The Geography of Development

Klaus Desmet

Southern Methodist University

Dávid Krisztián Nagy

Centre de Recerca en Economia Internacional

Esteban Rossi-Hansberg

Princeton University

We develop a dynamic spatial growth theory with realistic geography. We characterize the model and its balanced-growth path and propose a methodology to analyze equilibria with different levels of migration frictions. Different migration scenarios change local market size, innovation incentives, and the evolution of technology. We bring the model to the data for the whole world economy at a $1° \times 1°$ geographic resolution. We then use the model to quantify the gains from relaxing migration restrictions. Our results indicate that fully liberalizing migration would increase welfare about threefold and would significantly affect the evolution of particular regions of the world.

I. Introduction

An individual's place of residence is essential in determining her productivity, income, and well-being. A person's location, however, is neither a

We thank Treb Allen, Costas Arkolakis, Maya Buchanan, Angus Deaton, Gordon Hanson, Tom Holmes, Chang-Tai Hsieh, Erik Hurst, Steve Redding, Frédéric Robert-Nicoud, Andrés Rodríguez-Clare, Jesse Shapiro, Joe Shapiro, Áron Tóbiás, Rob Townsend, and five referees for comments and discussions. We also wish to thank Adrien Bilal, Mathilde Le Moigne,

Electronically published May 7, 2018
[*Journal of Political Economy*, 2018, vol. 126, no. 3]
© 2018 by The University of Chicago. All rights reserved. 0022-3808/2018/12603-0005$10.00

permanent characteristic nor a fully free choice. People tend to flee undesirable and low-productivity areas to go to places that offer better opportunities, but these choices are often hindered by an assortment of restrictions. An obvious example is the effort to stop undocumented migration to Europe, the United States, and most developed countries. How do these restrictions affect the evolution of the world economy? How do they interact with today's production centers, as well as today's most desirable places to live, to shape the economy of the future? Any attempt to answer these questions requires a theory of development that explicitly takes into account the spatial distribution of economic activity, the mobility restrictions and transport costs associated with it, and the incentives for innovation implied by the world's economic geography. Once we have a basic understanding of the role of geography in development, we can start evaluating the impact of events that change this geography.

Constructing a theory to study the effect of geography on development requires incorporating some well-known forces as well as others that have received, so far, less attention. First, a particular location is unique because of where it is relative to other locations, which determines its costs of trading goods. Furthermore, each location has particular amenities that determine its desirability as a place to live and a particular productivity level that determines its effectiveness as a place to produce and work. This singularity of individual places underscores the importance of bringing the actual geography of the world, as measured by the location of land and water, as well as the distribution of other local spatial and economic characteristics, into the analysis. Second, migration across and within countries is possible but limited, partly because of institutional restrictions and partly because of social norms and other mobility costs. National borders restrain mobility well beyond the existing frictions within countries, but frictions within countries are also potentially large. Third, the distinct levels of labor productivity of locations, which reflect their institutions, infrastructure, education systems, and capital stocks, as well as location-specific technological know-how, evolve over time. Firms can invest in improving this local technology and infrastructure. Their incentives to do so depend on the size of the market to which they can sell their products. This market size is determined by the magnitude of transport costs and the location's geography relative to the potential customers of the product. Not all improvements in technology are local in nature or are the result of purposeful investments, though. Firms also benefit from the diffusion of the innovations and creativity of others. Fourth, a location's population density affects its productivity, its incentives to innovate, and, perhaps most important, its amenities. Large concentrations of people in, for example, ur-

Charly Porcher, and Maximilian Vogler for excellent research assistance. Data are provided as supplementary material online.

ban areas benefit from agglomeration effects but also suffer the undesirable costs from congestion.

Our aim is to study the evolution of the world economy at a rich level of geographic detail ($1° \times 1°$ resolution) over many years. So our analysis naturally involves many choices and compromises. The basic structure of the model is as follows. Each $1° \times 1°$ cell of the world contains firms that produce a variety of goods using location-specific technologies that employ labor and a local factor we refer to as land. Firms can trade subject to iceberg transport costs. Each location is endowed with amenities that enhance the quality of life in that cell. Agents have stochastic idiosyncratic locational preferences drawn independently every period from a Fréchet distribution. The static spatial equilibrium resembles the one proposed in Allen and Arkolakis (2014), but with migration, local factors, a trade structure à la Eaton and Kortum (2002), and heterogeneous preferences as in Kline and Moretti (2014).[1] The dynamic model uses this structure and allows firms to invest in improving local technology as in Desmet and Rossi-Hansberg (2014). Technological innovations depend on a location's market size, which is a function of transport costs and the entire spatial distribution of expenditure. Local innovations determine next period's productivity after taking into account that part of these technologies disseminate over space.

We not only characterize the distribution of economic activity in the balanced-growth path but are also able to compute transitions since, as in Desmet and Rossi-Hansberg (2014), the innovation decision can be reduced to a simple static problem due to land competition and technological diffusion. Our ability to prove the uniqueness of the equilibrium, to characterize the steady state, to identify the initial distribution of amenities, productivity levels, and migration costs in all locations, and to simulate the model, though novel, is importantly enhanced by the set of results in Zabreyko et al. (1975), first used in spatial models in the related static and single-country model of Allen and Arkolakis (2014). So we owe a substantial debt to that work.

Identifying the reasons why agents locate in a particular place necessarily requires taking a stand on the opportunities they have to move across locations in search of better living and work opportunities. Under the assumption of free mobility, a long tradition in urban and regional economics has identified productivity and amenities across locations using land prices and population counts (see, e.g., Roback [1982] following the hedonic approach of Rosen [1979]). More recently, another strand

[1] Since the Allen and Arkolakis (2014) framework is isomorphic to a setup with local factors and a production and trade structure as in Eaton and Kortum (2002), the key difference in the static part of the model is our introduction of migration costs.

of the literature has used income per capita and population counts, together with a spatial equilibrium model, to identify these same local characteristics (see, among many others, Desmet and Rossi-Hansberg [2013], Allen and Arkolakis [2014], Fajgelbaum and Redding [2014], and Behrens et al. [2017]).

We follow this second strand of the literature but note that the static nature of these papers yields a decomposition that depends crucially on parameters that are likely to evolve with the level of development of the economy. That is, the parameters used to identify amenities and productivities depend on the particular time period when the exercise was done. As a result, they do not represent stable parameters that one can use to obtain meaningful conclusions in a dynamic context. In fact, in our theoretical framework, the relationship between income, population density, and amenities evolves over time as the economy becomes richer and slowly converges to a balanced-growth path. We provide empirical evidence consistent with these theoretical patterns, using data from different regions of the world, as well as from US counties and zip codes.

Another problematic assumption in this literature is free mobility, particularly because we are analyzing not only migration across regions within a country but also migration across all countries in the world. While the assumption of free mobility within countries is clearly imperfect, although perhaps acceptable in some contexts, it is hard to argue that people anywhere can freely move to the nicest and most productive places on earth. To see this, it suffices to have a casual look at the United States–Mexico border or the restrictions on African immigration in Europe. More important, ignoring mobility restrictions leads to unreasonable conclusions. Take the example of the Democratic Republic of the Congo, a country with the same population density as the United States but with real wages that are orders of magnitude lower. The only way in which standard economic geography models with free mobility can reconcile this fact is to assume that Congo has some of the best amenities on earth. Even though it is hard to take a definitive stand on what characteristics a country's individuals enjoy the most, and the heterogeneity in their preferences, basic evidence on health, education, governance, and institutions suggests that such a conclusion masks the fact that many people in Congo do not choose to live there, but instead are trapped in an undesirable location. Thus, we incorporate migration frictions within and across countries in our analysis.

Once we explicitly account for migration restrictions, and therefore utility differences across space, we get a more nuanced picture. To understand why, note that our theory identifies amenities only relative to utility at each location. Hence, in the absence of mobility restrictions, the large values of amenities relative to utility in Congo would show up as those regions having high amenity levels. But if the Congolese face high mobility costs, then the same large values of amenities relative to utility would

show up as Congo having a low utility level. Identifying the actual amenity levels therefore involves incorporating more data. To do so, we use survey data on Cantril ladder measures of subjective well-being from the Gallup World Poll.

