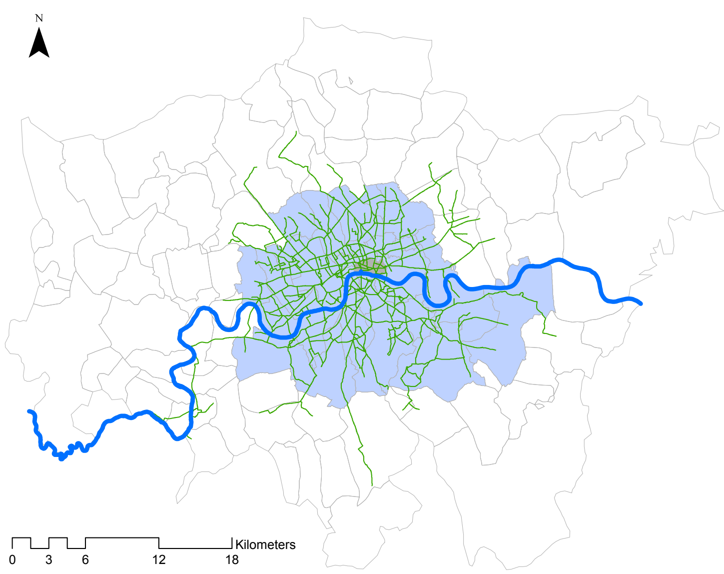

FIGURE J.17:
Omnibus and Tram Network in Greater London in 1901

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus and tram lines are shown in green.

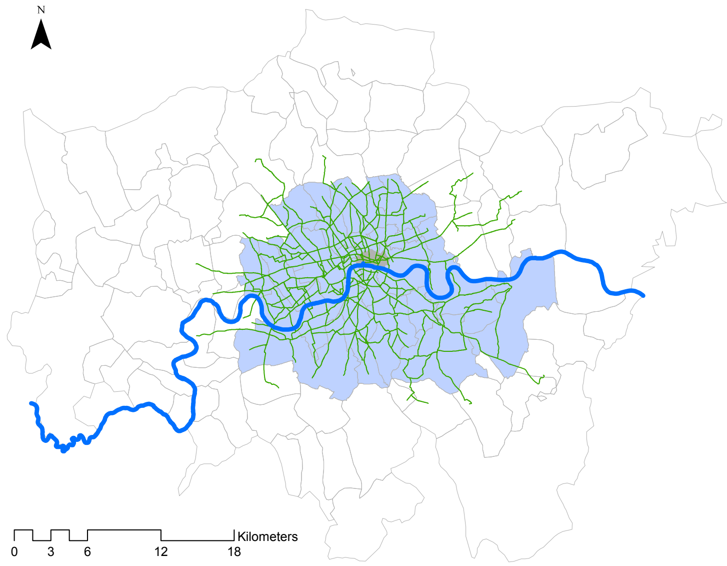

FIGURE J.18:
Omnibus and Tram Network in Greater London in 1911

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus and tram lines are shown in green.

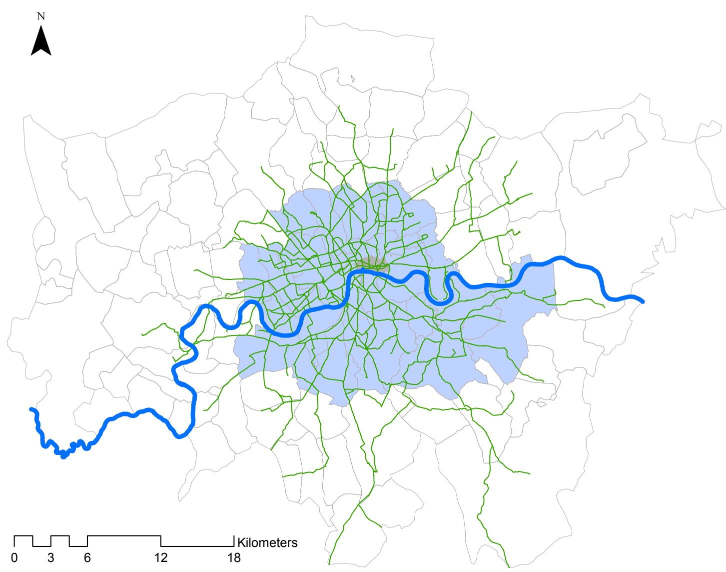

FIGURE J.19:
Omnibus and Tram Network in Greater London 1921

Note: Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; omnibus and tram lines are shown in green.

J8 Data on Travel Speeds using Alternative Transport Modes

As discussed in Section II. of the paper, at the beginning of the 19th-century, walking was the most common mode of transport, with average travel speeds in good road conditions of 3 miles per hour (mph). The state of the art technology for long distance travel was the stage coach, but it was expensive because of the multiple changes in teams of horses required over long distances, and hence was relatively infrequently used. Even with this elite mode of transport, poor road conditions limited average long distance travel speeds to around 5 mph (see for example Gerhold 2005).

