JUE Insight: Migration, Transportation Infrastructure, and the Spatial Transmission of COVID-19 in China*

Bingjing Li † Lin Ma ‡

April 5, 2021

Abstract

This paper evaluates the impacts of migration flows and transportation infrastructure on the spatial transmission of COVID-19 in China. Prefectures with larger bilateral migration flows and shorter travel distances with Hubei, the epicenter of the outbreak, experienced a wider spread of COVID-19. In addition, richer prefectures with higher incomes were better able to contain the virus at the early stages of community transmission. Using a spatial general equilibrium model, we show that around 28% of the infections outside Hubei province can be explained by the rapid development in transportation infrastructure and the liberalization of migration restrictions in the recent decade.

Keywords: COVID-19; spatial transmission; migration; transportation infrastructure; general equilibrium spatial model

JEL Classification: I10, I15, R10, O18, R12, R13,

*We thank the editor and two anonymous referees for their comments. We also thank Yang Xu for his excellent research assistance, Davin Chor and EeCheng Ong for helpful discussions, and the seminar participants in the Singapore University of Social Sciences for feedback. All the remaining errors are our own.

†Department of Economics, National University of Singapore. Email: bingjing.crystal@gmail.com.

‡Department of Economics, National University of Singapore. Email: malin84@gmail.com.

1 Introduction

The spatial transmission of COVID-19 in mainland China is unprecedented. Following the initial report of the novel coronavirus in Wuhan, 262 cities in 30 provinces reported cases of COVID-19 within the next 28 days. By the end of our sample period — February 22, 2020 — the number of infections outside Hubei province had reached 12,526. Part of the high transmissibility of COVID-19 is undeniably due to biological reasons (Petersen et al., 2020); nevertheless, it remains an open question as to what extent have socioeconomic factors, i.e., the unprecedented ease at which people travel and commute over long distances in China, contributed to the spread of the disease. Recognizing the links between connectivity and disease transmission broadens our understanding of the impacts of factor mobility, one of the central topics of interest in spatial economics. In this paper, we set out to answer this question.

The improved transportation infrastructure and liberalized migration policy, among many others, are the potential forces behind the increased mobility of people in the recent decades in China. The transportation infrastructure has expanded rapidly, as dense networks of roads, railways, and airports have significantly reduced travel distance. Ma and Tang (2020a) estimate that the average costs of passenger transportation have declined by around 70 percent between 1995 and 2015. The reduction in commuting costs not only increases the frequency of travel but also lowers the costs of medium- and long-term migrations. Meanwhile, the reform of the household registration system (*hukou*) has gradually lowered migration barriers in China (Tombe and Zhu, 2019; Fan, 2019). Many cities have relaxed the requirements to obtain local hukou, which improved the employment prospects of the migrants and, at the same time, elevated their access to public services such as education, healthcare, and social security. The steady decline in migration barriers and the improved transportation infrastructure have induced a phenomenal rise in internal migration. Gross migration flows rose from 64.5 million in 2000 to 129.0 million in 2015; gross flows specific to Hubei more than doubled from 4.2 million to 10.3 million during this period.

The changes in transportation networks and migration patterns could have played important roles in shaping the spread of COVID-19. The onset of the outbreak of COVID-19

was in the run-up to the Spring Festival, the period of travel fest expecting about 3 billion trips (Bloomberg News, 2020). The population outflow from Wuhan amounted to 4.3 million two weeks before the city-wide lockdown on January 23, 2020. Figure 1 shows the residual scatter plots of a multivariate regression of outflows from Wuhan to different prefectures in the two weeks before the lockdown (January 9–22, 2020). Cities with more emigrants to and more immigrants from Wuhan record greater outflows from Wuhan, reflecting that family reunions are the primary reasons for travel during the Spring Festival. In addition, the partial correlation of population outflow and travel distance is negative, which suggests that in addition to the movement of long-term migrants, short-term population movement, e.g., work-related travel, comprises a significant proportion of all trips.

We evaluate the role of the transportation infrastructure and the reduction in migration barriers in the context of the COVID-19 transmission in China. Specifically, we ask: without the recent changes in the transportation networks and migration policies, how would the transmission of COVID-19 be affected? In these counterfactual experiments, we hold constant the public health measures implemented during the COVID-19 pandemic. Our setting is unique. The spatial spread during our sample period originated from a single epicenter in Hubei (Jia et al., 2020). Due to the stringent public health measures and travel restrictions, there were few cross-transmissions among the regions outside the epicenter. In light of this pattern, our empirical focus is on the spatial relations specific to Hubei even though our spatial model accounts for all bilateral linkages.

We combine a disease transmission model and a general equilibrium spatial model incorporating trade in goods and migration flows, and conduct the analysis in three steps. First, guided by the viral transmission model, we find that prefectures with larger bilateral migration flows and shorter travel distances with Hubei experienced a greater spread of COVID-19. However, these factors affected only transmissions in the early stages when most cases were imported, indicating the travel ban's effectiveness and other measures restricting potential social interactions of return-migrants and visitors from Hubei with the local population. Local economic activities also influenced the speed of transmission, with two counteracting mechanisms. Prefectures with greater economic activities received more imported cases; however, higher-income prefectures were better able to contain the virus in

the early stages of community transmission.

In the second step, based on the spatial economic model, we quantify the effects of the expanding transportation network and the reduction in migration barriers over the period 2005–15 on migration flows, the spatial distribution of population, and income. The indirect general equilibrium effect on income is part of the total impact of the counterfactual policy shock, affecting both aggregate transmissions and spatial patterns. We find that had the transportation network reverted to the 2005 configuration, the travel distance with Hubei would increase by nearly 90 percent for the average prefecture, while the total population flow in and out of Hubei would decline by 14 percent. The reversion of the migration policy to the 2005 configuration would reduce the Hubei-related migration flow by 57 percent.

In the third and final step, we bring together the counterfactual changes in migration flows, population distribution, and income with the elasticities of incidence to these underlying variables, and simulate the counterfactual changes in the spread of COVID-19 outside Hubei. We find that the number of infections would have been lower by 15.31% by February 22, 2020, had there been *no* expansions in transportation networks between 2005 and 2015. The transportation infrastructure affects the spread of COVID-19 mainly by altering migration flows related to Hubei and short-term population movement; the quantitative effect of migration flows is around one-fourth that of short-term population movement. The counterfactual change in migration policies had a similar quantitative effect. On its own, the reversion of migration barriers to the 2005 configuration would have lowered the number of infections by 17.82%. If *both* transportation networks and migration policies had reverted to their 2005 configurations, the spread of COVID-19 would have been reduced by 28.21%. These findings indicate that the swift spatial spread of COVID-19 is partly facilitated by the tighter inter-regional linkages induced by the expanded transportation infrastructure and the reform in migration policies over the past two decades.

Given the low number of infections in China, the healthcare costs of better connectivity are likely to be orders of magnitude smaller than its economic benefits. Under our model, reverting the transportation networks and migration policies to 2005-levels would reduce the aggregate income by 3.60%, which equals to $321 billion, based on the estimates of Chinese GDP from the World Bank. On the other hand, the 28.21% reduction of the

incidences from reverting infrastructure and migration policies would lead to 3,517 fewer infections, 132 fewer hospitalizations, and 23 fewer fatalities, based on the estimates of hospitalization and fatality rates in Walker et al. (2020) and Verity et al. (2020). The costs of these hospitalizations and fatalities are between $35 and $173 million, depending on the estimates of the values of a statistical life as in Ashenfelter and Greenstone (2004) and Viscusi and Aldy (2003)[1]. However, the low costs of better mobility critically depend on the fact that the disease was efficiently controlled in China. Without effective containment policies, the number of infections would have been much higher, and so would the healthcare costs of better connectivity. For example, if China has 100 million cases of COVID-19, the economic costs of a 28.21% change are between $283 billion and $1.40 trillion, which are on par with the estimated benefit of better mobility (see Appendix B for details). One hundred million cases in China is not unimaginable; it puts China at a 7% population infection rate, similar to that in the U.S. in January 2021 (WHO, 2021). With these cautions in mind, we argue that the unintended and potentially fatal consequences of factor mobility should no longer be overlooked in the long and flourishing literature on transportation economics (Fogel, 1962; Allen and Arkolakis, 2014; Donaldson and Hornbeck, 2016; Donaldson, 2018; Allen and Arkolakis, 2019).

This study contributes to the literature on the health costs of transportation infrastructure, and more generally, to the long-run economic determinants of the transmission of disease. Adda (2016) employs quasi-experimental variation and a difference-in-differences design to evaluate the role of public transportation and expanding railways in France on viral transmission. We take a different approach by employing a quantitative spatial model that characterizes how transportation costs and migration barriers shape spatial links among prefectures to determine the spread of COVID-19 from Hubei. Our approach enables the computation of the national-level general equilibrium effects of shocks to economic fundamentals while relies more on the model's structure. The literature on COVID-19 also investigates the association between population mobility and spatial spread. Most of these studies focus on projecting the impacts of travel restrictions (Chinazzi et al., 2020), assessing community transmission risk (Jia et al., 2020), and evaluating the effectiveness of transmission control

[1]Appendix B provides more details on the cost-benefit estimation.

measures in containing the spread (Jia et al., 2020; Kraemer et al., 2020; Tian et al., 2020). In contrast, our study explores the roles of transportation infrastructure and migration policies — the fundamentals that determine population mobility — on disease transmission through the lens of a spatial economic model.[2]

Our work is also related to a broader literature that explores the propagation of shocks to economic fundamentals through spatial linkages. Allen and Arkolakis (2014) and Allen et al. (2020) propose a series of spatial general equilibrium models to study the interactions of goods and factor mobility. In the context of the Ricardian models, Caliendo et al. (2018, 2019) analyze the transmission of trade and migration shocks in a similar setup to our model. We highlight that in addition to the direct economic impacts usually documented in the literature, the mobility of people has an unintended spatial impact through disease transmission.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 examines the roles of migration flows, travel distance, and local economic activities on the spread of COVID-19 outside Hubei. Section 4 lays out a general equilibrium spatial model that computes the aggregate effects of counterfactual changes in transportation networks and migration policies. Section 5 quantifies the model, and Section 6 presents the counterfactual experiments. Section 7 concludes.

2 Data

Prefecture-level Data on COVID-19 Cases We collected prefecture-level data on reported COVID-19 cases with daily frequency from the Health Commissions of different prefectures. We exclude the data of the epicenter Hubei given that our study focuses on the spatial spread of the disease outside Hubei.[3] Our baseline analysis covers the period from January 28 to February 22, 2020 — 30 days after the lockdown of Hubei, when the spread was almost halted, as shown in Figure A.1 in the appendix. By then, there were 12,526

[2]The existing economic literature establishes the links between epidemics and economic activities, such as international trade (Oster, 2012), business cycles (Adda, 2016), and social networks (Fogli and Veldkamp, 2019).

[3]The exclusion is also due to the possibility that the data for Hubei in the early stages of the outbreak may underestimate the actual prevalence of infection.

reported cases of infections located across 267 prefectures outside Hubei.

Migration Flows The bilateral migration data come from a 10% subsample of the *1% Population Sampling Survey of China* (mini census) in 2015. The mini census data contains information on prefecture of residence and prefecture of *hukou* registration, based on which we code migration status and calculate bilateral migration flows. We employ the following prefecture-level measures for the empirical analysis: (i) the ratio of emigrants to Hubei to the local population in a particular prefecture, and (ii) the share of immigrants from Hubei in the local population in a particular prefecture. For the quantification analysis, we also employ the data on migration flows in 2005 from a 20% subsample of the *1% Population Sampling Survey of China* in 2005.

Transportation Networks The transportation network data come from Ma and Tang (2020a), which constructs the transportation networks from the digitized transportation maps that incorporate roads, railways, high-speed railways, and waterways. The distance is measured as the time required to travel between two points. We use the data from 2005 and 2015 in this paper, which are visualized in Figure A.2.

Other Data Sources We have used the following datasets in the quantification stage in Section 5. We use the *Investment Climate Survey* from the World Bank to calibrate the parameters related to internal trade. The *Population Census* in 2000 and 2010 were used to measure the initial population distribution.

3 Empirical Framework and Results

This section lays out an empirical model linking the spatial transmission of COVID-19 to economic fundamentals. Specifically, we consider the number of infections in a locality as a function of bilateral population flows with Hubei, which are determined by the bilateral long-term migration pattern with Hubei, the travel distance, and the size of the local economy. Guided by the disease transmission model in Appendix C, we estimate the following equation:

$$I_i(t) = \exp\Big(\beta_{0t} + \beta_{1t}X_i + \beta_{2\tau}M_i + \beta_{3t}\ln(Dist_i) + \beta_{4t}\ln(Pop_i) + \beta_{5t}\ln(GDPpc_i) + \nu_i(t)\Big), \quad (1)$$

where $I_i(t)$ is the number of infections at time t; X_i denotes the ratio of Hubei-bound emigrants to the local population in prefecture i, and M_i is the share of immigrants from Hubei in the local population in prefecture i; $Dist_i$ measures the travel distance between prefecture i and Hubei based on transportation networks in 2015;[4] and Pop_i and $GDPpc_i$ represent population size and GDP per capita in 2015, respectively. Equation (1) is estimated by a Poisson quasi-maximum likelihood count model (Wooldridge, 1999), with robust standard errors clustered by prefecture.

The time-varying coefficients $\beta_{\iota t}$ represent the cumulative effects of the underlying variables up to period t. We set the starting date $t = 0$ to January 28, 2020, five days after the lockdown was imposed in Wuhan and other cities in Hubei. At this time, most imported cases would have passed the incubation period, and would have been recorded. Therefore, the estimates $\beta_{\iota 0}$ reveal the effects of the underlying variables on the arrivals of imported cases. For the baseline analysis, we include in the sample the observations from January 28 to February 22, 2020, with time intervals of five days.[5] The differences in the coefficients, $\beta_{\iota t} - \beta_{\iota t-\delta}$, capture the effects of the underlying variables on the local transmission within an incubation period $[t - \delta, t]$, which also reflects the effectiveness of the prevention and control policies. For example, when travelers from Hubei are subjected to quarantine orders, $\beta_{1t} - \beta_{1t-\delta}$ and $\beta_{2t} - \beta_{2t-\delta}$ are expected to be zero in the later periods.

3.1 Empirical Results

Figure 2 reports the point estimates of $\beta_{\iota t}$ and their 90% confidence intervals. Appendix E demonstrates that the baseline findings are robust to a variety of alternative specifications.