The subjective well-being data are an evaluative measure that asks individuals to assess their lives on a ladder scale, from the worst possible to the best possible life they can envision for themselves. Deaton (2008) and Kahneman and Deaton (2010) argue that this measure correlates well with log income and does not exhibit a variety of well-known pathologies that afflict hedonic measures of subjective well-being or happiness. We therefore interpret this evaluative measure as giving us information on the welfare of individuals. Still, we need to convert this ladder with 11 steps into a cardinal measure of the level of utility. To do so, we match the relationship between the ladder measure and log income in the model and in the data. Using the cardinal measures of utility together with the amenity to utility ratios, we can recover the actual level of amenities for each cell of the world. As an overidentification check, we find that these estimated amenities correlate well with commonly used exogenous measures of quality of life.[2]

We can then use the evolution of population in the model together with data on population counts in each location for two subsequent periods to compute the cost of moving in and out of each location in the world. This identifies mobility costs between all locations, both within and across countries, as the product of an origin- and a destination-specific cost. These migration costs will be key to quantitatively assessing different migration policies, such as keeping costs constant in the future or fully liberalizing mobility.

We calibrate the rest of the model using data on the evolution of output across countries and other information from a variety of sources. We then perform a number of experiments in which we simulate the transition of the world economy to its balanced-growth path. The parameters we estimate guarantee that the equilibrium is unique and that the economy eventually converges to a balanced-growth path in which the geographic distribution of economic activity is constant. The transition to this balanced-growth path can, however, take very long. If current migration frictions do not change, it takes about 400 years for the economy to reach its balanced-growth path. The protracted length of this transition is the result of the sequential development of clusters due to the complexity of the world's geography. During these 400 years the world experiences

[2] This suggests that subjective well-being differences capture an essential part of utility differences across countries, although both concepts are unlikely to exactly coincide (see also Glaeser, Gottlieb, and Ziv 2016).

large changes. The world's real output growth rate increases progressively to around 2.9 percent by 2100 and then decreases back to 2.8 percent, while the growth rate of welfare increases from around 2.4 percent to 2.8 percent. The correlation between GDP per capita and population density also changes dramatically. The world goes from the current negative correlation of around −.41 to a high correlation of .65 in the balanced-growth path.[3] That is, in contrast to the world today, where many densely populated areas are poor, in the future the dense regions will be the wealthy regions.[4]

Before using our quantified framework as a tool to evaluate the role of migratory restrictions and potentially other spatial frictions, it is important to gauge its performance using data that were not directly used in the quantification. For this purpose we develop a method to solve our model backward, allowing us to compute the model-implied distribution of population in the past. We run the model backward for 130 years and compare the country population levels and growth rates predicted by the model to those in the Penn World Tables and Maddison (2001). The results are encouraging. The model does very well in matching population levels, and it also performs quite well in matching population growth rates. For example, we find that the correlation of the population growth rates between 1950 and 2000 implied by the model and those observed in the data is more than .7. This number declines somewhat for other time periods, but the model preserves predictive power, even going as far back as 1870, in spite of many historical shocks, such as World Wars I and II, not being included in the analysis.

Relaxing migration restrictions leads to large increases in output and welfare at impact. The growth rates of real GDP and welfare also unambiguously increase in the balanced-growth path, with the magnitude of the effect depending on the degree of liberalization. With the current frictions, about 0.3 percent of people migrate across countries and about 0.45 percent move across cells in a year today (which is matched exactly by the model), with this percentage converging to zero in the balanced-growth path. If, instead, we drop all restrictions, so there is free mobility, at impact 70.3 percent of the population moves across countries and 71.6 percent across cells. In present discounted value terms, complete liberalization yields output gains of 126 percent and welfare gains of 306 percent. Although this experiment is somewhat extreme and we also compute the effects of partial liberalizations, it illustrates the large magnitude of the gains at stake and it highlights the role of migration policies when thinking about the future of the world economy.

[3] The correlation is computed using $1° \times 1°$ land cells as units.

[4] Consistent with wealthy regions having a stronger correlation between population density and income per capita, in 2000 the correlation was only −.11 in Africa and as high as .50 in North America.

Different levels of migration restrictions put the world on alternative development paths in which the set of regions that benefit varies dramatically. With the current restrictions, we get a productivity reversal, with many of today's high-density, low-productivity regions in sub-Saharan Africa, South Asia, and East Asia becoming high-density, high-productivity regions, and North America and Europe falling behind in terms of both population and productivity. In contrast, when we relax migration restrictions, Europe and the eastern areas of the United States remain strong, with certain regions in Brazil and Mexico becoming important clusters of economic activity too.

The driving forces behind these results are complex since the world is so heterogeneous. One of the key determinants of these patterns is the correlation between GDP per capita and population density. As we mentioned above, the correlation is negative and weak today, and our theory predicts that, consistent with the evidence across regions in the world today, this correlation will become positive and grow substantially over the next six centuries, as the world becomes richer. Two forces drive this result. First, people move to more productive areas, and second, more dense locations become more productive over time since investing in local technologies in dense areas is, in general, more profitable. Migration restrictions shift the balance between these two mechanisms. If migration restrictions are strict, people tend to stay where they are, and today's dense areas, which often coincide with developing countries, become the most developed parts of the world in the future. If, in contrast, migration is free, then people move to the most productive, high-amenity areas. This tends to favor today's developed economies. Liberalizing migration improves welfare so much because it makes the high-productivity regions in the future coincide with the high-amenity locations. So relaxing migration restrictions eliminates the productivity reversal that we observe when migration restrictions are kept constant.

These results highlight the importance of geography, and the interaction of geography with factor mobility, for the future development path of the world economy. Any policy or shock that affects this geography can have potentially large effects through similar channels. One relative strength of our framework compared to the current literature is that, by explicitly modeling the evolution of local technology over time, it incorporates into the analysis the effect of spatial frictions on productivity. By doing so, it accounts for the future impact of migrants on local productivity and amenities, rather than for just their immediate impact on congestion and the use of local factors. The resulting growth effects can be large, suggesting that this interaction between spatial frictions and productivity should be an essential element in any analysis of the impact of migration restrictions.

The rest of the paper is organized as follows. Section II presents the model and proves the existence and uniqueness of the equilibrium. Section III provides a sufficient condition on the parameters for a unique balanced-growth path to exist. Section IV discusses the calibration of the model, including the inversion to obtain initial productivity and amenity values, the methodology to estimate migration costs, and the algorithm to simulate the model. Section V presents our numerical findings, which include the results for the benchmark calibration and the results for different levels of migration frictions. Section VI presents conclusions. Appendix A presents empirical evidence of the correlation of density and productivity, and it shows how our estimates of amenities correlate with exogenous measures of quality of life. It also discusses the robustness of our results to changes in different parameter values. Appendix B presents the proofs not included in the main text. Appendix C provides a summary of the data sources. Videos with simulations of the world economy for different migration scenarios are available online.

II. The Model

Consider an economy that occupies a closed and bounded subset S of a two-dimensional surface that has positive Lebesgue measure. A location is a point $r \in S$. Location r has land density $H(r) > 0$, where $H(\cdot)$ is an exogenously given continuous function that we normalize so that $\int_S H(r)\,dr = 1$. There are C countries. Each location belongs to one country; hence countries constitute a partition of S: $(S_1, \ldots, S_C)$. The world economy is populated by $\bar{L}$ agents who are endowed with one unit of labor, which they supply inelastically. The initial population distribution is given by a continuous function $\bar{L}_0(r)$.

A. *Preferences and Agents' Choices*

Every period agents derive utility from local amenities and from consuming a set of differentiated products according to constant elasticity of substitution preferences. The period utility of an agent i who resides in r this period t and lived in a series of locations $\bar{r}_- = (r_0, \ldots, r_{t-1})$ in all previous periods is given by

$$u_t^i(\bar{r}_-, r) = a_t(r)\left[\int_0^1 c_t^\omega(r)^\rho\,d\omega\right]^{1/\rho} \varepsilon_t^i(r)\prod_{s=1}^{t} m(r_{s-1}, r_s)^{-1}, \tag{1}$$

where $1/[1-\rho]$ is the elasticity of substitution with $0 < \rho < 1$, $a_t(r)$ denotes amenities at location r and time t, $c_t^\omega(r)$ denotes consumption of good ω at location r and time t, $m(r_{s-1}, r_s)$ represents the permanent

flow-utility cost of moving from r_{s-1} in period $s-1$ to r_s in period s, and $\varepsilon_t^i(r)$ is a taste shock distributed according to a Fréchet distribution. We assume that the log of the idiosyncratic preferences has constant mean proportional to Ω and variance $\pi^2\Omega^2/6$ with $\Omega < 1$. Thus,

$$\Pr[\varepsilon_t^i(r) \leq z] = e^{-z^{-1/\Omega}}.$$

A higher value of Ω indicates greater taste heterogeneity. We assume that $\varepsilon_t^i(r)$ is independent and identically distributed (i.i.d.) across locations, individuals, and time.