With the growth of urban populations, attempts to improve existing modes of transport led to the introduction of the horse omnibus from Paris to London in the 1820s, as discussed in Barker and Robbins (1963). The main innovation of the horse omnibus relative to the stage coach was increased passenger capacity for short-distance travel. However, the limitations of horse power and road conditions ensured that average travel speeds remained low. As reported in Table J.3, average travel speeds using horse omnibuses remained around 6 mph for routes through both central and outlying areas. A further innovation along the same lines was the horse tram, which was introduced in London in 1860. However, average travel speeds again remained low, in part because of road congestion. As reported in Table J.4, average travel speeds during both rush and slack hours using horse trams were little different from those using horse omnibuses. A later innovation was the replacement of the horse tram with the electric tram, with the first fully-operational services starting in 1901. As reported in Table J.4, although electric trams brought some improvement in travel speeds relative to horse trams, during the rush hours when most commuting occurred, average travel speeds remained between 5.5 and 7 mph.

In contrast, overground and underground railways brought a substantial improvement in average travel speeds, transforming the relationship between distance travelled and time taken. The world's first overground railway to be built specifically for passengers was the London and Greenwich railway, which connected what was then the village of Greenwich to Central London, and opened in 1836. As reported in Table J.5, average travel speeds using overground railways and steam locomotives were around 21 mph, with some variation depending on the track layout and number of intermediate stops. The world's first underground railway was the Metropolitan and District Railway, which connected the London termini of Paddington, Euston and Kings Cross with Farringdon Street in the City of London, and opened in 1863. When first opened, the Metropolitan and District Railway used steam locomotives, like its overground counterparts. The City and South London Railway was the first underground railway to use electric traction from its opening in 1890 onwards. As reported in Table J.6, average travel speeds using underground railways were slightly slower than overground railways at around 15 mph, reflecting both differences in the engineering conditions and frequency of intermediate stops. In London today, average travel speeds using overground and underground railways are little different from those reported in Tables J.5 and J.6.[16]

We assume relative weights for different modes of transport based on these average travel speeds in our quantitative analysis of the model, as summarized in Table J.7 below. Normalizing the weight for overground railways to 1 (21 mph), we assume the following weights for the other modes of transport: walking 7 (21/3 mph); horse and motor omnibuses and horse and electric trams 3.5 (21/6 mph); and underground railways 1.4 (21/15 mph).

[16] See "Commuter Journeys Slower than Before the War," David Millward, *Daily Telegraph*, 18th August 2015.

TABLE J.3:
Travel Speeds for Horse and Motor Omnibuses for 1907

	Route or Section of Route		Horse Omnibuses		Motor Omnibuses	
		Approximate Distance (Miles)	Time for Journey (minutes)	Speed (Miles Per Hour)	Time for Journey (minutes)	Speed (Miles Per Hour)
	Routes Through Central Areas					
1.	Liverpool Street to Wormwood Scrubs	7.2	70	6.2	56	7.7
	Wormwood Scrubs to Liverpool Street	7.3	74	5.9	61.5	7.1
2.	Bank to Shepherd's Bush	5.7	61	5.6	57.5	6
	Shepherd's Bush to Bank	5.7	62	5.5	45	7.6
3.	Oxford Circus to Kilburn	3.5	37	5.7	27.5	7.7
	Kilburn to Oxford Circus	3.5	34	6.2	26.5	8.1
4.	Bank to Putney	7	67	6.3	55	7.7
	Putney to Bank	7	75	5.6	61	6.8
5.	Bank to Hammersmith	6.5	61	6.4	49	8.0
	Hammersmith to Bank	6.4	61	6.3	56.5	6.8
	Average Speed			6.0		7.3
	Routes Through Outlying Areas					
6.	Clapham ("Plough") to Putney	3.9	37	6.3	27	8.6
	Putney to Clapham ("Plough")	3.9	40	5.8	26	8.9
7.	Shepherd's-Bush to Putney	3.2	31.5	6.1	20	9.6
	Putney to Shepherd's-Bush	3.2	29.5	6.4	20.5	9.4
	Average Speed			6.1		9.1
	Average Speed for all the Above Journeys			6.0		7.5

Source: London Statistics, 1907.

TABLE J.4:
Travel Speeds for Trams 1904

Type of Tram	Speed During Rush Hours (Miles Per Hour)	Speed During Slack Hours (Miles Per Hour)
Horse	2-5	5.5-8
Electric	5.5-7	8.5-11.5

Source: Royal Commission on London Traffic, 1904.