We find in Panel A that the prefectures with a higher share of Hubei-bound emigrants have on average more cases of COVID-19 infection. However, the cumulative effect remains

[4]The travel distance is the simple average of the distances between prefecture i and all prefectures in Hubei.

[5]Based on the findings in the epidemiological literature on COVID-19 (Guan et al., 2020; Li et al., 2020), we set the incubation period to five days.

stable over the sample period. This finding suggests that while cases were imported when Hubei-bound emigrants returned home for the Spring Festival, such imported cases did not engender further community transmissions in the later periods, perhaps due to the effective quarantine measures that were implemented.

As shown in Panel B, a higher share of immigrants from Hubei is also associated with a wider spread of the disease. Additionally, the associated imported cases resulted in local transmissions over an earlier period between January 28 and February 2, 2020, but the effect quickly diminished afterward. Panel C presents the estimates for the distance to Hubei, which reflect the effects of short-term population movement, such as business trips, before the lockdown in Hubei on disease transmission in the subsequent periods. As expected, prefectures closer to the epicenter had more imported cases at the start of the period, but the distance does not affect subsequent transmissions, consistent with the travel restrictions from and to Hubei. We observe a similar pattern for population size in Panel D. The lack of correlation between population size and the number of local transmissions indicates the effectiveness of a range of public health interventions aimed at minimizing interpersonal contact (Jia et al., 2020; Kraemer et al., 2020; Tian et al., 2020).

Last but not least, Panel E shows the effects of GDP per capita. *Ceteris paribus*, prefectures with higher incomes reported more imported cases due to tighter economic relationships with Hubei. Interestingly, as shown in Figure A.3, in an earlier period between January 28 and February 7, 2020, a higher income per capita was associated with a slower spread of the disease, indicating that higher-income regions were more capable of implementing transmission control measures promptly. Lower-income prefectures caught up in the later period, though, and the incidence rate as of February 12, 2020, was no longer correlated with income level.

In Section 6, we take the estimates of the underlying parameters β's as given and quantify the impacts of different counterfactual configurations of transportation networks and migration policies on the transmission of COVID-19. A change in transportation networks alters travel distance and migration flows, as well as spatial distributions of population and income across China through general equilibrium effects. Our regression analysis indicates that all these factors have independent effects on disease transmission. The following section

introduces a quantitative spatial model that computes the aggregate effect of counterfactual changes in transportation infrastructure and migration costs.

4 The Model

Our model is drawn from Ma and Tang (2020a), which extends Tombe and Zhu (2019) to allow for productivity agglomeration. The economy contains a mass $\bar{L} > 0$ of individuals and J cities indexed by $j = 1, 2, ..., J$. Individuals can migrate between the J cities within China subject to frictions. Individuals living in city j obtain utilities according to the following CES function:

$$U_j = \left(\int_0^1 \left(q_j(\omega) \right)^{\frac{\eta-1}{\eta}} d\omega \right)^{\frac{\eta}{\eta-1}}, \tag{2}$$

where ω indexes the goods and η is the elasticity of substitution.

The production side of the model follows Eaton and Kortum (2002): firms operate in perfectly competitive markets, and every city can produce every variety of ω. The production function for variety ω in city j is:

$$q_j(\omega) = A_j \cdot z_j(\omega) \cdot \ell_j,$$

where ℓ_j is the labor input. A_j is the city-specific productivity that depends on an exogenous component, $\bar{A}_j$, and the population to allow for agglomeration:

$$A_j = \bar{A}_j \cdot (L_j)^{\beta}, \tag{3}$$

where β is the agglomeration elasticity. The city-variety specific productivity, $z_j(\omega)$, is from an *i.i.d.* Frechet distribution with parameter θ:

$$F(z) = \exp\left[-(z)^{-\theta} \right].$$

Trade is subject to iceberg costs: for a unit of product to arrive in city i from city j, $\tau_{ij} > 1$

units of goods need to be produced and shipped.

The consumers in city i purchase from the supplier offering the lowest price for every variety ω:

$$p_i(\omega) = \min_{j=1,\cdots,J+1} \{p_{ij}(\omega)\},$$

where $p_{ij}(\omega)$ is the price of ω from city j at the market in city i.

4.1 Migration

Individuals decide on migration destinations to maximize utility. Denote V_j as the indirect utility of living in city j:

$$V_j = \frac{w_j}{P_j}, \tag{4}$$

where w_j is the nominal wage and P_j is the ideal price index in city j:

$$P_j = \left(\int_0^1 \left(p_j(\omega)\right)^{1-\eta} d\omega \right)^{\frac{1}{1-\eta}}.$$

In addition to the indirect utility, each worker also draws an idiosyncratic location preference for each city $\{e_j\}_{j=1}^J$ from an *i.i.d.* Frechet distribution with the CDF:

$$F(e_j) = \exp\left[-\left(e_j\right)^{-\kappa}\right],$$

where κ is the shape parameter. Lastly, moving from j to i also incurs a pair-specific cost, $\lambda_{ij} \geq 1$ with $\lambda_{ii} = 1$. If a worker moves from city j to i, the utility in the end is the combination of the location preference and the migration costs:

$$\frac{V_i \cdot e_i}{\lambda_{ij}}.$$

The costs of migration capture the financial costs of moving and commuting, the psychological costs of living in an unfamiliar environment, as well as the policy barriers that deter

migration, such as the *hukou* system.

Considering all the determinants of migration, a worker living in city j will migrate to city i if and only if doing so provides her with the highest utility among all the J locations:

$$\frac{V_i \cdot e_i}{\lambda_{ij}} \geq \frac{V_k \cdot e_k}{\lambda_{kj}}, \forall k = 1, 2, ..., J.$$

Conditional on V_j, the probability of an individual migrating from city j to i is:

$$m_{ij} = \frac{(V_i)^{\kappa} (\lambda_{ij})^{-\kappa}}{\sum_{m=1}^{J} (V_m)^{\kappa} (\lambda_{mj})^{-\kappa}}. \tag{5}$$

This probability is also the fraction of the individuals who migrate from city j to i due to the law of large numbers. Therefore, the migration flow from city j to city i is:

$$L_{ij} = (V_i)^{\kappa} (\lambda_{ij})^{-\kappa} (\Pi_j)^{-\kappa} \bar{L}_j, \tag{6}$$

where Π_j is the expected utility of a worker who lives initially in j:

$$\Pi_j = \left[\sum_{m=1}^{J} (V_m)^{\kappa} (\lambda_{mj})^{-\kappa} \right]^{\frac{1}{\kappa}}.$$

4.2 The Equilibrium

Given the parameters of the model, the equilibrium is defined as a vector of prices $\{w_j, p_j(\omega)\}$, a vector of quantities $\{q_j(\omega)\}$, and a population distribution $\{L_j\}$ such that:

- Every individual maximizes his utility by choosing the location and the consumption bundle.
- Every firm maximizes its profit.
- The labor market in each location clears.
- Trade is balanced.

Appendix D provides details of the equilibrium conditions.

5 Quantification

We quantify the model to 291 prefecture-level cities in China around the year 2015. The sample is determined by data availability, and we focus on the year 2015 as it is the latest year in which the *1% Population Sampling Survey* is available. The quantification strategy aims to capture the migration flows into and out of the Hubei province and the broad pattern of migration flows inside China as well. Table 1 summarizes all the parameters.

Common Parameters The following common parameters come from the literature. Following Redding and Turner (2015), we set the agglomeration elasticity, β, to 0.1. We set the elasticity of substitution to, $\eta = 6$, which is a value in the middle of plausible ranges.[6] The trade elasticity is $\theta = 4$ as in Simonovska and Waugh (2014). The elasticity of migration, κ, varies between 1.4 and 3.3 in the literature; we set $\kappa = 2.0$ following Hsieh and Moretti (2019).[7] Appendix E provides robustness checks and shows that the quantitative results are robust to the alternative parameter values.

Initial Population Distribution The initial population distribution comes from the *Population Census* in the year 2010. We use the total population, including the urban and rural populations in each prefecture, as the initial population.

Migration and Trade Costs We assume that the migration costs from j to i, denoted as λ_{ij} takes the following functional form:

$$\lambda_{ij} = \bar{\lambda} \cdot \lambda_i \cdot \lambda_j \cdot \left(T_{ij}^p\right)^{\xi}. \tag{7}$$

Migration frictions depend on the national migration policy, $\bar{\lambda}$, and the location-specific entry and exit barriers, λ_i and λ_j. Migration frictions are also related to the underlying

[6] The elasticity of substitution usually ranges between 5 and 10 in the literature (Anderson and van Wincoop, 2004).

[7] For example, Monte et al. (2018) estimate κ to be 3.3 in the context of the U.S. and Hsieh and Moretti (2019) set it to 2.0. Bryan and Morten (2019) estimate it to be 2.7 using Indonesian data. Stillwell et al. (2014) finds κ to be between 1.4 and 2.2 in in a number of the European countries based on reduced-form estimations.

passenger transportation networks between the cities, T_{ij}^p, up to an elasticity of ξ. We first focus on the transportation network, $\left(T_{ij}^p\right)^{\xi}$.

We use the transportation network, T_{ij}^p, in 2015 from Ma and Tang (2020a). The parameter ξ governs the elasticity of λ_{ij} with respect to the infrastructure, T_{ij}^p. We follow the same estimation methods in Ma and Tang (2020a), with the updated bilateral migration matrix in 2015. The migration flow in equation (6) can be transformed to:

$$\log\left(L_{ij}\right) = \kappa \log\left(V_i\right) - \kappa \log\left(\Pi_j\right) + \log\left(\bar{L}_j\right) - \kappa \log \bar{\lambda} - \kappa \log \lambda_i - \kappa \log \lambda_j - \kappa\xi \log\left(T_{ij}^p\right).$$

This equation leads to a reduced-form estimation with origin and destination fixed effects. The two fixed effects, δ_i and δ_j, absorb all the variables in the expression above except for the last term:

$$\log\left(L_{ij}\right) = \delta_i + \delta_j - \kappa\xi \log\left(T_{ij}^p\right).$$

We estimate the equation using OLS, with the migration flow data from the *1% Population Sampling Survey* in 2015. The regression estimates $\kappa\xi$ to be 0.40. With the calibrated κ at 2.0, the estimated ξ equals 0.20. We also estimate the equation with an instrumental variable for T_{ij}^p to alleviate the concerns of endogenous placements of infrastructure. To do so, we follow Faber (2014) to construct the Minimum-Spanning Tree instruments. The point estimate is only slightly higher at $\kappa\xi = 0.42$ and with corresponding ξ at 0.21. In the baseline model, we use the OLS estimate.

The trade costs matrix is also based on Ma and Tang (2020a). The trade cost matrix is assumed to be:

$$\tau_{ij} = \bar{\tau} \cdot \left(T_{ij}^g\right)^{\psi},$$

where $\bar{\tau}$ is the overall trade frictions, T_{ij}^g is the underlying goods transportation network, and ψ is the elasticity of the iceberg costs to T_{ij}^g. We take the values of $\left(T_{ij}^g\right)^{\psi}$ directly from Ma and Tang (2020a) and estimate $\bar{\tau}$ in our context.

City-Level Productivity, $\bar{A}_j$ We follow the methods in Ma and Tang (2020b), which implements Donaldson and Hornbeck (2016) in the context of China to estimate the city-level productivity. The details are explained in Appendix D.

5.1 Joint Calibration

The remaining four parameters call for joint-calibration: the overall migration and trade barriers, $\bar{\lambda}$ and $\bar{\tau}$, and the origin and destination-specific migration barriers, λ_i and λ_j. As the COVID-19 outbreak stems from a single epicenter of Hubei, the migration flows between prefectures outside of Hubei are irrelevant to the virus's spatial spread. For this reason, we only impose λ_i and λ_j on the prefectures within Hubei province, and assume that $\lambda_i = \lambda_j = 1.0$ for all the migration flows outside of Hubei. To simplify notation, we use λ_{IN} to denote the common migration friction of moving into any prefecture in Hubei, and λ_{OUT} to denote the friction of moving out of Hubei.

We jointly calibrate these four parameters to four moments in the data. The first moment is the internal-trade-to-GDP ratio of 0.625 from the *Investment Climate Survey* conducted by the World Bank. This moment identifies the overall internal trade barrier, $\bar{\tau}$. The second moment is the overall stay-rate of 89 percent. This moment is defined as one minus the fraction of migrants in the entire population as computed from the *Population Sampling Survey*. This moment pins down $\bar{\lambda}$. The other two moments also come from the same survey: the outflow rate of all prefectures in Hubei province at 14.7 percent, and the inflow rate at 3.5 percent. The outflow (inflow) rate is defined as the total outflow (inflow) population as a fraction of the initial population of Hubei. The outbound and inbound migration barriers, λ_{OUT} and λ_{IN}, are respectively backed out from these two moments.

Our model is calibrated to match the population flow into and out of the Hubei province in 2015. Moreover, we can also match the bilateral population flows between prefectures in Hubei and prefectures outside Hubei due to the detailed geographic information incorporated in the T_{ij}^{p} matrix. Appendix D discusses the out-of-sample model fit.

6 Quantitative Results

In this section, we illustrate the impact of transportation networks and migration policies on disease transmission through the lens of our model. In the model, transportation networks were captured by the T_{ij}^p and T_{ij}^g matrices, and the migration policies are summarized in the $\Lambda = \{\bar{\lambda}, \lambda_{\text{IN}}, \lambda_{\text{OUT}}\}$ vector. To counterfactually simulate the population flow, we first need to estimate these objects under the counterfactual scenario.

To back-out the policy parameters, we re-calibrate the model to the state of the Chinese economy around the year 2005. Following the same strategy, we use the data from the *2005 Population Survey*, the initial population from the census in 2000, and the T_{ij}^p and T_{ij}^g matrices in 2005 from Ma and Tang (2020a) to calibrate the counterfactual. We also re-estimate the $\bar{A}_j$ vector in the year 2005. All the other parameters are the same as in the 2015 calibration. These parameters are reported in Table 1.

The migration policy has been substantially liberalized over time, as seen in the table. Between 2005 and 2015, the national migration multiplier, $\bar{\lambda}$, fell by 54 percent, while the Hubei-specific frictions fell by 11 – 43 percent. The decline in these estimated policies is driven by the surge of internal migration in China, as reflected in the two *Population Surveys*. In the 2005 survey, the aggregate stay rate was around 94.4 percent, and it declined to 89 percent in 2015. Similarly, the outflow rate of Hubei province doubled from 7.4 percent to 14.7 percent, and the inflow rate more than quadrupled from 0.8 percent to 3.5 percent. These data patterns are broadly consistent with the reforms in the urbanization policy during that time, as discussed in detail in Hsu and Ma (2021).