Agents discount the future at rate β, and so the welfare of an individual i in the first period is given by $\Sigma_t \beta^t u_t^i(r_{t-}^i, r_t^i)$, where r_t^i denotes her location choice at t, r_{t-}^i denotes the history of locations before t, and r_0^i is given. Amenities take the form

$$a_t(r) = \bar{a}(r)\bar{L}_t(r)^{-\lambda}, \tag{2}$$

where $\bar{a}(r) > 0$ is an exogenously given continuous function, $\bar{L}_t(r)$ is population per unit of land at r in period t, and λ is a fixed parameter, where $\lambda \geq 0$.[5] Thus, we allow for congestion externalities in local amenities as a result of high population density, with an elasticity of amenities to population given by $-\lambda$.

An agent earns income from work, $w_t(r)$, and from the local ownership of land.[6] Local rents are distributed uniformly across a location's residents.[7] So if $R_t(r)$ denotes rents per unit of land, then each agent receives land rent income $R_t(r)/\bar{L}_t(r)$. Total income of an agent in location r at time t is therefore $w_t(r) + R_t(r)/\bar{L}_t(r)$. Agents cannot write debt contracts with each other. Thus, every period agents simply consume their income, and so

$$\begin{aligned} u_t^i(\bar{r}_-, r) &= \frac{a_t(r)}{\prod_{s=1}^t m(r_{s-1}, r_s)} \frac{w_t(r) + R_t(r)/\bar{L}_t(r)}{P_t(r)} \varepsilon_t^i(r) \\ &= \frac{a_t(r)}{\prod_{s=1}^t m(r_{s-1}, r_s)} y_t(r)\varepsilon_t^i(r), \end{aligned}$$

where $y_t(r)$ denotes the real income of an agent in location r, and $P_t(r)$ denotes the ideal price index at location r in period t, where

[5] This is consistent with the positive value of λ we find in our estimation, although the theory could in principle allow for a negative number, in which case amenities would benefit from positive agglomeration economies.

[6] We drop the i superscript here because all agents in location r earn the same income.

[7] See Caliendo et al. (2018) for alternative assumptions on land ownership and their implications.

$$P_t(r) = \left[\int_0^1 p_t^{\omega}(r)^{-\rho/(1-\rho)} d\omega\right]^{-(1-\rho)/\rho}.$$

Every period, after observing their idiosyncratic taste shock, agents decide where to live subject to permanent flow-utility bilateral mobility costs $m(s, r)$. These costs are paid in terms of a permanent percentage decline in utility. In what follows we let $m(s, r) = m_1(s)m_2(r)$, with $m(r, r) = 1$ for all $r \in S$. These assumptions guarantee that there is no cost to staying in the same place and that the utility discount from moving from one place to another is the product of an origin-specific and a destination-specific discount. Furthermore, these assumptions also imply that $m_1(r) = 1/m_2(r)$. Hence, a migrant who leaves a location r will receive a benefit (or pay a cost) $m_1(r)$, which is the inverse of the cost (or benefit) $m_2(r) = 1/m_1(r)$ of entering that same location r. As a result, a migrant who enters a country and leaves that same country after a few periods will end up paying the entry migration costs only while being in that country. This happens because, although the flow-utility mobility costs are permanent, the cost of entering is compensated by the benefit from leaving.

Our theory focuses on net, rather than gross, migration flows, since local population levels are what determines innovation and hence the evolution of the global economy. Important, these assumptions on migration costs do not impose a restriction on our ability to match data on changes in population. As we will later show, because the migration cost between any pair of locations in the world consists of an origin- and a destination-specific cost, we can use observations on population levels in each location for two subsequent periods to exactly identify all migration costs. We summarize these assumptions in assumption 1.

Assumption 1. Bilateral moving costs can be decomposed into an origin- and a destination-specific component, so $m(s, r) = m_1(s)m_2(r)$. Furthermore, there are no moving costs within a location, so $m(r, r) = 1$ for all $r \in S$.

Independently of the magnitude of migration costs, preference heterogeneity implies that the elasticity of population with respect to real income adjusted by amenities is not infinite. This elasticity is governed by the parameter Ω, which determines the variance of the idiosyncratic preference distribution. Conditional on a location's characteristics, summarized by $a_t(r)y_t(r)$, a location with higher population has lower average flow utility, since the marginal agent has a lower preference to live in that location. In that sense, preference heterogeneity acts like a congestion force, an issue we will return to later on.

Assumption 1 implies that the location choice of agents depends only on current variables and not on their history or the future characteristics of the economy. The value function of an agent living at r_0 in period 0,

after observing a distribution of taste shocks in all locations, $\bar{\varepsilon}_1^i \equiv \varepsilon_1^i(\cdot)$, is given by

$$\begin{aligned} V(r_0, \bar{\varepsilon}_1^i) &= \max_{r_1} \left[\frac{a_1(r_1)}{m(r_0, r_1)} y_1(r_1)\varepsilon_1^i(r_1) + \beta E\left(\frac{V(r_1, \bar{\varepsilon}_2^i)}{m(r_0, r_1)} \right) \right] \\ &= \frac{1}{m_1(r_0)} \max_{r_1} \left[\frac{a_1(r_1)}{m_2(r_1)} y_1(r_1)\varepsilon_1^i(r_1) + \beta E\left(\frac{V(r_1, \bar{\varepsilon}_2^i)}{m_2(r_1)} \right) \right] \\ &= \frac{1}{m_1(r_0)} \left\{ \max_{r_1} \left[\frac{a_1(r_1)}{m_2(r_1)} y_1(r_1)\varepsilon_1^i(r_1) \right] \right. \\ &\qquad \left. + \beta E\left(\max_{r_2} \left[\frac{a_2(r_2)}{m_2(r_2)} y_2(r_2)\varepsilon_2^i(r_2) + \frac{V(r_2, \bar{\varepsilon}_3^i)}{m_2(r_2)} \right] \right) \right\}, \end{aligned}$$

where the second and third lines use assumption 1. Hence, since $[a_1(r_1)/m_2(r_1)]y_1(r_1)\varepsilon_1^i(r_1)$ depends only on current variables and taste shocks, the decision of where to locate in period 1 is independent of the past history and future characteristics of the economy. That is, the value function adjusted for the value of leaving the current location $V(r, \varepsilon^i)/m_2(r)$ (which is equal to $V(r, \varepsilon^i)\, m_1(r)$ by assumption 1) is independent of the current location r. This setup implies that the location decision is a static one and that we do not need to keep track of people's past location histories, a feature that enhances the tractability of our framework substantially.

Consistent with this, we can show that an individual's flow utility depends only on her current location and on where she was in period 0 (which is not a choice). Using (1) and taking logs, the period t log utility of an agent who resided in r_0 in period 0 and lives in r_t in period t is

$$\tilde{u}_t^i(r_0, r_t) = \tilde{u}_t(r_t) - \tilde{m}_1(r_0) - \tilde{m}_2(r_t) + \tilde{\varepsilon}_t^i(r_t),$$

where $\tilde{x} = \ln x$ and $u_t(r)$ denotes the utility level associated with local amenities and real consumption, so

$$u_t(r) = a_t(r)\left[\int_0^1 c_t^\omega(r)^\rho d\omega\right]^{1/\rho} = a_t(r)y_t(r). \tag{3}$$

Note that $u_t(r)$ summarizes fully how individuals value the production and amenity characteristics of a location. Hence, it is a good measure of the desirability of a location, and it will be one of the measures we use to evaluate social welfare. However, it does not include the mobility costs incurred to get there or the idiosyncratic preferences of individuals who live there.