TABLE J.5:
Travel Speeds for Overground Railways for 1907 from 8-9am

Railway and Terminus	Number of inward suburban trains 8-9am	Average speed of inward suburban trains 8-9am
Great Central (Marylebone)	5	29.9
London and North Western (Euston)	7	28.7
South-Eastern and Chatham		
(London-Bridge)	3	28.5
(Holborn)	8	20.3
(Victoria)	9	19.0
(Charing Cross)	12	18.0
(Cannon Street)	10	17.1
(Ludgate Hill, St. Paul's, Moorgate-Street and Farringdon-Street)	16	16.4
Great Western (Paddington)	10	27.2
Midland		
(St. Pancras)	5	26.2
(Moorgate Street)	6	15.3
Great Northern		
(King's Cross)	16	23.3
(Moorgate-Street)	7	15.6
London, Tilbury and Southend (Fenchurch Street)	6	21.3
London and South-Western (Waterloo)	27	21.2
Great Eastern		
(Liverpool Street)	38	20.3
(Fenchurch Street)	14	16.0
London, Brighton and South Coast		
(London-Bridge)	28	20.2
(Victoria)	15	18.3
North London (Broad Street)	28	18.1

Source: London Statistics, 1907.

TABLE J.6:
Travel Speeds for Underground Railways for 1907

Railway	Total Length (miles)	Number of Intermediate Stations	Time for Journey (minutes)	Approximate Speed per Hour (miles)
Great Northern, Piccadilly and Brompton	8.9	19	32	16.7
Baker Street and Waterloo	4.25	9	16	15.9
Great Northern and City	3.44	4	13	15.9
Central London	5.75	11	23	15.0
Metropolitan and District				
(Inner Circle)	13	26	50	15.6
(Ealing-Whitechaptel)	12.74	24	51	15.0
City and South London	7.45	13	30	14.9
Charing Cross, Euston and Hampstead				
(Charing Cross - Golder's Green)	5.93	10	24	14.8
(Charing Cross - Highgate)	4.28	10	18	14.3

Source: London Statistics, 1907.

TABLE J.7:
Assumed Travel Time Weights for Each Transport Mode

Transport Mode	Travel Time Weight	Assumed Average Speed
Overground Railway	1	21 mph
Underground Railway	$1.4 = 21/15$	15 mph
Omnibus / Tram	$3.5 = 21/6$	6 mph
Walking	$7 = 21/3$	3 mph

We compute bilateral travel times between borough and parish centroids in ArcGIS using the travel time weights in Table J.7. We assume that workers can only interchange with the overground / underground railway network and omnibus/tram network at stations. We geolocated overground / underground railway stations using latitude and longitude coordinates and collected data on their year of opening/closing from a historical atlas of British railways (Cobb 2003). Comprehensive historical data are not available for the location of omnibus / tram stops for each year of our sample. Therefore, we approximate the location of omnibus / tram stops by assuming equally-spaced stops 1 km apart along all omnibus / tram lines. In solving for the least cost path between the centroids of each pair of boroughs and parishes, we allow workers to walk in a straight line from each centroid to any overground/underground railway station or omnibus/tram stop within 10 km, since the nearest station or stop need not be along the least cost path. When changing between different modes of transport, we assume that workers incur a travel time cost of 3 minutes.

J9 Commuting Data for Henry Poole Tailors

Prior to the 1921 census, there is no comprehensive data on commuting flows in London. To provide some evidence on commuting in the years before 1921, we consulted a number of company archives. While many archives contain lists of employees, most companies do not seem to have recorded the home addresses of their employees. An exception is Henry Poole Tailors, a high-end bespoke tailoring firm, which was founded in 1802, and has been located on London's Savile Row since 1828. Savile Row and its immediate surroundings have been a concentration of bespoke tailoring firms in London for several hundred years. The archives of Henry Poole contain ledgers with the names and home addresses of their employees going back to 1857.

This data has previously been used by the historian David Green in Green (1988). He kindly made available to us his raw data with his transcriptions of the ledgers surviving in the archives of Henry Poole. The data has a simple format. It lists the name of the employee and the year in which he or she joined the company. It also records a history of home addresses for each person. Unfortunately, for most address changes, no year is recorded. We therefore follow Green (1988) and only use the first reported home address and assume that it is the address at which the person lived in the year in which he or she joined Henry Poole Tailors.

The data is organised in two ledgers. The first ledger contains employees that joined the company between 1857 and 1877. For these employees, David Green did not transcribe the exact year in which they joined the company. The second ledger mainly contains employees who were hired after 1893 and David Green transcribed the data in this ledger up to 1914. However, this second ledger also contains a number of employees who in some cases were hired decades before the 1890s. For employees in this second ledger who were hired before the 1890s, it is uncertain whether the first recorded address is their home address at the time they were hired, or their home address at the time this second ledger was started. Therefore, we ignore these observations. We use the first ledger to construct commuting distances for workers who joined the company during a first twenty-year period from 1857 to 1877.[17] We use the second ledger to construct commuting distances for workers who joined the company during a second twenty-year period from 1891 to 1911.

David Green geolocated the home addresses of the Henry Poole employees in these two ledgers by hand on printed maps and then worked out the distances to the workshop of Henry Poole on Savile Row. The straight-line distances that he measured have unfortunately been lost. We therefore geolocated each address using Google Maps. In a large number of cases, these addresses do not exist any more today, because of street name changes or changes in London's street layout. To track street name changes, we used *Bruce's List of London Street Name Changes*, which contains a record of all street name changes in London between 1857 and 1966 collected by Bruce Hunt. A copy of this book is available at www.maps.thehunthouse.com. If addresses were located on streets that were renamed or no longer exist today for other reasons, the location of the address was traced with the help of the geo-referenced historical maps of London provided on www.oldmapsonline.org.