6.1 Counterfactuals: Migration Flows, Population, Real Income and Welfare

In the rest of the section, we present three sets of counterfactual simulations. In the first "constant network" simulation, we use the T_{ij}^p and T_{ij}^g matrix in the year 2005 and keep all the other parameters the same as in the 2015 baseline. In the second counterfactual, "constant policy", we use the Λ vector in 2005. In the third simulation, "constant network and policy", we revert T_{ij}^p, T_{ij}^g, and Λ back to 2005. The counterfactual results, together

with the baseline model, are presented in Table 2. With the older transportation network, the total population flow in and out of Hubei province declines by 13 percent. The mild response of migration flow is expected because distance-related costs are a minor obstacle for migrants (Morten and Oliveira, 2016). The tightening of migration policies, on the other hand, induces a sharp decline in population flow. Reverting the policy vector to the year 2005 reduces the Hubei-related population flow by 57 percent. Lastly, combining both changes leads to a 63 percent reduction in population flow.

When transport network and migration policies are altered, spatial distributions of population and income change as well. Figure A.4 in the appendix shows the distribution of changes in population and wage rates in different counterfactual scenarios. The induced change in welfare is significantly larger than that in real income. For example, the population-weighted average decline in real wage under the case of "constant network & policy" is 3.60%, while the decline in welfare is 7.41%.[8]

6.2 Counterfactuals: Disease Transmission

Given the counterfactual migration flows, travel distance, population, and income per capita, we simulate the incidence of COVID-19 in prefecture i according to:

$$I_i(t)^{CF} = I_i(t) \exp\left(\hat{\beta}_{1t}\Delta X_i + \hat{\beta}_{2\tau}\Delta M_i + \hat{\beta}_{3t}\Delta \ln(Dist_i) + \hat{\beta}_{4t}\Delta \ln(Pop_i) + \hat{\beta}_{5t}\Delta \ln(GDPpc_i)\right),$$

where $\Delta X_i = X_i^{CF} - X_i$ and X_i^{CF} represents the counterfactual ratio of Hubei-bound emigrants to the local population in prefecture i. Other variables are defined analogously. $\hat{\beta}_{\iota t}$'s are the estimates obtained from Section 3.1. The counterfactual total number of infections outside Hubei is computed as $I(t)^{CF} = \sum_i I_i(t)^{CF}$.

Panel (b) in Table 2 presents the counterfactual trends of COVID-19 under the three scenarios discussed above and contrasts them with the actual data. For brevity, we focus on the actual and counterfactual spreads by February 22, 2020. We find that had the transportation networks been the same as in 2005, the total number of reported infections

[8] Welfare in prefecture i is defined as $\frac{w_i}{P_i} m_{ii}^{-1/\kappa}$, where m_{ii} is the stay rate as defined in equation (5). The term $m_{ii}^{-1/\kappa}$ captures the welfare loss that comes from the migration frictions as shown in Tombe and Zhu (2019).

would have been lower by 1,908, amounting to 15.31% of the total reported infections. In the counterfactual scenario where the migration policy remained the same as in 2005, the number of reported infections would have been lower by 2,221, which is 17.82% of the total reported infections. Lastly, had both transportation networks and migration policies reverted to their 2005 configurations, the number of infections would have been lower by 3,517, which is 28.21% of the total reported infections. Figure A.5 in the appendix presents the spatial distribution of the counterfactual declines in the number of infections across prefectures. We find that the spread could have been reduced more in the coastal areas and in the regions that are geographically closer to Hubei in the counterfactuals.

The quantitative importance of transportation networks and migration policies is similar in explaining the overall spread of COVID-19 outside Hubei. However, the two factors affect the disease spread through different channels, as revealed by the decomposition exercises in Table 2. Panel (b.1) finds that under the case of "constant network", the direct effect of an increase in travel distance decreases the number of total reported infections by 11.27%, while the induced decrease in migration flows leads to only a 3.44% reduction. These estimates are consistent with the findings in Section 6.1 that migration flows declines slightly in response to a reversion of the transportation infrastructure to the 2005 configuration. Hence, the rapid expansion of transportation infrastructure in China mainly affects disease transmission by increasing short-term population movement rates rather than altering medium- and long-term migration patterns. Columns (8) and (9) reveal the roles of changes in income and population distributions induced by the changing transportation network. We find that such general equilibrium effects lower the reported number of COVID-19 cases by 1.21%. Panel (b.2) shows that the counterfactual changes in migration policies mainly affect the disease spread through changing migration flows, while the general equilibrium effects lead to a moderate reduction of 1.44%. Lastly, as shown in Panel (b.3), under the case of "constant network and policy", the induced changes in migration flows, travel distance, and income and population reduce the reported number of infections by 16.89%, 11.27%, and 2.54%, respectively.

7 Conclusion

We evaluate the impacts of migration flows and transportation infrastructure on the spatial transmission of COVID-19 in China. Using the daily data of reported cases at the prefecture level and the bilateral migration data from the mini census, we show that cities with larger bilateral migration flows and shorter travel distances with Hubei experienced a greater spread of COVID-19. In addition, wealthier prefectures with higher incomes were better able to contain the virus in the early stages. We then evaluate the contribution of the rapid development in transportation infrastructure and the liberalization of migration restrictions in the recent decade using a general equilibrium spatial model. We show that the increased mobility of people following the expansion of transportation networks and easing migration policies explain around 28% of the infections outside Hubei province. The strong link between disease transmission, migration policy, and transportation networks documented in this paper highlights the need to incorporate epidemiological elements in the models of the spatial economy in future research.

References

Adda, Jérôme, "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data," *The Quarterly Journal of Economics*, 2016, *131* (2), 891–941.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *1085*, 1139.

_ and _ , "The Welfare Effects of Transportation Infrastructure Improvements," Working Paper 25487, National Bureau of Economic Research January 2019.

_ , _ , and Yuta Takahashi, "Universal Gravity," *Journal of Political Economy*, 2020, *128* (2), 393–433.

Anderson, James E. and Eric van Wincoop, "Trade Costs," *Journal of Economic Literature*, 2004, *42* (3), 691–751.

Ashenfelter, Orley and Michael Greenstone, "Using mandated speed limits to measure the value of a statistical life," *Journal of political Economy*, 2004, *112* (S1), S226–S267.

Bloomberg News, "China Will Rack Up Three Billion Trips During Worlds Biggest Human Migration," Available at https://www.bloomberg.com/news/articles/2020-01-20/china-readies-for-world-s-biggest-human-migration-quicktake 2020.

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *The Review of Economic Studies*, 2018, *85* (4), 2042–2096.

_ , Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019.

Cameron, A Colin and Pravin K Trivedi, *Regression analysis of count data*, Vol. 53, Cambridge university press, 2013.

Chinazzi, Matteo, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun et al., "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, 2020.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, April 2018, *108* (4-5), 899–934.

_ **and Richard Hornbeck**, " Railroads and American Economic Growth: A Market Access Approach," *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, Sep. 2002, *70* (5), 1741–1779.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 03 2014, *81* (3), 1046–1070.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–88.

Fogel, Robert William, "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings," *The Journal of Economic History*, 1962, *22* (2), 163–197.

Fogli, Alessandra and Laura Veldkamp, "Germs, Social Networks and Growth," *Forthcoming in Review of Economic Studies*, 2019.

Guan, Weijie, Zhengyi Ni, Yu Hu, Wenhua Liang, Chunquan Ou, Jianxing He, Lei Liu, Hong Shan, Chunliang Lei, David SC Hui et al., "Clinical Characteristics of 2019 Novel Coronavirus Infection in China," *New England Journal of Medicine*, 2020.

Hsieh, Chang-Tai and Enrico Moretti, "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, April 2019, *11* (2), 1–39.

Hsu, Wen-Tai and Lin Ma, "Urbanization Policy and Economic Development: A Quantitative Analysis of Chinas Differential Hukou Reforms," *Regional Science and Urban Economics*, 2021, p. 103639.

Jia, Jayson S., Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A. Christakis, "Population Flows Drives Spatio-Temporal Distribution of COVID-19 in China," *Nature*, 2020.

Kraemer, Moritz UG, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis du Plessis, Nuno R Faria, Ruoran Li, William P Hanage et al., "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, 2020.

Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong et al., "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia," *New England Journal of Medicine*, 2020.

Ma, Lin and Yang Tang, "The Distributional Impacts of Transportation Networks in China," Technical Report, working paper 2020.

_ **and** _ , "Geography, trade, and internal migration in China," *Journal of Urban Economics*, 2020, *115*, 103181. Cities in China.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, December 2018, *108* (12), 3855–90.

Morten, Melanie and Jaqueline Oliveira, "Paving the Way to Development: Costly Migration and Labor Market Integration," Working Paper 22158, National Bureau of Economic Research April 2016.

Oster, Emily, "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa," *Journal of the European Economic Association*, 2012, *10* (5), 1025–1058.

Petersen, Eskild, Marion Koopmans, Unyeong Go, Davidson H Hamer, Nicola Petrosillo, Francesco Castelli, Merete Storgaard, Sulien Al Khalili, and Lone Simonsen, "Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics," *The Lancet infectious diseases*, 2020.

Redding, Stephen J. and Matthew A. Turner, "Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, pp. 1339 – 1398.

Simonovska, Ina and Michael E. Waugh, "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 2014, *92* (1), 34 – 50.

State Council, "Fighting COVID-19: China in Action," Technical Report White Paper, The State Council Information Office of the People's Republic of China, http://www.scio.gov.cn/zfbps/32832/Document/1681809/1681809.htm June 2020.

Stillwell, John, Konstantinos Daras, Martin Bell, and Nik Lomax, "The IMAGE studio: A tool for internal migration analysis and modelling," *Applied Spatial Analysis and Policy*, 2014, *7* (1), 5–23.

Tian, Huaiyu, Yonghong Liu, Yidan Li, Chieh-Hsi Wu, Bin Chen, Moritz UG Kraemer, Bingying Li, Jun Cai, Bo Xu, Qiqi Yang et al., "An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China," *Science*, 2020.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, May 2019, *109* (5), 1843–72.

Verity, Robert, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson,

Patrick GT Walker, Han Fu et al., "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet Infectious Diseases*, 2020.

Viscusi, W Kip and Joseph E Aldy, "The value of a statistical life: a critical review of market estimates throughout the world," *Journal of Risk and Uncertainty*, 2003, *27* (1), 5–76.

Walker, PGT, C Whittaker, O Watson et al., "The Global Impact of COVID-19 and Strategies for Mitigation and Suppression: WHO Collaborating Centre for Infectious Disease Modelling," *MRC Centre for Global Infectious Disease Analysis, Abdul Latif Jameel Institute for Disease and Emergency Analytics, Imperial College London*, 2020.

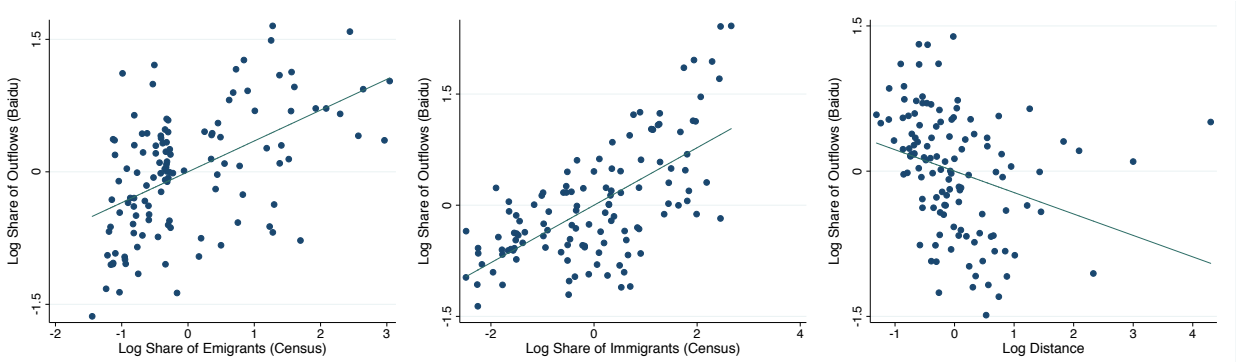
WHO, "Weekly epidemiological update 12 January 2021," Technical Report Situation Report, World Health Organization January 2021.

Wooldridge, Jeffrey M, "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, 1999, *90* (1), 77–97.

_ , *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.

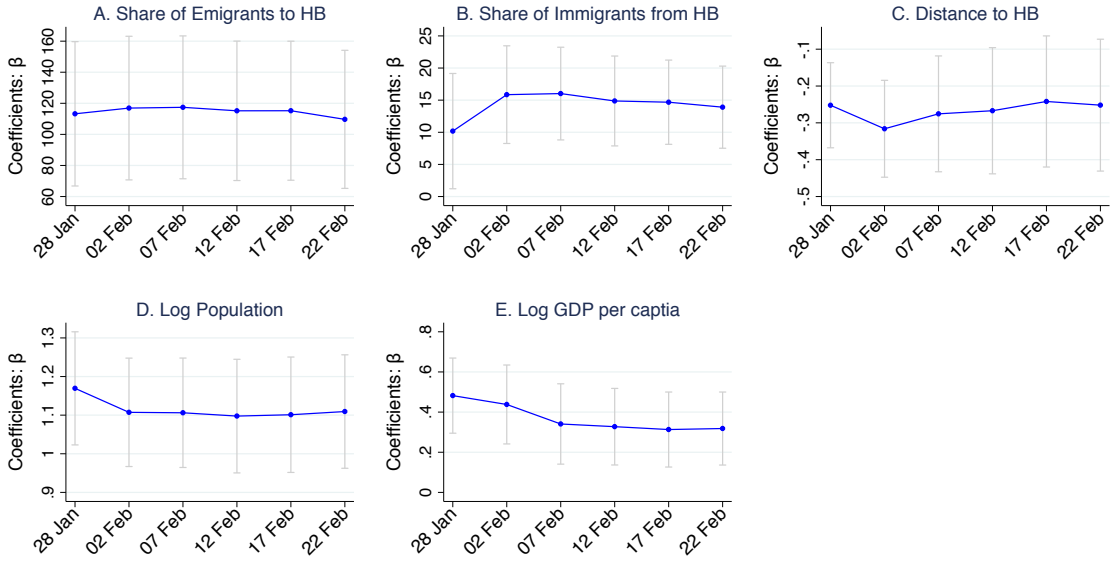
Tables and Figures

Figure 1: Residual Scatter Plots



Note: The residual scatter plots are of the multivariate regression $\ln OutFlow_i = \gamma_0 + \gamma_1 \ln X_i + \gamma_2 \ln M_i + \gamma_3 \ln Dist_i + \nu_i$, where $\ln OutFlow_i$ denotes the share of population outflows from Wuhan to prefecture i in the two weeks before the lockdown (January 9–22, 2020); X_i denotes the ratio of emigrants to Hubei to the population in prefecture i in 2015; M_i is the share of immigrants from Hubei in the local population in prefecture i in 2015; $Dist_i$ measures the travel distance between prefecture i and Hubei based on transportation networks in 2015. We discuss the data sources in Section 2. The green straight line is the best-fitted line.