As an alternative measure to evaluate social welfare, we include taste shocks into our measure, while still ignoring the direct utility effects of migration costs. As we will argue, it is more meaningful to leave out mo-

bility costs because very often the lack of migration between two locations reflects a legal impossibility of moving (as well as lack of information or ex ante psychological impediments) rather than an actual utility cost once the agent has moved. Hence, including migration costs when evaluating the social welfare effects of liberalizing migration restrictions would tend to grossly overestimate the gains. To compute this alternative measure, suppose therefore that individuals move across locations assuming they have to pay the cost $m(\cdot, \cdot)$ but that the ones that move get reimbursed for the whole stream of moving costs ex post. Then the expected period t utility of an agent i who resides in r is

$$\begin{aligned} &E(u_t(r)\varepsilon_t^i(r) | i \text{ lives in } r) \\ &= \Gamma(1-\Omega) m_2(r) \left[\int_S u_t(s)^{1/\Omega} m_2(s)^{-1/\Omega} ds \right]^{\Omega}, \end{aligned} \tag{4}$$

where Γ denotes the gamma function. Section A in appendix B provides details on how to obtain this expression.

We now derive expressions of the shares of people moving between locations. The density of individuals residing in location s in period $t-1$ who prefer location r in period t over all other locations is given by

$$\begin{aligned} \Pr(\tilde{u}_t(s, r) \geq \tilde{u}_t(s, v) \ \forall \ v \in S) &= \frac{\exp([\tilde{u}_t(r) - \tilde{m}_2(r)]/\Omega)}{\int_S \exp([\tilde{u}_t(v) - \tilde{m}_2(v)]/\Omega) dv} \\ &= \frac{u_t(r)^{1/\Omega} m_2(r)^{-1/\Omega}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv}. \end{aligned} \tag{5}$$

This corresponds to the fraction of the population in location s that moves to location r,

$$\frac{\ell_t(s, r)}{H(s)\bar{L}_{t-1}(s)} = \frac{u_t(r)^{1/\Omega} m_2(r)^{-1/\Omega}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv}, \tag{6}$$

where $\ell_t(s, r)$ denotes the number of people moving from s to r in period t (or that stayed in r for $\ell_t(r, r)$) and $\bar{L}_{t-1}(s)$ denotes the total population per unit of land in s at $t-1$. The number of people living at r at time t must coincide with the number of people who moved there or stayed there, so

$$H(r)\bar{L}_t(r) = \int_S \ell_t(s, r) ds.$$

Using (6), this equation can be written as

$$\begin{aligned} H(r)\bar{L}_t(r) &= \int_S \frac{u_t(r)^{1/\Omega} m_2(r)^{-1/\Omega}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv} H(s)\bar{L}_{t-1}(s)ds \\ &= \frac{u_t(r)^{1/\Omega} m_2(r)^{-1/\Omega}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv} \bar{L}. \end{aligned} \tag{7}$$

B. *Technology*

Firms produce a good $\omega \in [0, 1]$ using land and labor. A firm using $L_t^\omega(r)$ production workers per unit of land at location r at time t produces

$$q_t^\omega(r) = \phi_t^\omega(r)^{\gamma_1} z_t^\omega(r) L_t^\omega(r)^\mu$$

units of good ω per unit of land, where $\gamma_1, \mu \in (0, 1]$. A firm's productivity is determined by its decision on the quality of its technology—what we call an innovation $\phi_t^\omega(r)$—and an exogenous local and good-specific productivity shifter $z_t^\omega(r)$. To use an innovation $\phi_t^\omega(r)$, the firm has to employ $\nu\phi_t^\omega(r)^\xi$ additional units of labor per unit of land, where $\xi > \gamma_1/[1 - \mu]$. The exogenous productivity shifter $z_t^\omega(r)$ is the realization of a random variable that is i.i.d. across goods and time periods. It is drawn from a Fréchet distribution with cumulative distribution function

$$F(z, r) = e^{-T_t(r)z^{-\theta}},$$

where $T_t(r) = \tau_t(r)\bar{L}_t(r)^\alpha$, and $\alpha \geq 0$ and $\theta > 0$ are exogenously given. The value of $\tau_t(r)$ is determined by an endogenous dynamic process that depends on past innovation decisions in this location and potentially in others, $\phi_\cdot^\omega(\cdot)$.

We assume that the initial productivity $\tau_0(\cdot)$ is an exogenously given positive continuous function. Conditional on the spatial distribution of productivity in period $t - 1$, $\tau_{t-1}(\cdot)$, the productivity at location r in period t is given by

$$\tau_t(r) = \phi_{t-1}(r)^{\theta\gamma_1} \left[\int_S \eta\tau_{t-1}(s)ds\right]^{1-\gamma_2} \tau_{t-1}(r)^{\gamma_2}, \tag{8}$$

where η is a constant such that $\int_S \eta dr = 1$ and $\gamma_1, \gamma_2 \in [0, 1]$. If $\gamma_2 = 1$ and population density is constant over time, this implies that the mean of

$z_t^\omega(r)$ is $\phi_{t-1}(r)^{\gamma_1}$ times the mean of $z_{t-1}^\omega(r)$.[8] That is, the distribution of productivity draws is shifted up by past innovations, but with decreasing returns if $\gamma_1 < 1$. If $\gamma_2 < 1$, the dynamic evolution of a location's technology also depends on the aggregate level of technology, $\int_S \eta \tau_t(s)\,ds$.[9]

Later we will see that assuming $\gamma_1, \gamma_2 \in (0, 1)$ helps with the convergence properties of the model since we can have local decreasing returns but economywide linear technological progress. If $\gamma_2 = 1$, the evolution of local technology is independent of aggregate technology, and as we will show below, in a balanced-growth path in which the economy is growing, economic activity could end up concentrating in a unique point. In contrast, if $\gamma_1 = \gamma_2 = 0$, only the aggregate evolution of technology matters, there are no incentives to innovate, and the economy stagnates.

Across locations $z_t^\omega(r)$ is assumed to be spatially correlated. In particular, we assume that the productivity draws for a particular variety in a given period are perfectly correlated for neighboring locations as the distance between them goes to zero. We also assume that at large enough distances the draws are independent. This implies that the law of large numbers still applies in the sense that a particular productivity draw has no aggregate effects. Formally, let $\zeta_t^\omega(r, s)$ denote the correlation in the draws $z_t^\omega(r)$ and $z_t^\omega(s)$ and let $\delta(r, s)$ denote the distance between r and s. We assume that there exists a continuous function $s(d)$, where $\delta(r, s(d)) = d$ such that $\lim_{d \to 0} \zeta_t^\omega(r, s(d)) \to 1$. Furthermore, $\zeta_t^\omega(r, s) = 0$ for $\delta(r, s)$ large enough. One easy example is having land divided into regions of positive area, where $\zeta_t^\omega(r, s) = 1$ within a region and $\zeta_t^\omega(r, s) = 0$ otherwise.

[8] To obtain the mean of the standard Fréchet distribution $F(z) = e^{-T_t z^{-\theta}}$, first write down the density function $f(z) = \theta T_t z^{-\theta-1} e^{-T_t z^{-\theta}}$. The mean is then $\int_0^\infty z f(z)\,dz = \int_0^\infty \theta T_t z^{-\theta} e^{-T_t z^{-\theta}}\,dz$. Remember that $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\,dy$. Redefine $T_t z^{-\theta} = y$, so that $dy = -\theta T_t z^{-\theta-1}\,dz$. Substitute this into the previous expression, so that

$$\int_\infty^0 \frac{T_t \theta z^{-\theta}}{-\theta T_t z^{-\theta-1}} e^{-y}\,dy = \int_\infty^0 -z e^{-y}\,dy = T_t^{\frac{1}{\theta}} \int_0^\infty y^{-\frac{1}{\theta}} e^{-y}\,dy = T_t^{\frac{1}{\theta}} \Gamma\left(\frac{\theta - 1}{\theta}\right),$$

where $T_t = \tau_t L_t^\alpha$. If $\gamma_2 = 1$, we have $\tau_t = \phi_{t-1}^{\theta\gamma_1} \tau_{t-1}$. Assuming the labor force does not change over time, we can write

$$\frac{T_t}{T_{t-1}} = \frac{\tau_t}{\tau_{t-1}} = \phi_{t-1}^{\theta\gamma_1},$$

so that $T_t = \phi_{t-1}^{\theta\gamma_1} T_{t-1}$. Hence, the expectation is given by

$$E(z_t) = T_t^{\frac{1}{\theta}} \Gamma\left(\frac{\theta - 1}{\theta}\right) = \phi_{t-1}^{\gamma_1} T_{t-1}^{\frac{1}{\theta}} \Gamma\left(\frac{\theta - 1}{\theta}\right) = \phi_{t-1}^{\gamma_1} E(z_{t-1}).$$

[9] The assumption that η does not depend on r or s implies that any location benefits from any other location's technology, irrespectively of their distance. We need this assumption to characterize the balanced-growth path of the economy in Sec. III. However, we note that all the results in Sec. II, such as the existence and the uniqueness of the equilibrium, carry over to the case in which technology diffusion has a spatial scope. We examine the robustness of our numerical results to this case in Sec. D of app. A.