Of the employees contained in the first ledger covering the period 1857 to 1877, David Green geolocated the home addresses of 162 workers. We managed to geolocate 156 of these addresses.[18] For some addresses, only a street name and house number is provided, with no further information, and a given street name can appear multiple times in London. Therefore, for each address, we have documented how confident we are that we have correctly geolocated this address, by assigning a confidence level of low, low to medium, medium, medium to high, or high, depending on the further information provided in the ledger (such as the name of a suburb) and the frequency of use of the street name in London. In this first ledger, there are 135 addresses for which our confidence that the geolocation is correct is medium to high or high. We use these 135 addresses for our analysis, for which the median and 95th percentile commuting distances are 1.9 and 4.8 kilometers respectively.

Of the employees contained in the second ledger, David Green concentrated on those who were employed by Henry Poole Tailors up to and including 1899. We instead considered all 190 employees whose addresses David Green transcribed from the second ledger. Of these employees, the year in which they were hired by Henry Poole was missing for 10 names. We drop a further 18 names for employees who were hired before 1878 according to the second ledger and also appear in the first ledger. The first address reported for these workers differs across the two ledgers, suggesting that they were added to the second ledger at the time this ledger was created, and instead of transferring across their entire address history, only their address at the time they were added to the second ledger was transcribed.[19] There were a further 10 names for which we could not geolocate their address. We also drop any other workers who were hired before 1891 or after 1911. Of the remaining workers, we were able to geolocate 84 of these addresses with a medium-to-high or high confidence. We use these 84 addresses for our analysis, for which the median and 95th percentile commuting distances are 4.9 and 16.2 kilometers respectively.

[17] There are a number of employees who appear in both ledgers, as discussed further below.

[18] Of these 156 workers, Mr S. Codling reports living in Newcastle, and is probably a representative or salesmen of Henry Poole Tailors. Therefore, we drop him from the sample, as he cannot have commuted to London on a daily basis.

[19] We also drop John Blanchard Ash, who lives in Dover when he is hired by Henry Poole Tailors, and whose second recorded address is in Brussels. While we are not certain of his role, he is likely to have been a salesman or foreign representative who did not work in London. Finally, we also drop Rene Jones, who lives in Paris, and is described as an assistant to the Paris branch.

J10 Parameter Calibration

We calibrate the share of housing in household expenditure $(1 - \alpha)$ based on historical data for our sample period. In particular, we consulted the Parliamentary Paper entitled "Tabulations of the Statements Made by Men Living in Certain Selected Districts of London in 1887." Table D of this parliamentary paper reports the "Average Weekly Rent," "Average Weekly Earnings" and "Proportion of Earnings spent in Rent" for 29,451 working class men, grouped into 35 occupations, from four different districts in the North, South, East and West of London. The average of "Proportion of Earnings spent in Rent" reported across the 35 occupations is 0.25, with a minimum of 0.21 for "carman, carter" and a maximum of 0.33 for "hawkers and costermongers." Based on these figures, we set the share of housing in household expenditure equal to 0.25.

As a robustness check, we examined data from the New Dictionary of Statistics (Webb 1911). This publication reports data on wages for a number of different occupations, which comes originally from the "Abstract of Labour Statistics of the United Kingdom" (1908) and can be converted to shillings per week. This publication also reports average rents for tenements used by the working classes. These rents come originally from a 1908 inquiry by the Board of Trade and are reported in shillings per week for tenements of two to six rooms, respectively. Using the rent for a two and six room tenement, we can compute an upper and lower bound of the rent share for each occupation. Taking the simple average across the 17 occupations, the lower bound is 0.15 and the upper bound in 0.33. The midpoint of these is 0.24, which corroborates the average given in the Parliamentary Paper discussed above.

As a further robustness check, we consulted two additional historical sources. First, we examined the London budget survey data in Eichengreen (2010) ("New Survey of London Life and Labor, 1929-1931" (NSOL), ICPSR 8539), and calculated the share of rent in earnings using a more recent dataset of the NSOL data (UK Data Service, SN: 3758), which includes a much larger sample of the surviving records. These NSOL data have three main limitations for our analysis. First, they come from a time period after the end of our sample (1929-31). Second, they correspond to a random sample of working-class households in poorer Eastern boroughs of London, and hence it is unclear how representative these data are for all of London. Third, a growing secondary literature has raised data quality concerns about the sample survey data in the NSOL, including for example Abernethy (2013). Consistent with these concerns, we find values for rent as a share of income that are equal to 0 and more than 1. After excluding observations with rent as a share of income of less than 0.10 or more than 0.90, we find a mean share of rent in income of 21 percent, roughly in line with the average given in the Parliamentary Paper discussed above.