Figure 2: Estimates of Cumulative Effects: $\beta_{\iota t}$



Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\beta_{\iota t}$ coefficients in equation (1).

Table 1: Calibration Results

(a) Fixed Parameters

Name	Value	Source	Note
β	0.1	Redding and Turner (2015)	The agglomeration elasticity
θ	4.0	Simonovska and Waugh (2014)	Trade elasticity
κ	2.0	Hsieh and Moretti (2019)	Migration elasticity
η	6.0	Anderson and van Wincoop (2004)	Elasticity of substitution
$\bar{A}_j$	-	Ma and Tang (2020b)	City-level productivity

(b) Calibrated Parameters

Name	2005	2015	Note
$\bar{\lambda}$	1415.24	645.24	Overall Migration Barrier
λ_{IN}	7.77	4.41	Entry Barrier, Hubei
λ_{OUT}	0.85	0.76	Exit Barrier, Hubei
$\bar{\tau}$	10.65	13.46	Overall Trade Barrier

Note: This table summarizes the calibrated model parameters. Panel (a) presents the parameters that come from the literature. Panels (b) presents the jointly calibrated parameters.

Table 2: Counterfactual Experiments

(a) Population Flow: Baseline v.s. Counterfactual Simulations (Thousands)

Case	Total Flow (1)	Outflow (2)	Inflow (3)	Fraction of Baseline (4)
Baseline	10253.11	8247.61	2005.50	1.00
Constant Network	8971.20	7339.09	1632.11	0.87
Constant Policy	4359.36	3797.87	561.49	0.43
Constant Network & Policy	3779.17	3325.76	453.41	0.37

(b) Incidence of COVID-19 Outside Hubei: Actual v.s. Counterfactual Simulations

	Actual	Counterfactual							
		All Factors		Migration Flows		Travel Distance		GDP and Pop	
	Cases	Cases	Decline %	Cases	Decline %	Cases	Decline %	Cases	Decline %
Date	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel b.1: Constant Network (Transportation Networks Same as 2005)									
28/Jan/2020	2349	1975	15.94	2267	3.47	2078	11.53	2311	1.64
02/Feb/2020	5873	4794	18.37	5656	3.69	5055	13.93	5777	1.63
07/Feb/2020	9368	7829	16.43	9024	3.67	8228	12.17	9245	1.31
12/Feb/2020	11266	9468	15.96	10863	3.58	9937	11.79	11124	1.26
17/Feb/2020	12068	10263	14.95	11637	3.57	10767	10.78	11921	1.22
22/Feb/2020	12464	10556	15.31	12035	3.44	11059	11.27	12313	1.21
Panel b.2: Constant Policy (Migration Policies Same as 2005)									
28/Jan/2020	2349	1953	16.85	1983	15.58	2349	0.00	2315	1.43
02/Feb/2020	5873	4748	19.15	4829	17.77	5873	0.00	5776	1.65
07/Feb/2020	9368	7564	19.25	7685	17.97	9368	0.00	9226	1.52
12/Feb/2020	11266	9156	18.73	9297	17.47	11266	0.00	11101	1.47
17/Feb/2020	12068	9817	18.65	9969	17.39	12068	0.00	11890	1.47
22/Feb/2020	12464	10243	17.82	10397	16.59	12464	0.00	12285	1.44
Panel b.3: Constant Network & Policy									
28/Jan/2020	2349	1695	27.83	1978	15.81	2078	11.53	2279	2.99
02/Feb/2020	5873	4005	31.81	4809	18.12	5055	13.93	5688	3.14
07/Feb/2020	9368	6532	30.28	7652	18.31	8228	12.17	9114	2.71
12/Feb/2020	11266	7946	29.47	9261	17.79	9937	11.79	10971	2.62
17/Feb/2020	12068	8622	28.55	9931	17.71	10767	10.78	11757	2.58
22/Feb/2020	12464	8947	28.21	10359	16.89	11059	11.27	12148	2.54

Note: This table reports the results of three counterfactual experiments: "Constant Network" refers to the counterfactual using the T_{ij}^{p} and T_{ij}^{g} matrices in 2005; "Constant Policy" refers to the counterfactual using the Λ parameters in 2005; "Constant Network & Policy" refers to the counterfactual using both the T_{ij}^{p} and T_{ij}^{g} matrices and the Λ parameters in 2005. Panel (a) summarizes the population flows in and out of the Hubei province in the baseline and the counterfactual simulations. Panel (b) reports the actual spread of reported COVID-19 cases over time, and the spreads under three counterfactual scenarios. Columns 4 to 9 decompose the overall counterfactual changes reported in columns 2 and 3 into different components: (i) changes induced by counterfactual changes in bilateral migration flows specific to Hubei (i.e., X_i and M_i in equation (1)); (ii) changes induced by counterfactual changes in bilateral distance with Hubei (i.e., $\ln(Dist_i)$ in equation (1)); (iii) changes induced by counterfactual changes in population and GDP per capita (i.e., $\ln(Pop_i)$ and $\ln(GDPpc_i)$ in equation (1)).

Online Appendix

A Additional Tables and Figures

Figure A.1: Spread of COVID-19 Over Time

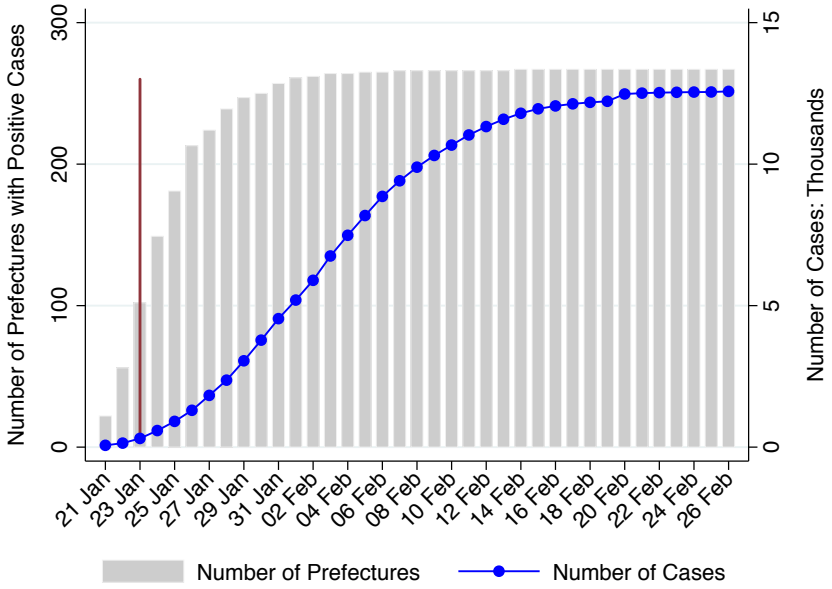
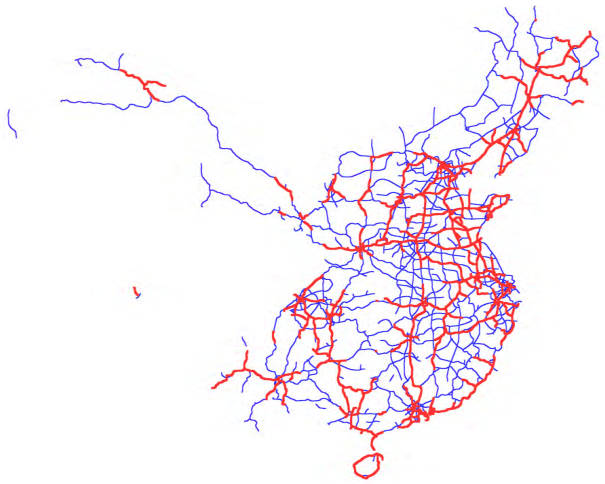
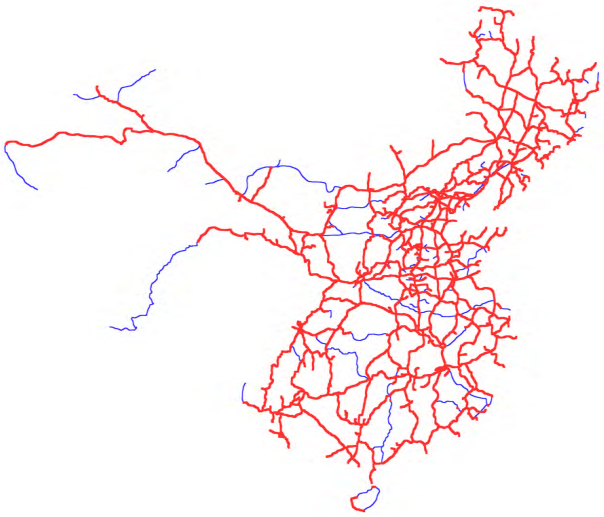


Figure A.2: Transportation Networks: 2005 v.s. 2015

(a) Highways



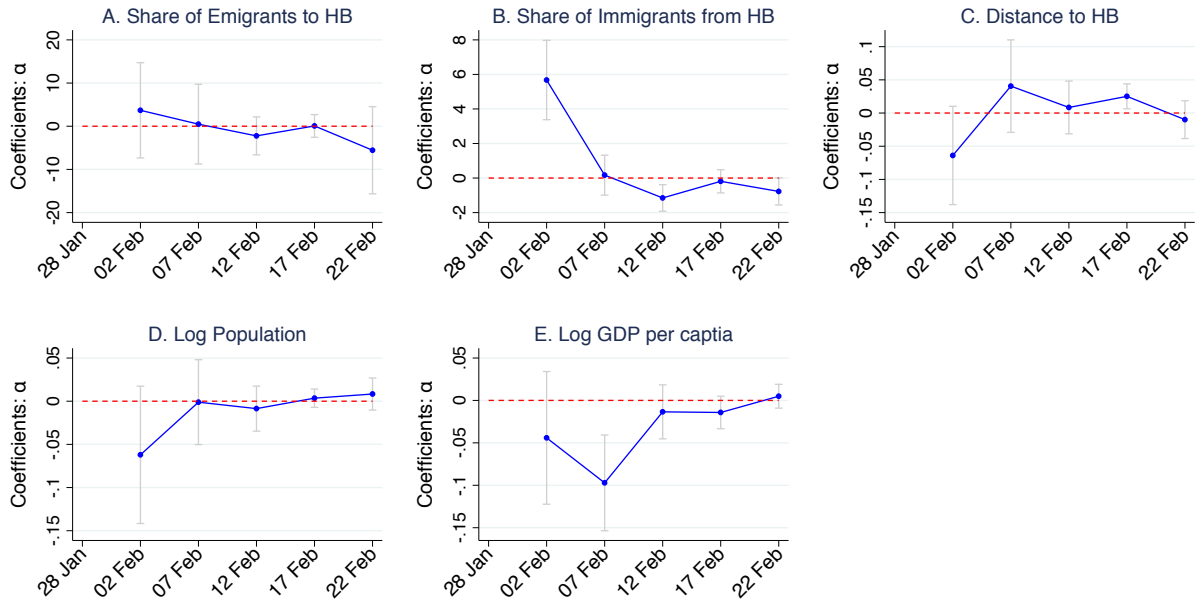
(b) Railways



(c) National Roads

Note: This figure shows the transportation networks for highways (panel A), railways (panel B), and national roads (panel C) in China, respectively. The networks in 2005 are represented by the thick red lines, and the networks in 2015 are represented by thin blue lines.

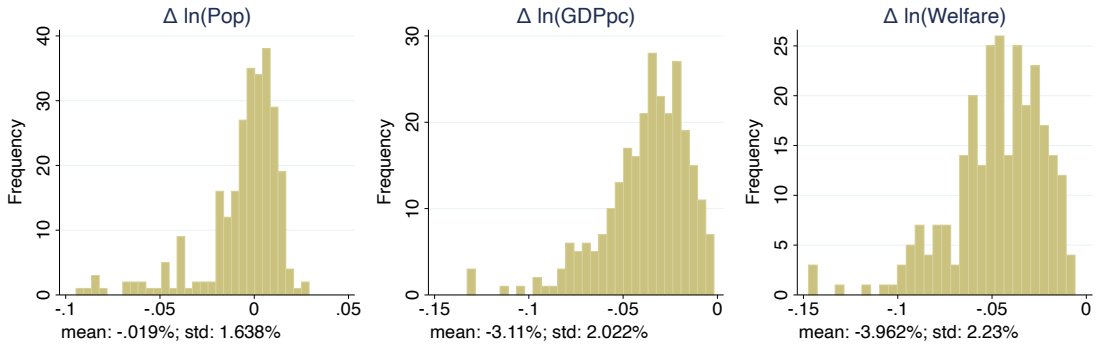
Figure A.3: Estimates of Policy Parameters: $\alpha_{\iota t}$



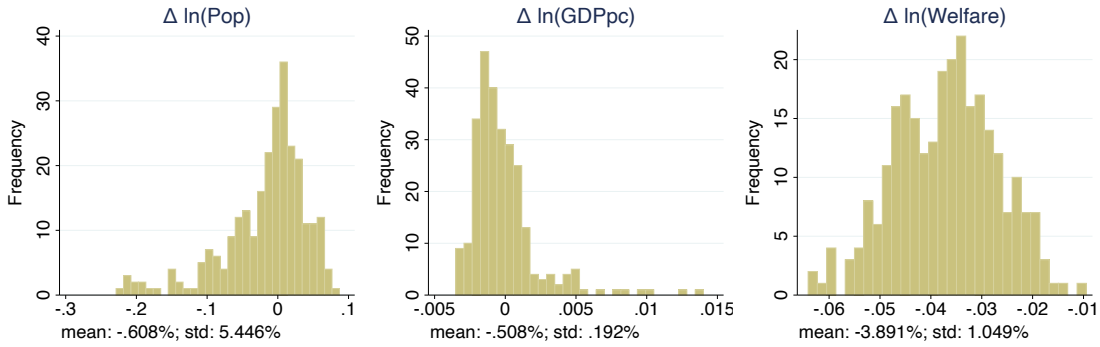
Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\alpha_{\iota t}$ coefficients of equation (C.5). These estimates are obtained by $\hat{\alpha}_{\iota t} = \hat{\beta}_{\iota t} - \hat{\beta}_{\iota t-\delta}$, where $\hat{\beta}_{\iota t}$'s are obtained from equation (1). The standard errors of $\hat{\alpha}_{\iota t}$'s are computed based on the variance-covariance matrix of $\hat{\beta}_{\iota t}$'s, and the confidence intervals are constructed accordingly.