Since firm profits are linear in land, for any small interval with positive measure there is a continuum of firms that compete in prices (à la Bertrand). Note that the spatial correlation of the productivity draws, as well as the continuity of amenities and transport costs in space, implies that the factor prices and transport costs faced by these firms will be similar in a small interval. Hence, Bertrand competition implies that their pricing will be similar as well. As the size of the interval goes to zero, these price differences converge to zero, leading to an economy in which firms face perfect local competition.

Local competition implies that firms will bid for land up to the point at which they obtain zero profits after covering their investment in technology, $w_t(r)\nu\phi_t^\omega(r)^\xi$.[10] So even though this investment in technology affects productivity in the future through equation (8), the investment decision at any given point can safely disregard this dependence given the absence of future profits regardless of the level of investment. This implies that the solution to the dynamic innovation decision problem is identical to a sequence of static innovation decisions that maximize static profits. Firms innovate in order to maximize their bid for land, win the land auction, and produce. This decision affects the economy in the future, but not the future profits of the firm, which are always zero. The implication, as discussed in detail in Desmet and Rossi-Hansberg (2014), is that we need to solve only the static optimization problem of the firm and can disregard equation (8) in the firm's problem.

Therefore, after learning their common local productivity draw, $z_t^\omega(r)$, a potential firm at r maximizes its current profits per unit of land by choosing how much labor to employ and how much to innovate,

$$\max_{L_t^\omega(r),\phi_t^\omega(r)} p_t^\omega(r,r)\phi_t^\omega(r)^{\gamma_1} z_t^\omega(r) L_t^\omega(r)^\mu - w_t(r)L_t^\omega(r) - w_t(r)\nu\phi_t^\omega(r)^\xi - R_t(r),$$

where $p_t^\omega(r,r)$ is the price charged by the firm of a good sold at r, which is equivalent to the price the firm charges in another location net of transport costs. The two first-order conditions are

$$\mu p_t^\omega(r,r)\phi_t^\omega(r)^{\gamma_1} z_t^\omega(r) L_t^\omega(r)^{\mu-1} = w_t(r) \tag{9}$$

and

$$\gamma_1 p_t^\omega(r,r)\phi_t^\omega(r)^{\gamma_1-1} z_t^\omega(r) L_t^\omega(r)^\mu = \xi w_t(r)\nu\phi_t^\omega(r)^{\xi-1}. \tag{10}$$

So a firm's bid rent per unit of land is given by

[10] Because in any location there are many potential entrants with access to the same technology, the bidding is competitive. There is no need to be more specific about the auction.

$$R_t(r) = p_t^\omega(r, r)\phi_t^\omega(r)^{\gamma_1} z_t^\omega(r) L_t^\omega(r)^\mu - w_t(r)L_t^\omega(r) - w_t(r)\nu\phi_t^\omega(r)^\xi, \tag{11}$$

which ensures all firms make zero profits. Using (9) and (10) gives

$$\frac{L_t^\omega(r)}{\mu} = \frac{\xi\nu\phi_t^\omega(r)^\xi}{\gamma_1}. \tag{12}$$

Then total employment at r for variety ω, $\bar{L}_t^\omega(r)$, is the the sum of production workers, $L_t^\omega(r)$, and innovation workers, $\nu\phi_t^\omega(r)^\xi$, so

$$\bar{L}_t^\omega(r) = L_t^\omega(r) + \nu\phi_t^\omega(r)^\xi = \frac{L_t^\omega(r)}{\mu}\left[\mu + \frac{\gamma_1}{\xi}\right]. \tag{13}$$

Note also that

$$R_t(r) = \left[\frac{\xi(1-\mu)}{\gamma_1} - 1\right] w_t(r)\nu\phi_t^\omega(r)^\xi, \tag{14}$$

so bid rents are proportional and increasing in a firm's investment in technology, $w_t(r)\nu\phi_t^\omega(r)^\xi$, as we argued above.

In equilibrium firms take the bids for land by others, and therefore the equilibrium land rent, as given and produce in a location if their land bid is greater than or equal to the equilibrium land rent. Hence, in equilibrium, in a given location, the number of workers hired per unit of land and the amount of innovation done per unit of land are identical across goods. We state this formally in the following result.

Lemma 1. The decisions of how much to innovate, $\phi_t^\omega(r)$, and how many workers to hire per unit of land, $\bar{L}_t^\omega(r)$, are independent of the local idiosyncratic productivity draws, $z_t^\omega(r)$, and so are identical across goods ω.

Proof. Since in equilibrium $R_t(r)$ is taken as given by firms producing at r, the proof is immediate by inspecting (12), (13), and (14). QED

Lemma 1 greatly simplifies the analysis, as it will provide us with a relation between $p_t^\omega(r, r)$ and $z_t^\omega(r)$ similar to the one in Eaton and Kortum (2002) in spite of firms being able to innovate. Combining the equations above yields an expression for the price of a good produced at r and sold at r:

$$p_t^\omega(r, r) = \left[\frac{1}{\mu}\right]^\mu \left[\frac{\nu\xi}{\gamma_1}\right]^{1-\mu} \left[\frac{\gamma_1 R_t(r)}{w_t(r)\nu(\xi(1-\mu) - \gamma_1)}\right]^{(1-\mu)-(\gamma_1/\xi)} \frac{w_t(r)}{z_t^\omega(r)}. \tag{15}$$

To facilitate subsequent notation, we rewrite the above expression as

$$p_t^\omega(r, r) = \frac{mc_t(r)}{z_t^\omega(r)}, \tag{16}$$

where $mc_t(r)$ denotes the input cost in location r at time t, namely,

$$mc_t(r) \equiv \left[\frac{1}{\mu}\right]^{\mu}\left[\frac{\nu\xi}{\gamma_1}\right]^{1-\mu}\left[\frac{\gamma_1 R_t(r)}{w_t(r)\nu(\xi(1-\mu)-\gamma_1)}\right]^{(1-\mu)-(\gamma_1/\xi)} w_t(r). \tag{17}$$

It is key to understand that from the point of view of the individual firm, this input cost $mc_t(r)$ is given. As a result, expression (16) describes a straightforward relation between the productivity draw $z_t^{\omega}(r)$ and the price $p_t^{\omega}(r, r)$. As in Eaton and Kortum (2002), this is what allows us in the next subsection to derive probabilistic expressions of a location's price distribution, its probability of exporting to other locations, and its share of exports.

C. *Prices, Export Probabilities, and Export Shares*

Let $\varsigma(s, r) \geq 1$ denote the iceberg cost of transporting a good from r to s. Then, the price of a good ω, produced in r and sold in s, will be

$$p_t^{\omega}(s, r) = p_t^{\omega}(r, r)\varsigma(s, r) = \frac{mc_t(r)\varsigma(s, r)}{z_t^{\omega}(r)}. \tag{18}$$

We impose the following assumption on transport costs.

ASSUMPTION 2. $\varsigma(\cdot, \cdot) : S \times S \rightarrow \mathbb{R}$ is symmetric.

As we derive formally in appendix B, the probability density that a given good produced in an area r is bought in s is given by

$$\pi_t(s, r) = \frac{T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta}}{\int_S T_t(u)[mc_t(u)\varsigma(u, s)]^{-\theta} du} \quad \text{all } r, s \in S. \tag{19}$$

The price index of a location s, as we also show in appendix B, is then given by

$$P_t(s) = \Gamma\left(\frac{-\rho}{(1-\rho)\theta} + 1\right)^{-(1-\rho)/\rho}\left\{\int_S T_t(u)[mc_t(u)\varsigma(s, u)]^{-\theta} du\right\}^{-1/\theta}. \tag{20}$$

D. *Trade Balance*

We impose trade balance location by location since there is no mechanism for borrowing from or lending to other agents. Market clearing requires total revenue in location r to be equal to total expenditure on goods from r. Total revenue at r is

$$w_t(r)H(r)\left[L_t(r) + \nu\phi_t(r)^{\xi}\right] + H(r)R_t(r) = \frac{1}{\mu} w_t(r)H(r)L_t(r),$$

where the last equality comes from (12), (13), and (14). As in Eaton and Kortum (2002), the fraction of goods that location s buys from r, $\pi_t(s, r)$, is equal to the fraction of expenditure on goods from r, so that the trade balance condition can be written as

$$w_t(r)H(r)\bar{L}_t(r) = \int_S \pi_t(s, r)w_t(s)H(s)\bar{L}_t(s)ds \quad \text{all } r \in S, \tag{21}$$

where the superscript ω can be dropped because the number of workers does not depend on the good a firm produces, and L can be replaced by $\bar{L}$ because the proportion of total workers to production workers is constant across locations.