Second, we consulted Feinstein (1998), which provides figures for the share of rent in household expenditures for 1788/92, 1828/32 and 1858/62. The figure for 1788/92 is 0.1 and is taken from a budget survey carried out by Eden (1797) and Davies (1795). As these predate our period and represent a predominantly rural society, it is unclear how relevant these estimates are for our sample. The figures for 1828/32 and 1858/62 are taken from MacKenzie (1921). This study is chiefly occupied with the diets of representative households and estimating changes in living standards over time, in terms of caloric intake. As such, its focus was not the estimation of household rents. In the author's own words, MacKenzie (1921) states "from the data thus collected hypothetical budgets were constructed to show the possible standard attained by each representative family in each of the three years 1860, 1880 and 1914. Owing to the scarcity of data available no attempt at a quantitative measure of the changes has been made. The most one could hope for was to obtain an indication of the magnitude of the rise which has undoubtedly taken place in the standard of living of all classes." For this reason, we focus on the first two historical data sources discussed above.

We assume a value for the share of labor in production costs of $\beta^L = 0.55$, which lies in the middle of the range of 0.43-0.63 considered for the period 1770-1860 in Antrás and Voth (2003), and is close to the average share of labor in income of 0.56 over the period from 1856-1913 in Table 6.1 of Matthews, Feinstein and Odling-Smee (1982). We allocate the remaining 0.45 share of production costs between (i) machinery and equipment and (ii) land and building structures using the data from Table 6.1 of Matthews, Feinstein and Odling-Smee (1982). From 1856-1913, the average reported share of profits in income in this table is 0.27, which includes the return to investments in both machinery and equipment and building structures in the non-farm sector. Based on this figure, we assume a share of machinery and equipment in production costs of $\beta^M = 0.20$, which implies a share of land and building structures in production costs of $\beta^H = 0.25$, such that $\beta^L + \beta^M + \beta^H = 1$. This assumption implies a share of machinery and equipment in non-labor production costs of 44 percent (0.20 divided by 0.45), which lies close to the value of 45 percent reported for the aggregate economy (GDP) in Table 5 of Valentinyi and Herrendorf (2008).

J11 Data for Other Cities

In this section of the online appendix, we report additional information about the data sources and definitions for the reduced-form evidence on other cities reported in Section I5 of this online appendix.

J11.1 Paris

We define downtown Paris as the post-1860 districts ("arrondissements") 1 to 4, which are Louvre, Bourse, Temple and Hôtel-de-Ville. The boundaries of the arrondissements in Paris changed substantially in 1860 as part of the Haussmann reforms. To create a consistent population time series for our definition of downtown Paris back to 1800 we digitized a 1843 map by Xavier Girard ("Plan de la ville de Paris divise en 12 arrondissements, en 48 quartiers indiquant tout les changements faits et projetes," published by J. Goujoun and J. Andriveau). The map shows that each of the 12 per-1860 arrondissements consisted of four "quartiers". We map these quartiers to the boundaries of the post-1860 arrondissements. For the years before 1860 we approximate the sum of the population of the post-1860 arrondissements 1 to 4 with the sum of the population of the following 25 pre-1860 quartiers: Arcis, Arsenal, Banque, Bonne-Nouvelle, Cité, Feydeau, Hôtel de Ville, Île Saint Louis, Lombards, Louvre, Mail, Marais, Marché Saint Jean, Marchés, Mont de Piété, Montmartre, Montorgueil, Palais de Justice, Palais Royal, Porte Saint Denis, Saint Eustache, Saint Honoré, Saint Martin des Champs, Sainte Avoye, and Tuileries. Finally, we define the metropolitan area of Paris as the "petite courande" which consists of the departements of Paris, Hauts de Seine, Seine-Saint-Denis, and Val-de-Marne.

Data for the post-1860 arrondissements and the petite courande is available at http://www.demographia.com/db-paris-arr1999.htm. The data for the petite courande before 1860 and the pre-1860 quartiers was taken from the 1860 publication "Recherches statistiques sur la Ville de Paris et le départment De La Seine - Recueil de tableaux dressés et réunis d'aprés les ordres de G.-E. Haussmann". To collect the opening dates of all overground and underground rail stations in the petite courande, we examined the history of all underground and overground railway lines in the petite courande together with their stations using the French pages of Wikipedia. Our measure of the cumulative number of stations accounts for the opening and (very few) closings stations. If a location has both a underground and an overground railway stations, these are counted as separate stations in our measure of stations.

J11.2 Berlin

We define downtown Berlin as the six most central wards ("Stadtteile") of Berlin, which are Berlin, Köln, Friedrichswerder, Neukölln, Dorotheenstadt and Friedrichstadt.[20] The boundaries of these central wards remained essentially unchanged between 1800 and 1925. We define the metropolitan area of Berlin as Greater Berlin ("Gross Berlin" as defined in the "Gross Berlin Gesetz" of 1920), which includes the following districts: Charlottenburg, Cöpenick, Friedrichshain, Kreuzberg, Lichtenberg, Mitte, Neukölln, Pankow, Prenzlauer Berg, Reinickendorf, Schöneberg, Spandau, Steglitz, Tempelhof, Tiergarten, Treptow, Wedding, Weißensee, Wilmersdorf and Zehlendorf. Finally, we have obtained data for the pre-1920 administrative city of Berlin ("Stadt Berlin"), the area of which was divided into the post-1920 districts Mitte, Tiergarten, Wedding, Prenzlauer Berg, Friedrichshain and Kreuzberg.