Figure A.4: Counterfactual Changes in Population, Real GDP per capita, and Welfare

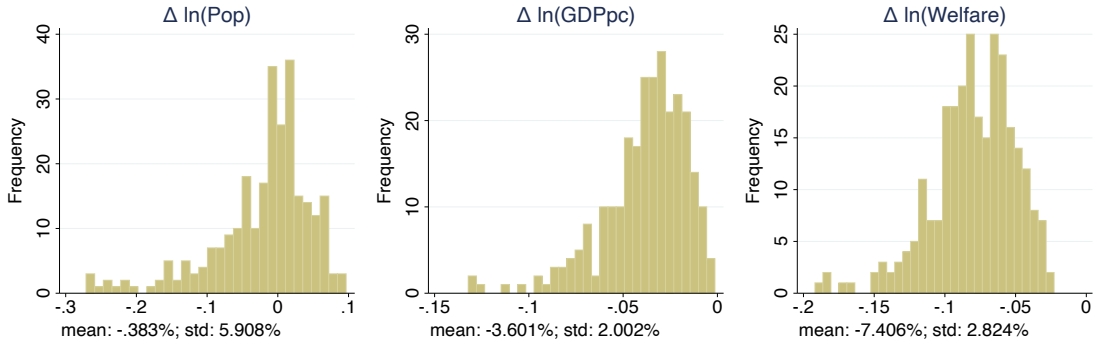
(a) Constant Network



(b) Constant Policy



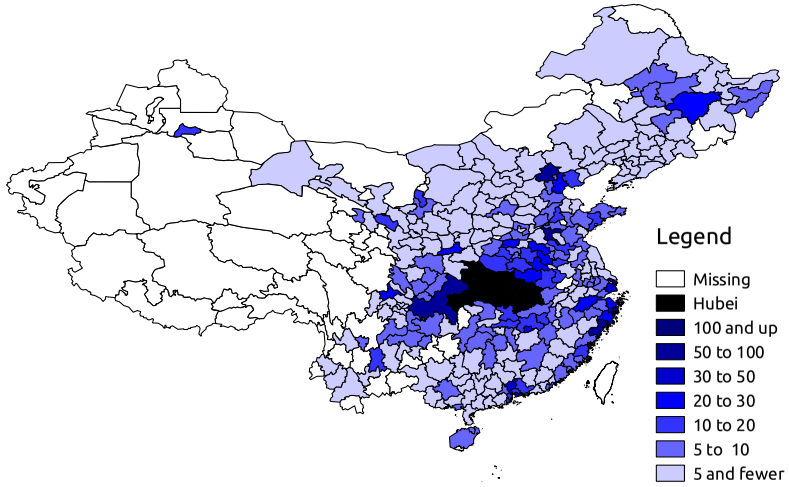
(c) Constant Network & Policy



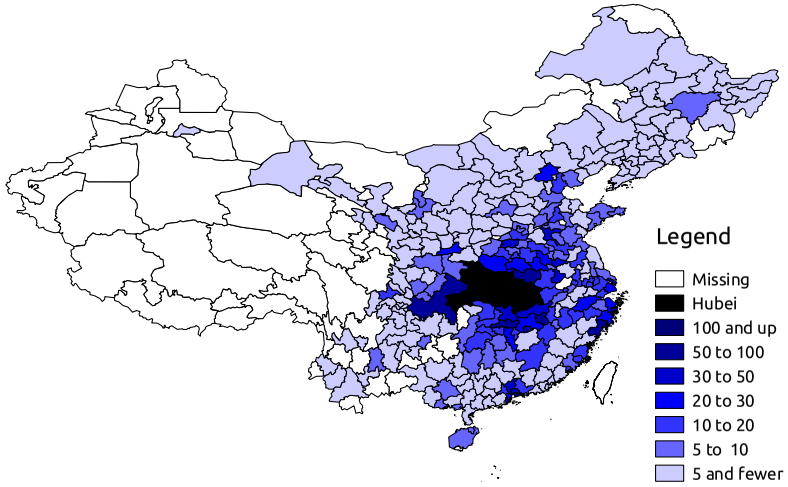
Note: This figure reports the distributions of counterfactual changes in population and income under three counterfactual scenarios: "Constant Network" refers to the counterfactual using the T_{ij}^p and T_{ij}^g matrices in 2005; "Constant Policy" refers to the counterfactual using the Λ parameters in 2005; "Constant Network & Policy" refers to the counterfactual both the T_{ij}^p and T_{ij}^g matrices and the Λ parameters in 2005. Under each graph two summary statistics are presented. In the first column, the mean is the average counterfactual population growth weighted by the population in the baseline equilibrium. In the middle column (respectively, the last column), the mean change in real wage (respectively, welfare) is the percentage change in the population weighted average of real wage (respectively, welfare) from the baseline to the counterfactual equilibrium, which represents the change in real wage (respectively, welfare) at the national level. All columns report the standard deviation of the counterfactual changes weighted by the population in the baseline equilibrium.

Figure A.5: Incidence of COVID-19 Outside Hubei: Counterfactual Declines by 22/Feb/2020

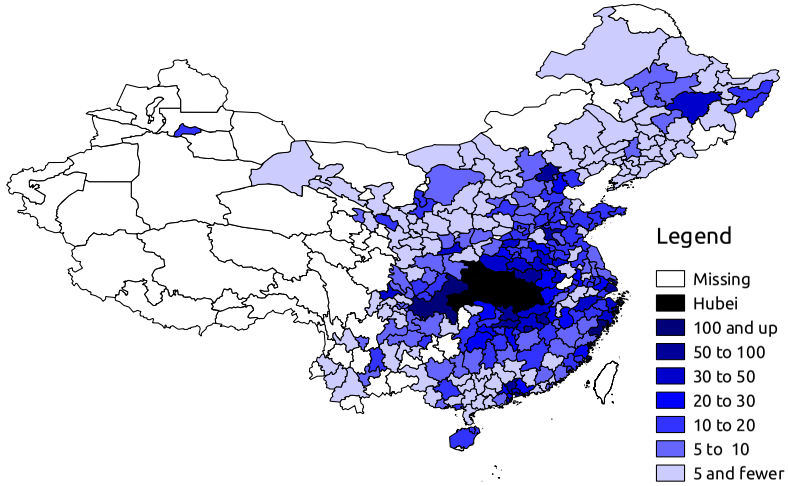
(a) Constant Network



(b) Constant Policy



(c) Constant Network & Policy



Note: This figure shows the transportation networks for highways (panel A), railways (panel B), and roads (panel C) in China, respectively. The networks in 2005 are represented by the thick red lines, and the networks in 2015 are represented by thin blue lines.

B Details of the Cost-Benefit Calculations in the Introduction

Benefits The GDP of China in 2015 was 8.914 trillion constant 2010 USD as reported by the World Bank, which leads to the cost estimate of $8914 * 0.036 = 321$ billion USD as reported in the main text.

Costs The reduction in the number of infections, 3517, is based the data in the last row of Table 2: $12464 - 8947 = 3517$. The hospitalization rate of COVID-19 in China is estimated to be 3.75% (Walker et al., 2020), which leads to $3517 * 0.0375 \approx 132$ reduction in hospitalizations. The infection fatality rate in China was estimated to be 0.66% (Verity et al., 2020), which leads to $3517 * 0.0066 \approx 23$ fewer fatalities.

The value of a statistical life estimated in the literature ranges from \$1.5 million to \$7.5 million (Ashenfelter and Greenstone, 2004; Viscusi and Aldy, 2003). According to the official report, the cost of hospitalization during the pandemic was 23,000 RMB (\$3,600) (State Council, 2020). Using the lower estimate of a statistical life at \$1.5 million, the costs are computed as $23 * 1.5 + 132 * 0.0036 \approx 35$ million USD; and at the higher estimate of \$7.5 million: $23 * 7.5 + 132 * 0.0036 \approx 173$ million USD. In these calculations, we ignored the loss of human capital due to the lack of data; however, such omissions are unlikely to alter the relative sizes between the healthcare costs and economic benefits of better connectivity due to the sheer gap between the two.

Costs in the Hypothetical Case A 28.21% reduction of 100 million cases is 28.2 million fewer cases. The reduction leads to 1.06 million fewer hospitalizations and 186 thousand fewer fatalities, using the same estimates of hospitalization and fatality rates as above. Applying the same estimate of costs, the lower bound is $186000 * 1.5 + 1060000 * 0.0036 \approx \$283$ billion, and the upper bound, $186000 * 7.5 + 1060000 * 0.0036 \approx \$1.40$ trillion.

As of Jan 12th, 2021, the WHO reported 22.3 million cases in the U.S., which is 6.7% of the total population of 330.8 million. The hypothetical 100 million cases in China translate into a 7.2% population infection rate, based on China's population of 1,394 million.

C A Simple Model of Disease Transmission

In this appendix, we lay out a simple model of disease transmission that rationalizes the empirical model in Section 2 and provides structural interpretations of the regression coefficients.

The disease takes two stages to develop. In the first stage, all cases are imported from Hubei. The number of imported cases, $I_i(0)$, follows a Poisson distribution with the arrival rate λ_i given by:

$$\lambda_i = \exp\Big(\theta_0 + \theta_1 X_i + \theta_2 M_i + \theta_3 \ln(Dist_i) + \theta_4 \ln(Pop_i) + \theta_5 \ln(GDPpc_i)\Big), \tag{C.1}$$

where X_i denotes the ratio of Hubei-bound emigrants to the local population in prefecture i, and M_i is the share of immigrants from Hubei in the local population in prefecture i; $Dist_i$ measures the travel distance between prefecture i and Hubei based on transportation networks in 2015; and Pop_i and $GDPpc_i$ represent population size and GDP per capita in 2015, respectively. The arrival rate takes a gravity form, and is determined by the bilateral migration pattern with Hubei, the travel distance, which affects short-term population movement, and the size of the local economy.

In the second stage, the disease is transmitted locally. We consider the dynamics of the epidemic as follows:

$$\frac{dI_i(t)}{dt} = \gamma \Omega_i(t) I_i(t), \tag{C.2}$$

where γ denotes the rate at which new cases develop, which is exogenously determined by the infectiousness of COVID-19; $I_i(t)$ denotes the fraction of individuals who are infected at time t; $\Omega_i(t)$ is the fraction of individuals who are susceptible to contracting the disease, which is time varying and depends on avoidance behaviours and the public health measures controlling social interactions. Compared to the Standard Inflammatory Response (SIR) model which is often employed in the epidemiology literature, our framework abstracts away from the dynamics associated with the fraction of the population who have recovered from the disease and have immunity (R), and the fraction of the population who are deceased due to the disease (D). We do this because in prefectures outside Hubei during the sample period, both the percentage of the population with immunity and the percentage of the population who are deceased are negligible. Moreover, as is shown in Figure A.1, the spread outside

Hubei was almost halted by the end of the sample period, suggesting that R and D should play a minimal role in determining the transmission dynamics outside Hubei.

We assume that the fraction of susceptible individuals is determined by:

$$\Omega_i(t) = \tilde{\alpha}_{0t} + \tilde{\alpha}_{1t}X_i + \tilde{\alpha}_{2t}M_i + \tilde{\alpha}_{3t}\ln(Dist_i) + \tilde{\alpha}_{4t}\ln(Pop_i) + \tilde{\alpha}_{5t}\ln(GDPpc_i) + \tilde{\varepsilon}_i(t). \quad \text{(C.3)}$$

Again, we allow the size of the susceptible population to be affected by bilateral migration patterns and distance with Hubei, reflecting possible interactions such as family gatherings with relatives traveling from Hubei prior to the Spring Festival. Importantly, the $\tilde{\alpha}$ coefficients could vary over time, reflecting the effectiveness of measures to contain the transmission of the disease. For example, when travelers from Hubei are subjected to quarantine orders, the coefficients $\tilde{\alpha}_{1t}$ and $\tilde{\alpha}_{2t}$ are expected to be zero; when the travel ban from and to the epicenter is imposed, $\tilde{\alpha}_{3t}$ should decrease in magnitude; when prefectures outside Hubei adopt stringent transmission control measures such as social distancing or lockdowns, the coefficients $\tilde{\alpha}_{4t}$ and $\tilde{\alpha}_{5t}$ should shrink. Equation (C.2) is then rewritten as

$$\frac{dI_i(t)}{dt} = \Big(\alpha_{0t} + \alpha_{1t}X_i + \alpha_{2t}M_i + \alpha_{3t}\ln(Dist_i) + \alpha_{4t}\ln(Pop_i) + \alpha_{5t}\ln(GDPpc_i) + \varepsilon_i(t)\Big)I_i(t), \quad \text{(C.4)}$$

where $\alpha_{\iota t} = \gamma\tilde{\alpha}_{\iota t}$. In the following analysis, we refer to the α coefficients as policy parameters that capture the period-specific prevention and control policies governing the disease transmission at the prefecture level.