E. Equilibrium

We define a dynamic competitive equilibrium as follows.

Definition 1. Given a set of locations, S, and their initial technology, amenity, population, and land functions $(\tau_0, \bar{a}, \bar{L}_0, H) : S \to \mathbb{R}_{++}$, as well as their bilateral trade and migration cost functions $\varsigma, m : S \times S \to \mathbb{R}_{++}$, a competitive equilibrium is a set of functions $(u_t, \bar{L}_t, \phi_t, R_t, w_t, P_t, \tau_t, T_t) : S \to \mathbb{R}_{++}$ for all $t = 1, \ldots$, as well as a pair of functions $(p_t^i, c_t^i) : [0, 1] \times S \to \mathbb{R}_{++}$ for all $t = 1, \ldots$, such that for all $t = 1, \ldots$:

1. Firms optimize and markets clear. Namely, (9), (10), and (13) hold at all locations.
2. The share of income of location s spent on goods of location r is given by (17) and (19) for all $r, s \in S$.
3. Trade balance implies that (21) holds for all $r \in S$.
4. Land markets are in equilibrium, so land is assigned to the highest bidder. Thus, for all $r \in S$,

$$R_t(r) = \left[\frac{\xi - \mu\xi - \gamma_1}{\mu\xi + \gamma_1}\right] w_t(r)\bar{L}_t(r).$$

5. Given migration costs and their idiosyncratic preferences, people choose where to live, so (7) holds for all $r \in S$.
6. The utility associated with real income and amenities in location r is given by

$$\begin{aligned} u_t(r) &= a_t(r)\frac{w_t(r) + R_t(r)/\bar{L}_t(r)}{P_t(r)} \\ &= \bar{a}(r)\bar{L}_t(r)^{-\lambda}\frac{\xi}{\mu\xi + \gamma_1}\frac{w_t(r)}{P_t(r)} \quad \text{for all } r \in S, \end{aligned} \tag{22}$$

where the price index $P_t(\cdot)$ is given by (20).

7. Labor markets clear, so

$$\int_S H(r)\bar{L}_t(r)\,dr = \bar{L}.$$

8. Technology evolves according to (8) for all $r \in S$.

In what follows we prove results under the following assumption.

ASSUMPTION 3. $\bar{a}(\cdot)$, $H(\cdot)$, $\tau_0(\cdot)$, $\bar{L}_0(\cdot) : S \to \mathbb{R}_{++}$, and $m(\cdot,\cdot)$, $\varsigma(\cdot,\cdot) : S \times S \to \mathbb{R}_{++}$ are continuous functions.

Assumption 3 implies that there is no discontinuity in the underlying functions determining the distribution of economic activity in space. Since we can make these functions as steep as we want at borders or other natural geographic barriers, this assumption comes at essentially no loss of generality. We prove all the results below under this assumption. Of course, for the quantification and calibration of the model we will use a discrete approximation. Similar results for existence and uniqueness, involving the exact same parameter restrictions we impose below, can be established directly for the discrete case by adapting some of the arguments in Allen and Arkolakis (2014).

We can manipulate the system of equations that defines an equilibrium and, ultimately, reduce it to a system of equations that determines wages, employment levels, and utility, $u_t(\cdot)$, in all locations. In a given period t, the following lemma characterizes the relationship between wages, utility, and labor density, conditional on $\bar{a}(\cdot)$, $\tau_t(\cdot)$, $\bar{L}_{t-1}(\cdot)$, $\varsigma(\cdot,\cdot)$, $m(\cdot,\cdot)$, $H(\cdot)$, and parameter values. Appendix B presents all proofs not included in the main text.

LEMMA 2. For any t and for all $r \in S$, given $\bar{a}(\cdot)$, $\tau_t(\cdot)$, $\bar{L}_{t-1}(\cdot)$, $\varsigma(\cdot,\cdot)$, $m(\cdot,\cdot)$, and $H(\cdot)$, the equilibrium wage, $w_t(\cdot)$, population density, $\bar{L}_t(\cdot)$, and utility, $u_t(\cdot)$ schedules satisfy equations (7), as well as

$$w_t(r) = \bar{w}\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\frac{\theta}{1+2\theta}} \tau_t(r)^{\frac{1}{1+2\theta}} H(r)^{-\frac{1}{1+2\theta}} \bar{L}_t(r)^{\frac{\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta}{1+2\theta}} \tag{23}$$

and

$$\begin{aligned} &\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} \tau_t(r)^{-\frac{\theta}{1+2\theta}} H(r)^{\frac{\theta}{1+2\theta}} \\ &\quad \times \bar{L}_t(r)^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]} \\ &= \kappa_1 \int_S \left[\frac{\bar{a}(s)}{u_t(s)}\right]^{\frac{\theta^2}{1+2\theta}} \tau_t(s)^{\frac{1+\theta}{1+2\theta}} H(s)^{\frac{\theta}{1+2\theta}} \varsigma(r,s)^{-\theta} \\ &\quad \times \bar{L}_t(s)^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]} ds, \end{aligned} \tag{24}$$

where κ_1 is a constant.

We now establish conditions to guarantee that the solution to the system of equations (7), (23), and (24) exists and is unique. We can prove that there exists a unique solution $w_t(\cdot)$, $\bar{L}_t(\cdot)$, and $u_t(\cdot)$ that satisfies (7), (23), and (24) if

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} \leq \lambda + 1 - \mu + \Omega.$$

This condition is very intuitive. It states that the static agglomeration economies associated with the local production externalities (α/θ) and the degree of returns to innovation (γ_1/ξ) do not dominate the three congestion forces. These three forces are governed by the value of the negative elasticity of amenities to density (λ), the share of land in production, and therefore the decreasing returns to local labor ($1 - \mu$), and the variance of taste shocks (Ω). Of course, the condition is stated as a function of exogenous parameters only. We summarize this result in the following lemma.

LEMMA 3. A solution $w_t(\cdot)$, $\bar{L}_t(\cdot)$, and $u_t(\cdot)$ that satisfies (7), (23), and (24) exists and is unique if $\alpha/\theta + \gamma_1/\xi < \lambda + 1 - \mu + \Omega$. Furthermore, the solution can be found with an iterative procedure.

The two lemmas above imply that we can uniquely solve for $w_t(\cdot)$, $\bar{L}_t(\cdot)$, and $u_t(\cdot)$ given the allocation in the previous period. For $t = 0$, using the initial conditions $\tau_0(\cdot)$ and $\bar{L}_0(\cdot)$, we can easily calculate all other equilibrium variables using the formulas described in the definition of the equilibrium. We can then calculate next period's productivity $\tau_1(\cdot)$ at all locations using equation (8). Applying the algorithm in lemma 3 for every time period then determines a unique equilibrium allocation over time. The following proposition summarizes this result.

PROPOSITION 1. An equilibrium of this economy exists and is unique if $\alpha/\theta + \gamma_1/\xi \leq \lambda + 1 - \mu + \Omega$.