Data for the population of the metro area of Berlin and the pre-1920 administrative city of Berlin are reported in the 2017 edition of the Statistical Yearbook for Berlin.[21] Leyden (1933) reports data for the wards of Berlin starting from 1867 until 1925. We have extended his time series with data for 1861 and 1803. The data for 1861 are reported in a summary table of the 1884 population census for Berlin ("Resultate der Berliner Volkszählung vom 3. December 1884 - Im Auftrage der städtischen Volkszählungs-Commission", p. 8) while the results for 1803 are reported in a historical overview in the 1890 population census for Berlin ("Die Bevölkerungs- und Wohnungs-Aufnahme vom 1. Dezember 1890 in der Stadt Berlin - Im Auftrag der Städtischen Deputation für Statistik", page XIX).[22]

Data on employment at the place of residence in each of the 20 post-1920 districts of Berlin is reported in the 1925 employment census ("Volks-, Berufs-, Betriebsählung vom 16. Juni 1925: Berufszählung - Die berufliche und soziale Gliederung der Bevölkerung im den Ländern und Landesteilen, Heft 3, Stadt Berlin", Statistik des Deutschen Reiches, Band 403). Data for 1925 on employment at the workplace in each post-1920 districts of Berlin is reported in the 1925 firm census ("Volks-, Berufs-, Betriebsählung vom 16. Juni 1925: Gewerbliche Betriebszählung - Die gewerblichen Niederlassungen und die technischen Betriebseinheiten in den Ländern und Landesteilen, Ost- und Mitteldeutschland", Statistik des Deutschen Reiches, Band 415). Workplace employment for the pre-1920 administrative city of Berlin is reported in the equivalent publications of the 1907, 1895, 1882 and 1875 firm censuses. To collect the opening dates of all overground and underground rail stations in Greater Berlin, we examined the history of all underground and overground railway lines in Greater Berlin using the German pages of Wikipedia.

J11.3 Boston

Our definition of downtown Boston follows the American Community Survey definition, which includes Chinatown and the Leather District, and includes the following 2010 census tracts: 070200, 070101, 030300 and 020303. Our definition of the metropolitan area of Boston follows the U.S. Census definition of the Boston-Cambridge-Newton, MA-NH Metro Area, including the following counties: Suffolk MA, Norfolk MA, Plymouth MA, Middlesex MA, Essex MA, Worcester MA, Bristol MA, Barnstable MA, Windham CT, Rockingham NH, Strafford NH, Merrimack

[20] The most central ward of Berlin is called Berlin itself.

[21] The 2017 Yearbook for Berlin does not report data for the metro area of Berlin before 1816 but only data for the pre-1920 administrative city of Berlin. To estimate the population of the metro area of Berlin in 1803 we use the ratio of the population of the pre-1920 administrative city of Berlin and the metro area of Berlin in 1816, which is 0.88, to scale up the population of the pre-1920 administrative city of Berlin to the population of the metro area in 1803.

[22] The 1890 population census for Berlin on page XVII discusses that the ward level data for the years immediately after 1803 have been lost and from 1816 Berlin counted population at the level of police districts rather than wards. The boundaries of the police districts unfortunately did not overlap with the wards and were changed substantially over the next decades. As a results ward level population data is only available from 1861 onwards again.

NH, Belknap NH, Hillsborough NH, Bristol RI, Kent RI, Newport RI, Providence RI, and Washington RI. We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the Boston metro area.

Data on the average length of the journey-to-work (in miles) for a sample of attorneys with offices in Boston at 15-year intervals from 1911-1971, as reported in Jackson (1987). The data are from the Boston City Directories for 1911, 1926, 1941, 1956 and 1971. The sample was chosen by Jackson (1987) by taking every tenth attorney in the directory for a given year until a total of 76-78 attorneys was reached.

J11.4 Chicago

We define downtown Chicago as Chicago Loop, including the following 2010 census tracts: 839000, 839100, 320400, 320600 and 320100. Our definition of the metropolitan area of Chicago follows the U.S. Census definition of the Chicago-Naperville-Elgin, IL-IN-WI Metro Area, including the following counties: Cook IL, DeKalb IL, Du Page IL, Grundy IL, Kane IL, Kendall IL, McHenry IL, Will IL, Lake IL, Jasper IN, Lake IN, Newton IN, Porter IN, and Kenosha WI. We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the Chicago metro area.

J11.5 New York

We define downtown New York as the parts of Lower Manhattan with the longest histories of European settlement, including all 2010 census tracts South of a line following Canal Street: 000201, 000202, 000600, 000700, 000800, 000900, 001001, 001002, 001200, 001300, 001401, 001402, 001501, 001502, 001600, 001800, 002000, 002100, 002201, 002202, 002400, 002500, 002601, 002602, 002700, 002800, 002900, 003001, 003002, 003100, 003200, 003300, 003400, 003601, 003602, 003700, 003800, 003900, 004000, 004100, 004200, 004300, 004500, 004700, 004900, 005501, 005502, 005700, 005900, 006100, 006300, 006500, 006700, 006900, 007100, 007300, 007500, 007700, 007900, 031703, 031704, and 031900.