In the empirical analysis, we consider a discrete time version of equation (C.4), which is given by

$$\begin{aligned} I_i(t) &= I_i(t-\delta)\exp\Big(\alpha_{0t} + \alpha_{1t}X_i + \alpha_{2t}M_i + \alpha_{3t}\ln(Dist_i) + \alpha_{4t}\ln(Pop_i) + \alpha_{5t}\ln(GDPpc_i) + \varepsilon_i(t)\Big) \\ &= I_i(0)\exp\sum_{\tau=\delta,\ldots,t-\delta,t}\Big(\alpha_{0\tau} + \alpha_{1\tau}X_i + \alpha_{2\tau}M_i + \alpha_{3\tau}\ln(Dist_i) + \alpha_{4\tau}\ln(Pop_i) + \alpha_{5\tau}\ln(GDPpc_i) + \varepsilon_i(\tau)\Big), \end{aligned} \quad \text{(C.5)}$$

where δ represents the incubation time. Based on the findings in the epidemiological literature on COVID-19, we set the incubation period to five days. As shown in the second line of the equation, the number of infections in time t is determined by the number of initial imported cases and the cumulative increments up to time t. Substituting equation (C.1) into (C.5), we arrive at our estimation equation:

$$I_i(t) = \exp\Big(\beta_{0t} + \beta_{1t}X_i + \beta_{2\tau}M_i + \beta_{3t}\ln(Dist_i) + \beta_{4t}\ln(Pop_i) + \beta_{5t}\ln(GDPpc_i) + \nu_i(t)\Big), \quad \text{(C.6)}$$

where $\beta_{\iota t} = \theta_\iota + \sum_{\tau=\delta,\ldots,t} \alpha_{\iota\tau}$ represents the cumulative effects of the underlying variables up to period t, and $\nu_i(t) = \varepsilon_i(0) + \sum_{\tau=\delta,\ldots,t} \varepsilon_i(\tau)$. With the estimates of $\beta_{\iota t}$, we can back out the period-specific policy parameters $\alpha_{\iota t} = \beta_{\iota t} - \beta_{\iota t-\delta}$, which capture the dynamics of the policy interventions.

Guided by the model, we estimate the relationship between the number of infections and the economic fundamentals by a Poisson quasi-maximum likelihood count model (Wooldridge, 1999) based on equation (C.6).[9] The baseline estimates of $\beta_{\iota t}$ are reported in Figure 2 in the main text. Figure A.3 presents the estimates of the policy parameters $\alpha_{\iota t}$, which reveal the effects of the variables of interest on the disease propagation within a time interval.

D Details about the Spatial Model

D.1 Equilibrium Conditions and Model Solution

The solution of the model utilizes the equilibrium conditions as specified in Section 4. Conditional on a guess of equilibrium population distribution $\{L_j\}$, the solution of the model is similar to a standard Eaton-Kortum model. The price charged by the suppliers from city j in city i for variety ω is:

$$p_{ij}(\omega) = \frac{\tau_{ij} w_j}{z_j(\omega) A_j},$$

as determined by the **profit maximization problem of the firm**. The price paid for a particular variety ω in city i is:

$$p_i(\omega) = \min_j \left\{ \frac{\tau_{ij} w_j}{z_j(\omega) A_j} \right\}.$$

[9]The Poisson quasi-maximum likelihood count data model is generally preferred to alternative count data models (such as the negative binomial model), because the Poisson MLE estimator is consistent even when the error distribution is misspecified (i.e., the true distribution is not Poisson), provided that the conditional mean is specified correctly (Cameron and Trivedi, 2013; Wooldridge, 2002). Despite of this consideration, we demonstrates the robustness of the baseline findings to alternative specifications that estimate the relationship between the number of infections and the economic fundamentals in Appendix E.

Conditional on the Frechet distribution of productivity, the price index in city i is:

$$P_i = \left(\int_0^{\infty} p^{1-\eta} dG_i(p) \right)^{\frac{1}{1-\eta}} = \Psi \left(\sum_{j=1}^{J+1} (w_j \tau_{ij})^{-\theta} (A_j)^{\theta} \right)^{-1/\theta}$$

where $G_i(p)$ is the CDF of prices in city i, and Ψ is the Gamma function evaluated at $1 + (1 - \eta/\theta)$. The share of total expenditure in city i on the goods from city j is thus:

$$\pi_{ij} = \frac{(w_j \tau_{ij})^{-\theta} (A_j)^{\theta}}{\sum_{k=1}^{J+1} (w_k \tau_{ik})^{-\theta} (A_k)^{\theta}}.$$

The expression for the bilateral trade flow from j to i is thus:

$$X_{ij} = X_i \pi_{ij} = w_i L_i \frac{(w_j \tau_{ij})^{-\theta} (A_j)^{\theta}}{\sum_{k=1}^{J+1} (w_k \tau_{ik})^{-\theta} (A_k)^{\theta}}.$$

From the last equation, it is straightforward to solve for the equilibrium wage rate in city j through a system of non-linear equations. To see this, note that the **trade balance condition** implies $X_j = \sum_{j=1}^{J+1} X_{ij} = w_j L_j$:

$$X_j = \sum_{j=1}^{J+1} X_{ij} \iff w_j L_j = \sum_{j=1}^{J+1} w_i L_i \frac{(w_j \tau_{ij})^{-\theta} (A_j)^{\theta}}{\sum_{k=1}^{J+1} (w_k \tau_{ik})^{-\theta} (A_k)^{\theta}}. \tag{D.1}$$

The equation above describes a system of $J + 1$ non-linear equations where the vector $\{w_j\}$ is the unknown. We solve this with a simple iteration algorithm. All the other endogenous variables are functions of the wage rate. In particular, with the solution of the wage rate, we can compute the indirect utility, V_j, followed by the migration probabilities according to equation (5), which we replicate here for completeness:

$$m_{ij} = \frac{(V_i)^{\kappa} (\lambda_{ij})^{-\kappa}}{\sum_{m=1}^{J} (V_m)^{\kappa} (\lambda_{mj})^{-\kappa}}.$$

Note that the equation above is also the solution of the **utility maximization problem of the individuals**. We don't need to separately solve for the **labor market clearing conditions** as they are implicitly guaranteed by the trade balance condition per Walras's

Law.

The algorithm above depends on a guess of $\{L_j\}$, and the migration flows from solving equation (5) also imply a new equilibrium vector of population. We iterate on the population distribution until convergence at 1.0E-5.

D.2 City-Level Productivity

We estimate the city-level productivity from the residual of the following regression:

$$\ln(w_j) = \delta_0 + \delta_1 \ln(L_j) + \delta_2 \ln(\mathrm{MA}_j) + \nu_j,$$

where L_j is the population of city j and w_j is the wage rate that is approximated by per capita GDP. This equation comes from the trade balance condition in equilibrium. One can manipulate equation (D.1) as:

$$(w_j)^{1+\theta} = (L_j)^{-1} \left[\sum_{j=1}^{J+1} w_i L_i \frac{(\tau_{ij})^{-\theta}}{\sum_{k=1}^{J+1} (w_k \tau_{ik})^{-\theta} (A_k)^{\theta}}\right] \left[\bar{A}_j (L_j)^{\beta}\right]^{\theta}$$

$$w_j = (L_j)^{\frac{\beta\theta-1}{\theta+1}} \left[\sum_{j=1}^{J+1} w_i L_i \frac{(\tau_{ij})^{-\theta}}{\sum_{k=1}^{J+1} (w_k \tau_{ik})^{-\theta} (A_k)^{\theta}}\right]^{\frac{1}{\theta+1}} (\bar{A}_j)^{\frac{\theta}{\theta+1}}.$$

Taking logarithms on both sides of the equation, and approximating the terms in the square bracket as $\mathrm{MA}_j = \sum_{j=1}^{J+1} w_i L_i (\tau_{ij})^{-\theta}$ following Donaldson and Hornbeck (2016), we arrive at the equation to back-out city-level productivity:

$$\ln w_j = \frac{\beta\theta - 1}{\theta + 1} \ln L_j + \frac{1}{\theta + 1} \ln \mathrm{MA}_j + \frac{\theta}{\theta + 1} \ln \bar{A}_j.$$

The term, MA_j, is the market access from location j that encompasses the physical transportation network and market size distribution in China.[10] Denoting the residual of regressing $\ln w_j$ on $\ln L_j$ and $\ln \mathrm{MA}_j$ as $\tilde{\nu}_j$, it is then straightforward to see that $\bar{A}_j = \exp\left(\tilde{\nu}_j \frac{\theta+1}{\theta}\right)$. We estimate this regression with our sample of 291 cities and use $\theta = 4$ to back out the

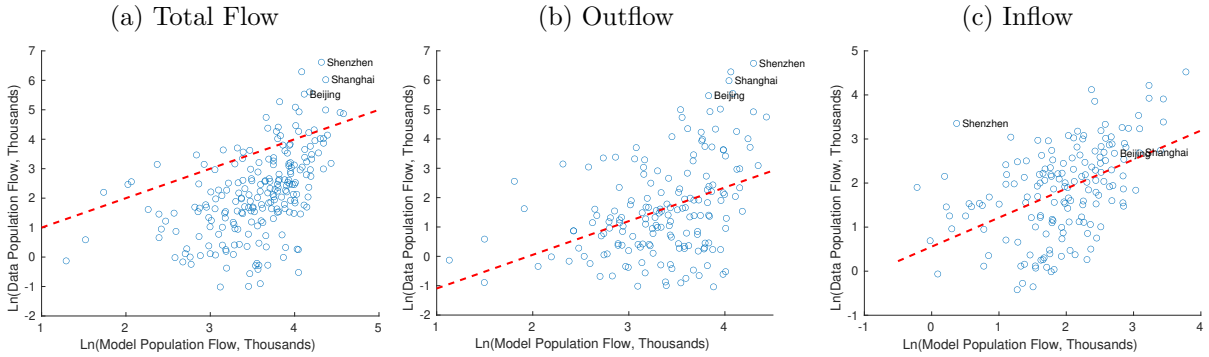
[10] The equation embodies the idea that the wage rate in a location depends on the agglomeration forces captured in L_j, the location advantage as summarized in MA_j, and the exogenous productivity in the residual term. See Appendix D for more details.

productivity term.

D.3 Out-of-Target Model Fit

Our model is calibrated to match the overall population flow into and out of the Hubei province around the year 2015. Consequently, these overall moments are exactly matched in the baseline quantification. Moreover, the T_{ij}^{p} matrix behind the bilateral migration costs captures the underlying geography and the transportation networks. The incorporation of such features allows us to match the bilateral population flows between prefectures in Hubei and prefectures outside Hubei, as shown in the three panels of Figure D.1 in the appendix. The baseline quantification of the model is able to fit the broad pattern of bilateral migration flows as the model prediction and the data are clustered around the 45-degree line. The popular destination cities among the outbound Hubei migrants in the data, such as Shenzhen, Shanghai, and Beijing, are also the top choices of the Hubei migrants in our model. The predicted outflow is more uniform across destinations compared to the data. This uniformity is mainly because the aggregation elasticity is relatively low ($\beta = 0.1$). As a result, the hot-spot cities in the data do not attract a sufficiently large population inflow in the baseline model.

Figure D.1: Bilateral Population Flow, Model v.s. Data



Note: This figure plots the model-predicted bilateral population flows between Hubei and the prefecture cities outside of Hubei, against the data. Only prefectures with positive population flows with the Hubei province are plotted. The red straight line is the 45 degree line.
Data source: *1% Population Sampling Survey of 2015*.

E Robustness Checks

E.1 Different Structural Parameters

In the baseline model, we calibrated β, κ, θ, and η using the common values from the literature. In this section, we check the robustness of the quantitative results with respect to these parameters. For each of the parameters listed above, we carry out two robustness checks: one with a value above the baseline level and the other below. In each of the robustness checks, we re-calibrate the four policy parameters ($\bar{\lambda}$, $\bar{\tau}$, λ_{IN}, and λ_{OUT}) and keep the other parameters the same as in the baseline version. We then re-compute the counterfactual results following the same steps as outlined in the paper. The parameter values in the robustness checks are reported in Table E.1. We report the number of infections under the "constant network and policy" counterfactual simulations in Figure E.1. The results under the other counterfactual simulations are similar and available upon request. As shown in the figure, the number of incidence under the counterfactual simulations is robust to these parameters' variations. Note that the counterfactual results do not vary at all to the changes in η. The irrelevance of η is because the elasticity of substitution only shifts the welfare level but does not interact with the trade and migration decisions, a feature common in the models based on Eaton and Kortum (2002).

E.2 Alternative Specifications

In this subsection, we demonstrate the robustness of the baseline findings to alternative empirical models of disease transmission.

Additional controls. We first augment the baseline regression model with additional control variables at the prefecture level, namely, agricultural employment share, manufacturing employment share, share of population aged 60 or above, and the share of male.[11] We consider that these variables are exogenous to the changes in transportation networks and migration policies in the quantitative exercises. Empirically, the question is whether there is evidence in the data that these factors are correlated with the main variables in the baseline

[11] These are aggregated variables based on the *1% Population Sampling Survey* in 2015, which are assembled and published by the provincial statistics bureau.

model, leading to biases in the estimates of $\beta_{\iota t}$. Figure E.2 reports the regression results. For the main variables of interest, the baseline estimates (red diamond points) always lie within the 90% confidence intervals of the estimates obtained from the augmented model (blue circle points). For the additional controls, all estimated coefficients are statistically insignificant. In columns (4) and (5) of Table E.2, we conduct the counterfactual simulations using the estimates of the augmented model and obtain very similar results. We take these findings as suggestive evidence that our baseline results are unlikely to be severely biased due to omitted factors that may have independent effects on local transmission.

Negative binomial regression. We adopt the negative binomial model to estimate the relationships between the number of infections and the economic fundamentals. The estimate of $\beta_{\iota t}$ are reported in Figure E.3. Except for the coefficients of $\ln(Pop_i)$, the baseline Poisson MLE estimates are always within the 90% confidence intervals of the negative binomial estimates. We then employ the negative binomial estimates to conduct counterfactual simulations. Columns (6) and (7) in Table E.2 show that if both transportation networks and migration policies had reverted to their 2005 configurations, the number of infections would have been lowered by 3,409 at the end of the sample period, which is 27.35% of the total reported infections. The results resemble the baseline findings.

Linear regression model with logarithm transformation. We also estimate the following linear model by the OLS regression

$$\ln(1+I_i(t)) = \beta_{0t}+\beta_{1t}X_i+\beta_{2t}M_i+\beta_{3t}\ln(Dist_i)+\beta_{4t}\ln(Pop_i)+\beta_{5t}\ln(GDPpc_i)+\varepsilon_i(t). \quad \text{(E.1)}$$

The consistency of OLS estimates does not depend on the assumption of the error distribution. Figure E.4 presents the OLS estimates. For X_i and $\ln(Pop_i)$, the OLS estimates are statistically different from the Poisson MLE estimates in the early sample period when there are more zero-value observations in the data. The ad hoc adjustment of left-hand-side variable for the zero counts appears to introduce biases.[12] The results of counterfactual

[12]We also consider the inverse hyperbolic sine transformation to adjust for zero counts, and obtain similar findings (available on request).

simulations based on the OLS estimates are reported in columns (8) and (9).[13] Compared to the baseline results, the linear specification with log transformation generates larger counterfactual changes in the earlier period. But the gap across different specifications diminishes over time.