III. The Balanced-Growth Path

In a balanced-growth path (BGP) of the economy, if one exists, all regions grow at the same rate. A BGP might not exist; instead, all economic activity might eventually concentrate in one point, or the economy may cycle without reaching a BGP. Given the evolution of technology in (8), the growth rate of $\tau_t(r)$ is given by

$$\frac{\tau_{t+1}(r)}{\tau_t(r)} = \phi_t(r)^{\theta\gamma_1}\left[\int_S \eta \frac{\tau_t(s)}{\tau_t(r)}\, ds\right]^{1-\gamma_2}.$$

Hence, in a BGP in which technology growth rates are constant, so $\tau_{t+1}(r)/\tau_t(r)$ is constant over time and space and $\tau_t(s)/\tau_t(r)$ is constant over time, the investment decision will be constant but different across

locations. Divide both sides of the equation by the corresponding equation for location s, and rearrange to get

$$\frac{\tau_t(s)}{\tau_t(r)} = \left[\frac{\phi(s)}{\phi(r)}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[\frac{\bar{L}(s)}{\bar{L}(r)}\right]^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}},$$

where the second equality follows from (12) and where we drop the time subscript to indicate that we refer to a variable that remains constant in the BGP. We can then use (7), (24), and the labor market clearing condition to derive an equation that determines the spatial distribution of $u_t(r)$ on the BGP. According to theorem 2.19 in Zabreyko et al. (1975), a unique positive solution to that equation exists if

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} + \frac{\gamma_1}{[1-\gamma_2]\xi} \leq \lambda + 1 - \mu + \Omega. \tag{25}$$

This condition is strictly more restrictive than the condition that guarantees the existence and uniqueness of an equilibrium in lemma 1 since it includes an extra positive term on the left-hand side. It is also intuitive. On the left-hand side we have the two static agglomeration effects: agglomeration externalities (α/θ) and improvements in local technology for today's production (γ_1/ξ). The third term, which appears in the condition for the BGP only, is related to the dynamic agglomeration effect from local investments in technology as well as diffusion ($\gamma_1/([1-\gamma_2]\xi)$). In fact, without diffusion, when $1-\gamma_2 = 0$, condition (25) is never satisfied and there is no BGP with a nondegenerate distribution of employment. On the right-hand side of condition (25) we have the parameters governing the three dispersion forces, namely, congestion through lower amenities (λ), congestion through lower land per worker ($1-\mu$), and dispersion because of taste shocks (Ω). So condition (25) simply says that in order for the economy to have a unique BGP, the dispersion forces have to be large enough relative to all agglomeration forces. Similarly, the condition in lemma 1 says that dispersion forces have to be strong enough relative to *static* agglomeration forces in order for an equilibrium to exist. The difference is that an equilibrium can exist even if condition (25) is violated since the *dynamic* agglomeration effect might lead economic activity to progressively concentrate in an area of measure zero.

We summarize the result in the following lemma.

LEMMA 4. If $\alpha/\theta + \gamma_1/\xi + \gamma_1/[[1-\gamma_2]\xi] \leq \lambda + 1 - \mu + \Omega$, then there exists a unique balanced-growth path with a constant distribution of employment densities $\bar{L}(\cdot)$ and innovation $\phi(\cdot)$. In the BGP $\tau_t(r)$ grows at a constant rate for all $r \in S$.

The condition that determines $\tau_t(r)$ in the BGP (which we write explicitly in the proof of lemma 4) guarantees that in the BGP welfare grows uniformly everywhere at the rate

$$\frac{u_{t+1}(r)}{u_t(r)} = \left[\frac{\tau_{t+1}(r)}{\tau_t(r)}\right]^{1/\theta}.$$

We can then use the equations above to show that the growth rate of world utility (or the growth rate of real output) is a function of the distribution of employment in the BGP.

LEMMA 5. In a balanced-growth path, under the conditions of lemma 4, aggregate welfare and aggregate real consumption grow according to

$$\begin{aligned}\frac{u_{t+1}(r)}{u_t(r)} &= \left[\frac{\int_0^1 c_{t+1}^{\omega}(r)^{\rho}\,d\omega}{\int_0^1 c_t^{\omega}(r)^{\rho}\,d\omega}\right]^{\frac{1}{\rho}} \\ &= \eta^{\frac{1-\gamma_2}{\theta}}\left[\frac{\gamma_1/\nu}{\gamma_1+\mu\xi}\right]^{\frac{\gamma_1}{\xi}}\left[\int_S \bar{L}(s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}}ds\right]^{\frac{1-\gamma_2}{\theta}}.\end{aligned} \tag{26}$$

Hence welfare and aggregate real output growth depend on population size and its distribution in space.

In a world with aggregate population growth the above result would result in strong scale effects in the BGP: growth of aggregate consumption would be an increasing function of world population in the BGP. There is some debate about whether such strong scale effects are consistent with the empirical evidence. In particular, Jones (1995) observes that over the course of the twentieth century there has been no acceleration in the growth of income per capita in the United States in spite of an important increase in its population. Given that in our model the world economy is not in the BGP and that population is constant, this issue is not of direct concern. However, if we were to allow for the world population to grow over time, it would be straightforward to eliminate strong scale effects in the BGP by making the cost of innovation an increasing function of the size of world population.[11]

IV. Calibration and Simulation of the Model

In order to compute the equilibrium of the model we need values for the 12 parameters used in the equations above, in addition to values for initial productivity levels and amenities for all locations, as well as bilateral

[11] In particular, assume that to introduce an innovation $\phi_t^{\omega}(r)$, a firm needs to employ $\tilde{\nu}\phi_t^{\omega}(r)^{\xi}\bar{L}$ units of labor per unit of land, where $\bar{L}$ is total world population. This alternative model is isomorphic to the benchmark model with $\nu = \tilde{\nu}\bar{L}$, and it implies a BGP growth rate of aggregate welfare and aggregate consumption that is no longer a function of the world's total population. In the alternative model, expression (26) would become a function of $\bar{L}(s)/\bar{L}$ rather than of $\bar{L}(s)$.

migration costs and transport costs between any two locations. Once we have numbers for all of these variables and parameters, we can compute the model with the simple iterative algorithm described in the proof of lemma 3.

Table 1 lists the parameter values and gives a brief explanation of how they are assigned. When assigning parameter values, we assume a model period to be 1 year, so we set $\beta = 0.965$.[12] We base some of the parameter values on those in the existing literature. We estimate other parameter values using our model. In what follows we start by briefly discussing some of the parameter values that come from the literature and then provide a detailed discussion of how we estimate the remaining parameters.

The elasticity of substitution, $1/(1 - \rho)$, is set to 4, similar to the 3.8 estimated in Bernard et al. (2003). We choose a trade elasticity, θ, equal to 6.5, somewhere in the middle between the 8.3 value estimated by Eaton and Kortum (2002) and the 4.6 value estimated by Simonovska and Waugh (2014). The labor share in production, μ, is set to 0.8. While higher than the standard labor share, this parameter should be interpreted as the nonland share. Desmet and Rappaport (2017) find a land share of 0.1 when accounting for the land used in both production and housing. Taking a broader view of land by including structures, this share increases to around 0.2, on the basis of a structures share slightly above 0.1, as calibrated by Greenwood, Hercowitz, and Krusell (1997). We therefore take the nonland share to be 0.8 but have checked that our main results are robust to alternative values of this parameter.

Equation (6) implies that Ω is the inverse of the elasticity of migration flows with respect to real income. In our specification, that elasticity is independent of the migration costs as those are captured by our estimate of m_2. We therefore focus on elasticity measures estimated in contexts in which there are no formal migration restrictions. On the basis of the study by Ortega and Peri (2013), who look at intra-EU migration, as well as Diamond (2016), Fajgelbaum et al. (2016), and Monte, Redding, and Rossi-Hansberg (2018), who consider intra-US migration, a reasonable value for that elasticity is 2, so we set $\Omega = 0.5$. In Section D of appendix A, we explore how our results depend on the particular value of Ω by recomputing the model for a higher value.

[12] We need to make individuals discount future consumption at a rate that is higher than the growth rate in the balanced-growth path. In our calibration, as well as in the different counterfactual scenarios with different migration costs, the growth rate of real consumption is never above 3.5 percent, so setting $\beta = 0.965$ results in well-defined present discounted values in all our exercises.