Our definition of the metropolitan area of New York follows the U.S. Census definition of the New York-Newark-Jersey City, NY-NJ-PA Metro Area, including the counties of New York NY (Manhattan), Bronx NY, Kings NY (Brooklyn), Queens NY, Richmond NY, Fairfield CT, Litchfield CT, New Haven CT, Bergen NJ, Essex NJ, Hudson NJ, Hunterdon NJ, Mercer NJ, Middlesex NJ, Monmouth NJ, Morris NJ, Ocean NJ, Passaic NJ, Somerset NJ, Sussex NJ, Union NJ, Dutchess NY, Nassau NY, Orange NY, Putnam NY, Rockland NY, Suffolk NY, Ulster NY, Westchester NY, and Pike PA.

We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from

Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the New York metro area.

Data on the average length of the journey-to-work (in miles) for a sample of attorneys with offices in Manhattan from 1825-1973 are reported in Jackson (1987). The samples were taken from Longworth's City Directories from 1825-1875, from Trow's New York City Directory from 1888-1917, from the New York City Directory of Lawyers for 1928, and the Columbia Law Register from 1938-1873.

J11.6 Philadelphia

Our definition of downtown Philadelphia follows that of the Center City District and Central Philadelphia Development Corporation, including the following 2010 census tracts: 001002, 001001, 000402, 000401, 001102, 000902, 000901, 000100, 000200, 000300, 000500, 000600, 000700, 000804, 001202, 001201, 000803, 000801, and 001101.

Our definition of the metropolitan area of Philadelphia follows the U.S. Census definition of the Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area, including the counties of Philadelphia PA, Berks PA, Bucks PA, Chester PA, Delaware PA, Montgomery PA, Kent DE, New Castle DE, Cecil MD, Atlantic NJ, Burlington NJ, Camden NJ, Cape May NJ, Cumberland NJ, Gloucester NJ, and Salem NJ.

We construct downtown population from 1880-1940 using the consistent historical time-series on the population of 2010 census tracts constructed by Lee and Lin (2018), where 1880 is the earliest year for which these data are available. To construct metro area population from 1880-1940, we first collapse the individual-level population census data from Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018) to the county level, before aggregating counties within the Philadelphia metro area.

Data on the fraction of workers commuting different distances (in miles) to workplaces in central Philadelphia from 1850-1880 are reported in Hershberg (1981). The data include the following occupations: Blacksmiths, Bookbinders, Cabinetmakers, Carpenters, Confectioners, Lawyers, and Physicians. The data were compiled by the Philadelphia Social History Project (PSHP) from the manuscript schedules of the U.S. Census of Population, the manuscript schedules of the U.S. Census of Manufacturing, City Business Directories, and City Street Directories.

References

Abernethy, Simon T., "Deceptive data? The New Survey of London Life and Labour 1928-31," CWPESH Working Paper No. 16, 2013.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83 (2015), 2127-2189.

Allen, Treb, and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129 (2014), 1085-1140.

Allen, Treb, Costas Arkolakis, and Xiangliang Li, "Optimal City Structure," Yale University, mimeograph, 2017.

Antrás, Pol, and Hans-Joachim Voth, "Factor Prices and Productivity Growth During the British Industrial Revolution," Explorations in Economic History, 40 (2003), 52-77.

Armington, Paul S., "A Theory of Demand for Products Distinguished by Place of Production," IMF Staff Papers, 16 (1969), 159-178.

Barker, T., "Towards an Historical Classification of Urban Transport Development Since the Later Eighteenth Century," Journal of Transport History, 1 (1980), 75-90.

Barker, T. C., and M. Robbins, A History of London Transport: The Nineteenth Century, Vol. 1 (London: George Allen and Unwin, 1963).

Cobb, Michael H., The Railways of Great Britain: a Historical Atlas, (Shepperton : Ian Allan Publishing, 2003).

Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein, Introduction to Algorithms, 3rd ed. (Cambridge, MA: MIT Press, 2009).

Crafts, Nicholas F. R., and Charles K. Harley, "Output Growth and the British Industrial Revolution: a Restatement of the Crafts-Harley View," Economic History Review, 45 (1992), 703-730.

Croome, Desmond F., and Alan A. Jackson, Rails Through the Clay: a History of London's Tube Railways, (London: Capital Transport Publishing, 1993).

Davies, D., The Case of Labourers in Husbandry, (London: Robinson, 1795).

Eaton, Jonathan, and Samuel S. Kortum, "Technology, Geography, and Trade," Econometrica, 70 (2002), 1741-1779.

Eden, F. M., The State of the Poor, (London: B. and J. White, 1797) .