Instrumentation strategy. One may concern that the travel distance to Hubei could be correlated with unobserved local socioeconomic factors that have independent effects on disease transmission. To address this potential problem, we follow Faber (2014) and construct the Minimum-Spanning Tree instrument. Intuitively, the instrument is based on the hypothetical routes that central planners would have constructed if the only policy objective had been to connect all targeted destinations on a single network subject to global construction cost minimization. Hence, it should be exogenous to the economic conditions of different locales. Since the baseline estimating equation is nonlinear, we perform this instrumentation strategy by adding the residuals from the first-stage regression and its squared term as a control function (CF) to the second stage (Wooldridge, 2002). The regression results are presented in Figure E.5. In Figure E.6, we conduct an IV estimation for the linear equation ((E.1)). We find that (i) the 90% confidence intervals of the estimates obtained from the CF model (respectively, the IV model) always contain the estimates from the baseline Poisson MLE model (respective, the OLS model); and (ii) for both specifications, the estimated coefficients of $\ln(Dist_i)$ based on the instrumentation strategy are somewhat larger in magnitude than the baseline estimates, although the differences are statistically insignificant. In columns (10)-(11), we conduct counterfactual simulations based on the estimates in Figure E.5. The counterfactual change in the cases of infection is slightly larger than the baseline estimate — if both transportation networks and migration policies had reverted to their 2005 configurations, the number of infections would have been declined by 32.75%. Columns (12) and (13) repeat the exercise based on the linear IV model, and obtain similar results.

The stability of the estimated coefficients obtained from the IV regression in Figure E.5 and the augmented model in E.2 suggests that, conditional on local income level and

[13] Using the OLS estimates, the counterfactual number of cases is calculated according to

$$I_i(t)^{CF} = \Big(1 + I_i(t)\Big) \exp\Big(\hat{\beta}_{1t}\Delta X_i + \hat{\beta}_{2\tau}\Delta M_i + \hat{\beta}_{3t}\Delta \ln(Dist_i) + \hat{\beta}_{4t}\Delta \ln(Pop_i) + \hat{\beta}_{5t}\Delta \ln(GDPpc_i)\Big) - 1,$$

population size, the travel distance with Hubei based on the pre-determined transportation networks could be exogenous to other unobserved determinants of local transmission. It is very challenging to find a valid instrument for GDP per capita. Therefore, the causal interpretation of the local income coefficient should be taken with caution. Although the potential confounding effects of the omitted variables are *a priori* ambiguous, we argue that they are unlikely to severely bias the quantitative results for the following reason. We find that the real income would have declined by 3.60% on average in the counterfactual scenario where the transportation networks and migration policy remained the same as in 2005. Based on the baseline estimate, this translates to a decline in the number of cases by 1.14% at the end of the sample period. This is gauged against the overall counterfactual change of 28.21% (column (3) of Table 2). Therefore, quantitatively, the baseline findings of the counterfactual experiments may not be too sensitive to the bias induced by the omitted variables that are correlated with local income, unless the bias is an order of magnitude larger than the baseline estimate. As is discussed above, the estimated coefficients of $\ln(GDPpc)$ remain stable at least to the additional controls such as local demographics and industry structure.

Additional robustness checks. In Figure E.7, we augment equation (1) with province $\times$ day fixed effects, which necessarily account for unobserved province-specific factors (e.g., local institutional quality, stringency of control measures) that may be correlated with the variables of interest and may have independent effects on the spread of the virus. Our baseline findings remain robust to this more stringent specification. To confirm that it is the migration flows specific to Hubei rather than migration flows per se that induced the propagation of COVID-19 in the early stages, we introduce additional controls of bilateral migrations with regions outside Hubei. The results are reported in Figure E.8. Reassuringly, the coefficients associated with migration flows specific to Hubei remain similar to those in the baseline regression. Moreover, bilateral migration flows with regions outside Hubei have economically and statistically insignificant effects on the spread of the virus.

Table E.1: Parameter Values in the Robustness Checks

Name	Baseline	Robustness Checks		Note
		Low	High	
β	0.1	0.05	0.2	Agglomeration elasticity
κ	2.0	1.4	3.3	Migration elasticity
θ	4.0	2.0	6.0	Trade elasticity
η	6.0	4.0	8.0	Elasticity of substitution

Note: This table lists the parameter values used in the baseline model and the robustness checks. We carry out two robustness checks per parameter with values both above and below the baseline level.

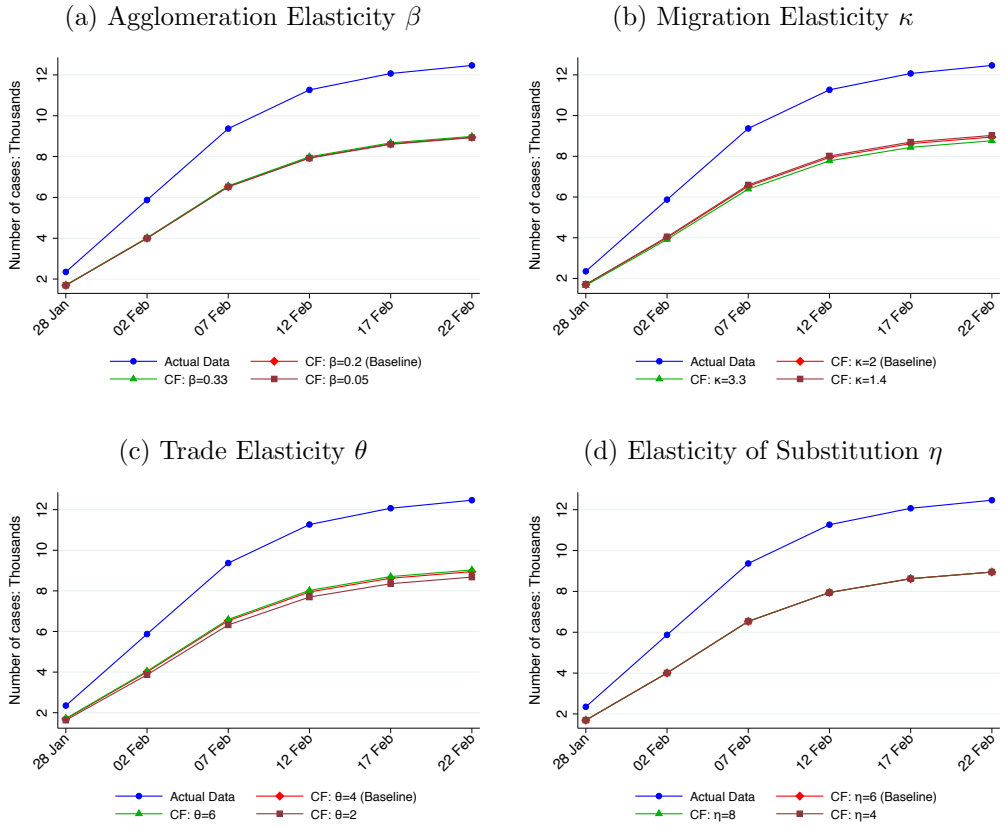
Table E.2: Incidence of COVID-19 Outside Hubei: Alternative Specifications (Constant Network & Policy)

Specification:	Actual	Counterfactual							
		Baseline Poisson		**Additional Controls**		**Negative Binomial**		**Linear** $\ln(\mathbf{1}+\mathbf{Case})$	
Date	Cases (1)	Cases (2)	Decline % (3)	Cases (4)	Decline % (5)	Cases (6)	Decline % (7)	Cases (8)	Decline % (9)
28/Jan/2020	2349	1695	27.83	1698	27.71	1636	30.35	1495	36.35
02/Feb/2020	5873	4005	31.81	4040	31.20	3935	33.00	3824	34.88
07/Feb/2020	9368	6532	30.28	6590	29.66	6542	30.16	6315	32.59
12/Feb/2020	11266	7946	29.47	7971	29.25	8009	28.91	7649	32.11
17/Feb/2020	12068	8622	28.55	8623	28.55	8669	28.16	8269	31.48
22/Feb/2020	12464	8947	28.21	8879	28.76	9055	27.35	8579	31.17

Specification:	**Poisson +CF**		**Linear+IV** $\ln(\mathbf{1}+\mathbf{Case})$	
Date	Cases (10)	Decline % (11)	Cases (12)	Decline % (13)
28/Jan/2020	1679	28.53	1667	29.05
02/Feb/2020	3897	33.65	3952	32.72
07/Feb/2020	6222	33.59	6352	32.19
12/Feb/2020	7516	33.29	7645	32.14
17/Feb/2020	8099	32.89	8246	31.67
22/Feb/2020	8382	32.75	8540	31.49

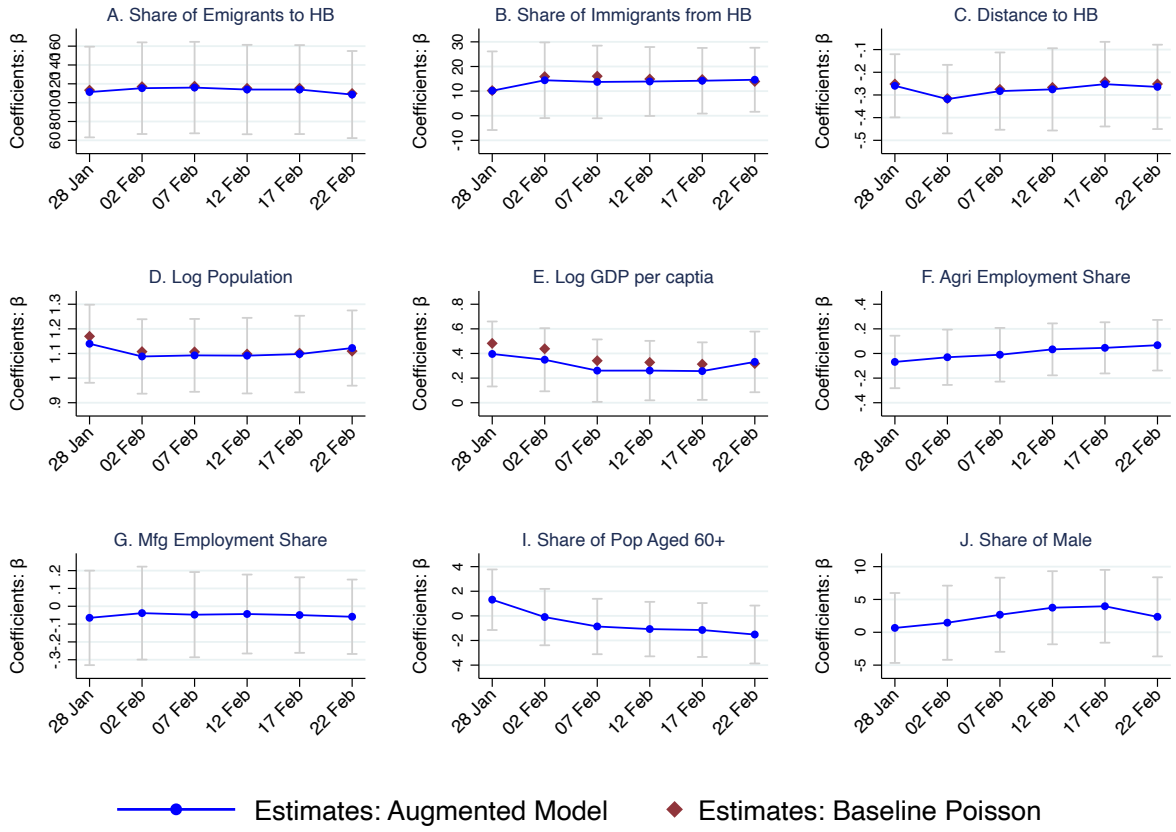
Note: This table reports the actual spread of reported COVID-19 cases over time, and the spreads under the counterfactual scenario "Constant Network & Policy" with both the T_{ij}^{p} and T_{ij}^{g} matrices and the Λ parameters remained the same as in 2005. Columns 2 and 3 repeat the baseline results. The remaining columns repeat the counterfactual simulation, but replace $\hat{\beta}_{ut}$ with the estimates from alternative regression models, namely: (i) the augmented model with additional controls in Figure E.2 (columns 4 and 5); (ii) the negative binomial model in Figure E.3 (columns 6 and 7); (iii) the linear model with dependent variable $\ln(1+\text{Case})$ in Figure E.4 (columns 8 and 9); (iv) the Poisson model with control functions in Figure E.5 (columns 10 and 11); and (v) the linear IV model in Figure E.6 (columns 12 and 13).

Figure E.1: Counterfactual Changes in COVID-19 Transmission under Different Parameterizations



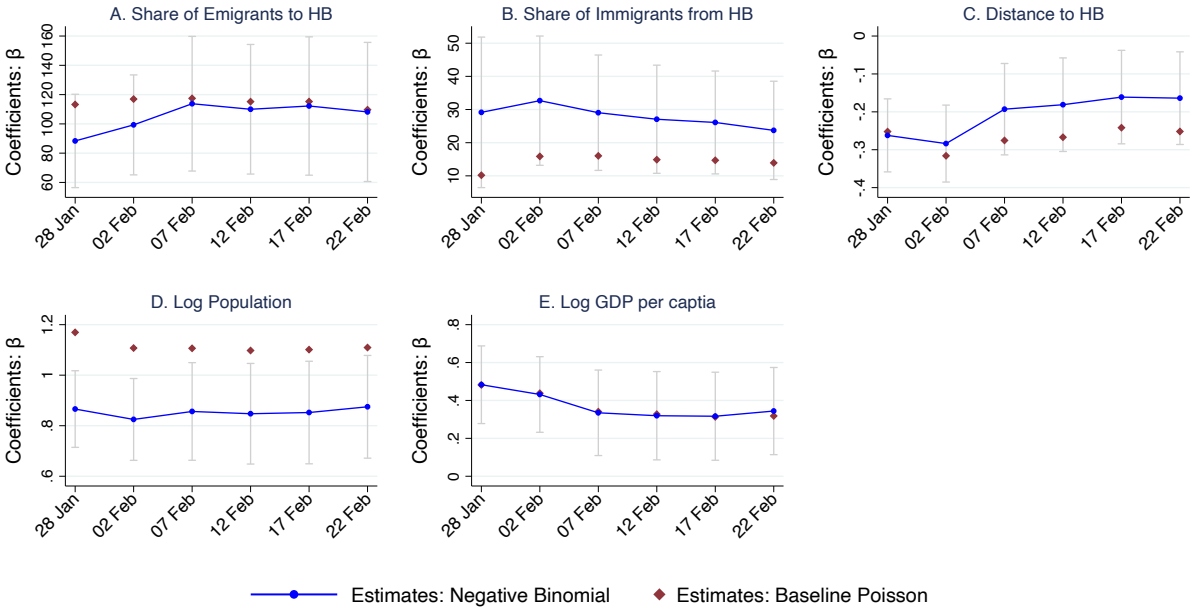
Note: This figure presents the number of infections in the robustness checks under the "constant networks and policy" counterfactual simulations. The results using the baseline parameter values are also reported as references.