Eichengreen, Barry, New Survey of London Life and Labor 1929-1931, ICPSR 8539 (Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2010).

England and Wales, "Integrated Census Microdata, I-CeM, 1851-1911," in UK Data Service, 2014, SN: 7481, K. Schurer, and E. Higgs ed., 1851, http://doi.org/10.5255/UKDA-SN-7481-1.

England and Wales, "Integrated Census Microdata, I-CeM, 1851-1911," in UK Data Service, 2014, SN: 7481, K. Schurer, and E. Higgs ed., 1881, http://doi.org/10.5255/UKDA-SN-7481-1.

England and Wales, Census of England and Wales: County of London, (London: His Majesty's Stationary Office, 1921).

Feinstein, Charles H., "Pessimism Perpetuated: Real Wages and the Standard of Living in Britain during and after the Industrial Revolution," Journal of Economic History, 58 (1998), 625-658.

Gerhold, D., Carriers and Coachmasters: Trade and Travel Before the Turnpikes, (Bodmin, Cornwall: Phillimore, 2005).

Great Britain, "Integrated Census Microdata (I-CeM), 1851-1911," in UK Data Service, 2014, SN: 7481, K. Schurer and E. Higgs, 1911, http://doi.org/10.5255/UKDA-SN-7481-1.

Green, David R., "Distance to Work in Victorian London: A Case Study of Henry Poole, Bespoke Tailors," Business History, 30 (1988), 179-194.

Helpman, Elhanan, "The Size of Regions," in Topics in Public Economics: Theoretical and Applied Analysis, D. Pines, E. Sadka, and I. Zilcha, ed. (Cambridge University Press, Cambridge, 1998).

Herschberg, Theodore, Philadelphia: Work, Space, Family, and Group Experience in the Nineteenth Century, (Oxford: Oxford University Press, 1981).

Jackson, Kenneth T., Crabgrass Frontier: The Suburbanization of the United States, (Oxford: Oxford University Press, 1987).

Kaplan, Greg, and Giovanni L. Violante, "A Model of the Consumption Response to Fiscal Stimulus Payments," Econometrica, 82 (2014), 1199-1239.

Kellet, John R., The Impact of Railways on Victorian Cities, (London: Routlege and Kegan Paul, 1969).

Krugman, Paul, "Increasing Returns and Economic Geography," Journal of Political Economy, 99 (1991), 483-499.

Lee, Sanghoon, and Jeffrey Lin, "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income," Review of Economic Studies, 85 (2018), 663-694.

Leyden, Friedrich, Gross-Berlin: Geographie einer Weltstadt, (Berlin: Gebr. Mann Verlag, 1933) .

London County Council, London Statistics, (London: London County Council, 1907).

London County Council, London Statistics, (London: London County Council, 1921).

Lucas, Robert E., and Esteban Rossi-Hansberg, "On the Internal Structure of Cities," Econometrica, 70 (2002), 1445-1476.

Mackenzie, W. A., "Changes in the Standard of Living in the United Kingdom, 1860-1914," Economica, 3 (1921), 211-230.

Matthews, Robert C. O., Charles H. Feinstein, and John C. Odling-Smee, British Economic Growth: 1856-1973, (Oxford: Oxford University Press, 1982).

Minnesota Population Center, "Integrated Public Use Microdata Series, International," Version 7.1, Minneapolis, MN: IPUMS, 2018, https://doi.org/10.18128/D020.V7.1.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," American Economic Review, 108 (2018), 3855-3890.

Obstfeld, Maurice, and Kenneth Rogoff, Foundations of International Macroeconomics, (Cambridge MA: MIT, 1996).

Parliamentary Papers, "Tabulations of the Statements Made by Men Living in Certain Selected Districts of London in 1887," (London: House of Commons, 1887).

Redding, Stephen J., "Goods Trade, Factor Mobility and Welfare," Journal of International Economics, 101 (2016), 148-167.

Redding, Stephen J. and Daniel M. Sturm, "The Costs of Remoteness: Evidence from German Division and Reunification," American Economic Review, 98 (2008), 1766-1797.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA," Minneapolis, MN: IPUMS, 2018.

Saiz, Albert, "The Geographic Determinants of Housing Supply," Quarterly Journal of Economics, 125 (2010), 1253-1296.

Salmon, James, Ten Years' Growth of the City of London, Report, Local Government and Taxation Committee of the Corporation, (London: Simpkin, Marshall, Hamilton, Kent and Company, 1891).

Santos Silva, J., and S. Tenreyro, "The Log of Gravity," Review of Economics and Statistics, 88 (2006), 641-658.

Stamp, Josiah, British Incomes and Property: The Application of Official Statistics to Economic Problems, (London: P. S. King and Son, 1922).

Valentinyi, Ákos, and Berthold Herrendorf, "Measuring Factor Income Shares at the Sectoral Level," Review of Economic Dynamics, 11 (2008), 820-835.

Webb, Augustus D., The New Dictionary of Statistics, (London: Routledge and Sons 1911).

Wrigley, Edward Anthony, The Early English Censuses, (Oxford: Oxford University Press, 2011).