Figure E.2: Estimates of Cumulative Effects: $\beta_{\iota t}$ (Including Additional Controls)



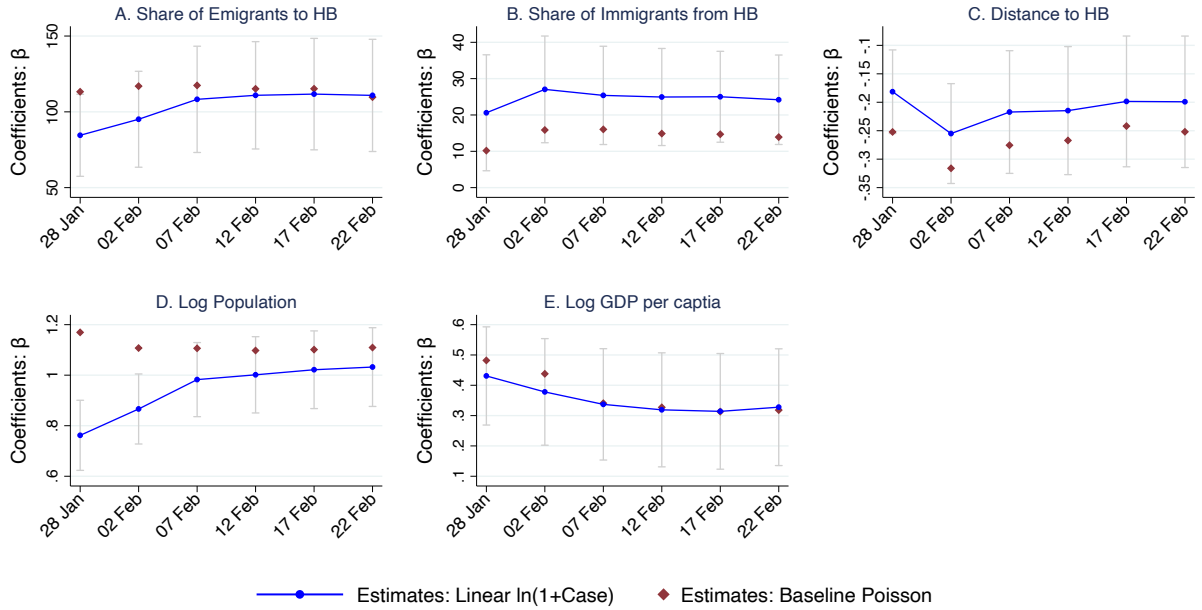
Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\beta_{\iota t}$ coefficients. For comparison purposes, the baseline estimates (in Figure 2) are reported in the figures and represented by the red diamond points.

Figure E.3: Estimates of Cumulative Effects: $\beta_{\iota t}$ (Negative Binomial)



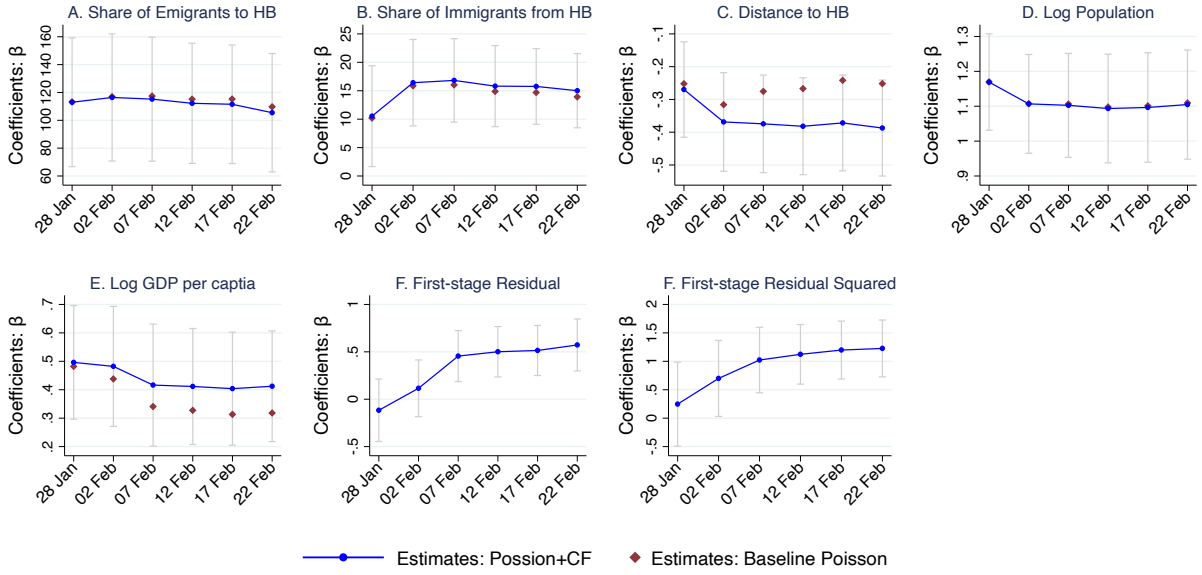
Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\beta_{\iota t}$ coefficients. For comparison purposes, the baseline estimates (in Figure 2) are reported in the figures and represented by the red diamond points.

Figure E.4: Estimates of Cumulative Effects: $\beta_{\iota t}$ (Linear Model with Log Transformation)



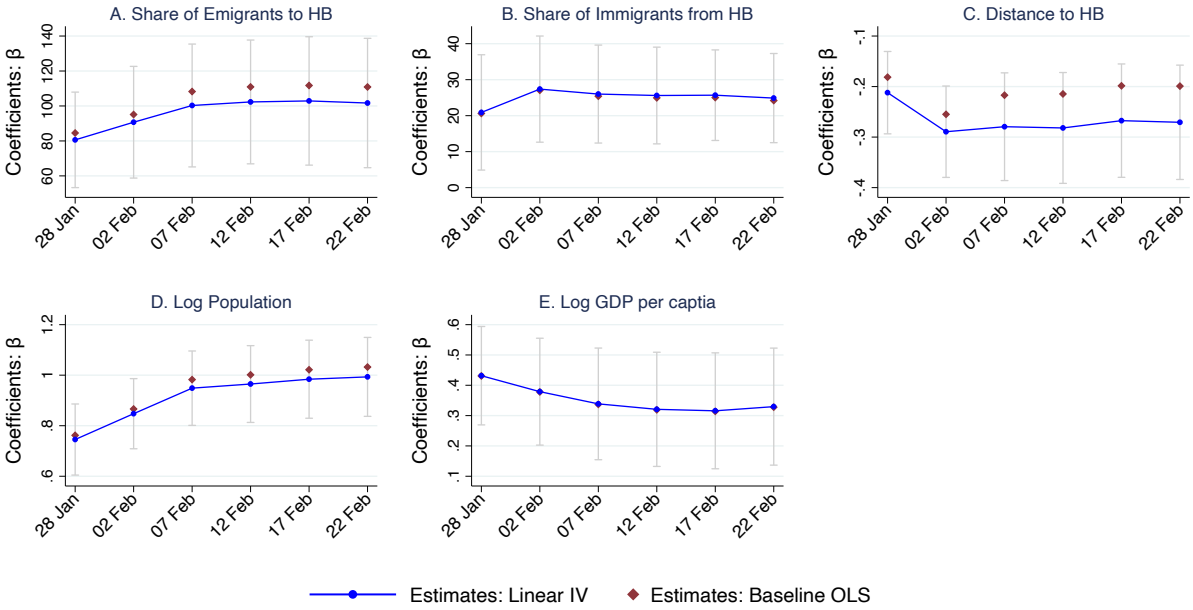
Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\beta_{\iota t}$ coefficients. For comparison purposes, the baseline estimates (in Figure 2) are reported in the figures and represented by the red diamond points.

Figure E.5: Estimates of Cumulative Effects: $\beta_{\iota t}$ (Poisson + Control Functions)



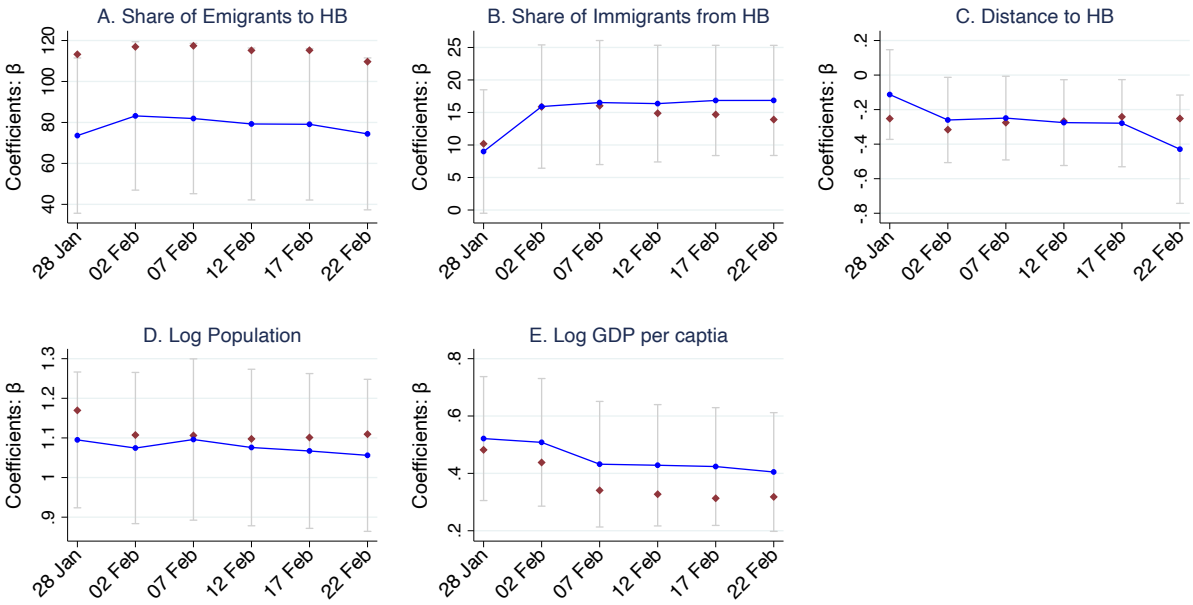
Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\beta_{\iota t}$ coefficients. For comparison purposes, the baseline estimates (in Figure 2) are reported in the figures and represented by the red diamond points.

Figure E.6: Estimates of Cumulative Effects: $\beta_{\iota t}$ (Linear + IV)



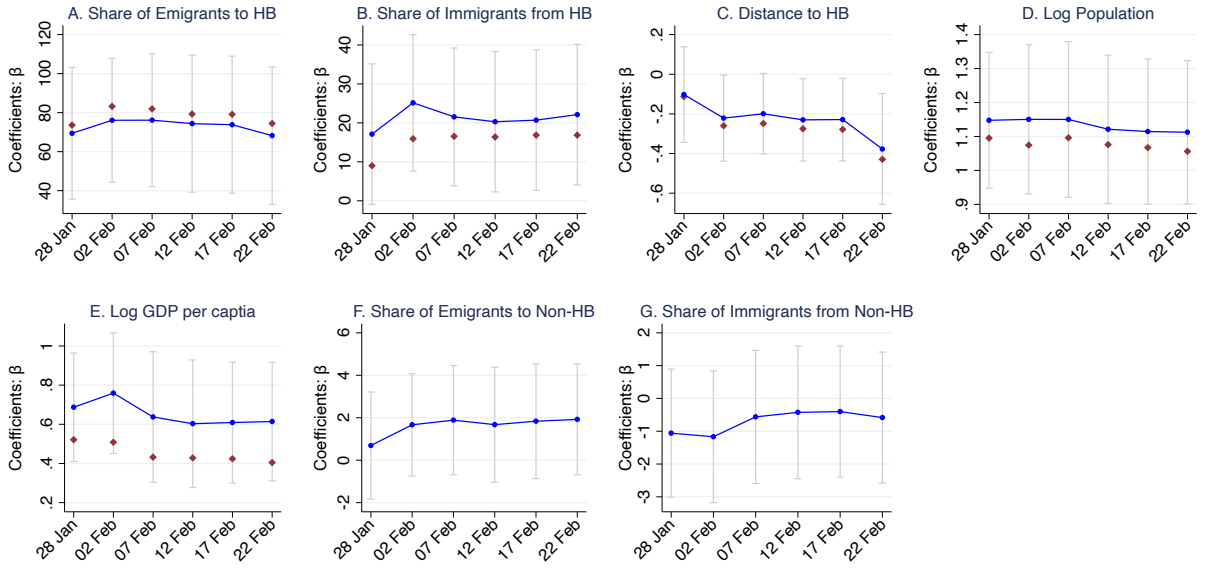
Note: This figure plots the point estimates and the corresponding 90% confidence intervals of the $\beta_{\iota t}$ coefficients. For comparison purposes, the estimates based on the OLS model (in Figure E.4) are reported in the figures and represented by the red diamond points.

Figure E.7: Estimates of Policy Parameters: β_{ut} (Controlling for Province×Date FEs)



Note: This figure reports the point estimates and the corresponding 90% confidence intervals of regression model (1) augmented with province × date fixed effects. For comparison purposes, the baseline estimates (in Figure 2) are reported in the figures and represented by the red diamond points.

Figure E.8: Estimates of Policy Parameters: β_{it} (Controlling for Migration Flows Outside Hubei)



Note: This figure reports the point estimates and the corresponding 90% confidence intervals of regression model (1) augmented with province × date fixed effects, and the bilateral migration flows with regions outside Hubei. The explanatory variable in panel F is the ratio of emigrants to regions outside Hubei to the local population in prefecture i in 2015. The explanatory variable in panel G is the share of immigrants from Hubei in the local population in prefecture i in 2015. For comparison purposes, the baseline estimates (in Figure E.7) are reported in the figures and represented by the red diamond points.