TABLE 1
PARAMETER VALUES

Parameter	Target/Comment
1. Preferences: $\Sigma_t \beta^t u_t(r)$, where $u_t(r) = \bar{a}(r)\overline{L_t}(r)^{-\lambda}(r)[\int_0^1 c_t^\omega(r)^\rho d\omega]^{1/\rho}$ and $u_0(r) = e^{\psi \bar{u}(r)}$	
$\beta = .965$	Discount factor
$\rho = .75$	Elasticity of substitution of 4 (Bernard et al. 2003)
$\lambda = .32$	Relation between amenities and population
$\Omega = .5$	Elasticity of migration flows with respect to income (Monte et al. 2018)
$\psi = 1.8$	Deaton and Stone (2013)
2. Technology: $q_t^\omega(r) = \phi_t^\omega(r)^{\gamma_1} z_t^\omega(r) L_t^\omega(r)^\mu$, $F(z, r) = e^{-T_t^\omega(r) z^{-\theta}}$, and $T_t^\omega(r) = \tau_t(r)\overline{L_t}(r)^\alpha$	
$\alpha = .06$	Static elasticity of productivity to density (Carlino et al. 2007)
$\theta = 6.5$	Trade elasticity (Eaton and Kortum 2002; Simonovska and Waugh 2014)
$\mu = .8$	Labor or nonland share in production (Greenwood et al. 1997; Desmet and Rappaport 2017)
$\gamma_1 = .319$	Relation between population distribution and growth
3. Evolution of productivity: $\tau_t(r) = \phi_{t-1}(r)^{\theta\gamma_1}[\int_S \eta \tau_{t-1}(s) ds]^{1-\gamma_2} \tau_{t-1}(r)^{\gamma_2}$ and $\psi(\phi) = \nu\phi^\xi$	
$\gamma_2 = .993$	Relation between population distribution and growth
$\xi = 125$	Desmet and Rossi-Hansberg (2015)
$\nu = .15$	Initial world growth rate of real GDP of 2%
4. Trade Costs	
$\varsigma_{\text{rail}} = .1434$	
$\varsigma_{\text{no_rail}} = .4302$	
$\varsigma_{\text{major_road}} = .5636$	
$\varsigma_{\text{other_road}} = 1.1272$	Allen and Arkolakis (2014)
$\varsigma_{\text{no_road}} = 1.9726$	
$\varsigma_{\text{water}} = .0779$	
$\varsigma_{\text{no_water}} = .0779$	
$\Upsilon = .393$	Elasticity of trade flows with respect to distance of −.93 (Head and Mayer 2014)

A. *Amenity Parameter*

The theory assumes that a location's amenities decrease with its population. As given by $a(r) = \bar{a}(r)\bar{L}(r)^{-\lambda}$, the parameter λ represents the elasticity of amenities to population. Taking logs gives us the following equation:

$$\log(a(r)) = E(\log(\bar{a}(r))) - \lambda \log \bar{L}(r) + \varepsilon_a(r), \tag{27}$$

where $E(\log(\bar{a}(r)))$ is the mean of $\log(\bar{a}(r))$ across locations, and $\varepsilon_a(r)$ is the location-specific deviation of $\log(\bar{a}(r))$ from the mean. Assuming that amenities are lognormally distributed across locations, we use data from Desmet and Rossi-Hansberg (2013) on amenities and population for 192 metropolitan statistical areas (MSAs) in the United States to estimate equation (27). One remaining issue is that not only does population affect amenities, but amenities also affect population. To deal with this problem of reverse causality, we use an MSA's exogenous productivity level as an instrument for its population. Desmet and Rossi-Hansberg provide estimates for the exogenous productivity of MSAs, which they define as the productivity that is not due to agglomeration economies. When using this as an instrument, the identifying assumption is that a location's exogenous productivity does not affect its amenities directly, but only indirectly through the level of its population. This is consistent with the assumptions of our model. Estimating (27) by two-stage least squares yields a value of $\lambda = 0.32$, which is what we report in table 1.

B. *Technology Parameters*

Our starting point is the economy's utility growth equation in the balanced-growth path (26). To exploit the cross-country variation in growth rates in the data, assume that all countries are in a balanced-growth path, but their growth rates may differ.[13] After taking logs and discretizing space into cells, we can rewrite (26) for country c as

$$\begin{aligned} \log u_{t+1}(c) - \log u_t(c) &= \log y_{t+1}(c) - \log y_t(c) \\ &= \alpha_1 + \alpha_2 \log \sum_{S_c} L_c(s)^{\alpha_3}, \end{aligned} \tag{28}$$

[13] Essentially, we are assuming that the relative distribution of population within countries has converged to what would be observed in a balanced-growth path, although international migration flows may still change the relative distribution of population across countries. As a result, growth rates may differ across countries, although each country is characterized by (26). As an alternative to this simplifying assumption, we could use the whole structure of the model calibrated for 1990 and then estimate the parameters that make the simulated model match the observed growth rates. This calculation requires enormous amounts of computational power and so is left for future research. The fact that adding population growth to the estimating equation does not substantially change the results alleviates this concern somewhat.

where α_1 is a constant and

$$\alpha_2 = \frac{[1 - \gamma_2]}{\theta},$$

$$\alpha_3 = \frac{\theta\gamma_1}{[1 - \gamma_2]\xi},$$

and country-level per capita growth is such that, in a steady state,

$$\log y_{t+1}(c) - \log y_t(c) = \log y_{t+1}(r) - \log y_t(r)$$

for all $r \in c$.[14] The theory therefore predicts that steady-state growth is a function of the following measure of the spatial distribution of population:

$$\sum_s L(s)^{\alpha_3}. \tag{29}$$

Assuming $\alpha_2 > 0$, then if $0 < \alpha_3 < 1$, steady-state growth is maximized when labor is equally spread across space; and if $\alpha_3 > 1$, steady-state growth is maximized when labor is concentrated in one cell. Before estimating (28), we normalize (29) in order to eliminate the effect of the number of cells differing across countries:

$$\frac{1}{NS}\sum_s L(s)^{\alpha_3}, \tag{30}$$

where NS is the number of cells in a country. To see what this normalization does, consider two examples. Country A has four cells: two have population levels L_1 and two have population levels L_2. Country B is identical to country A but is quadruple its size: it has 16 cells, of which eight have population levels L_1 and eight have population levels L_2. The above normalization (30) makes the population distribution measures of countries A and B identical.

To get empirical estimates for α_1, α_2, and α_3, we use cell population data from G-Econ 4.0 to construct a measure of (30) for four years: 1990, 1995, 2000, and 2005. We focus on countries with at least 20 cells, and for the data on real GDP per capita, we aggregate cell GDP and cell population from the G-Econ data set to compute a measure of real GDP per capita. This gives us 106 countries and three time periods.

[14] Equation (28) is consistent with there being no differences in population growth across countries in a balanced-growth path. If we were to allow for such differences, the first part of expression (28) should be written as

$$\log u_{t+1}(c) - \log u_t(c) = \log y_{t+1}(c) - \log y_t(c) - \lambda[\log L_{t+1}(c) - \log L_t(c)],$$

where $L_t(c) \equiv \Sigma_S L_t(s)$. As we will discuss later, this leads to very similar parameter values.

When estimating (28), we use the between-estimator; that is, we use the mean of the different variables. We do so because the dependent variable (growth) is rather volatile, whereas the independent variable of interest (the spatial distribution of population) is rather persistent. This suggests that most of the variation should come from differences between countries rather than from differences within countries. Moreover, (28) is a steady-state relation, so focusing on the average 5-year growth rates seems sensible. Our estimation gives values of $\alpha_2 = 0.00116$ and $\alpha_3 = 2.2$. Using the expressions for α_2 and α_3 following (28), this yields $\gamma_1 = 0.319$ and $\gamma_2 = 0.993$, which are the values reported in table 1. Section D of appendix A presents some sensitivity analysis of our results to the strength of technological diffusion, given by $1 - \gamma_2$.

If we were to allow for population growth, this would not affect the estimates as long as in the balanced-growth path population growth is the same across countries. If, however, population growth rates do differ across countries, equation (28) would become

$$\log y_{t+1}(c) - \log y_t(c) = \alpha_1 + \alpha_2 \log \sum_{S_c} L_c(s)^{\alpha_3} + \lambda[\log L_{t+1}(c) - \log L_t(c)].$$

Reestimating this equation and imposing a value of $\lambda = 0.32$, as estimated before, yields very similar results: $\alpha_2 = 0.00103$ and $\alpha_3 = 2.6$. This leaves the values of γ_1 and γ_2 virtually unchanged at, respectively, 0.335 and 0.993.

Finally, we choose the level of innovation costs ν so as to match a growth rate of real GDP of 2 percent in the initial period. This yields a value of $\nu = 0.15$. We also need to choose a value for the static agglomeration effect governed by α. Our model includes both static and dynamic agglomeration effects, so we choose a relatively low value of $\alpha = 0.06$, which corresponds to a static agglomeration effect of $\alpha/\theta = 0.01$. This value is similar to, although a bit smaller than, the one estimated in Carlino, Chatterjee, and Hunt (2007) or Combes et al. (2012), since our model also features a dynamic agglomeration effect. In Section D of appendix A, we present a robustness test with a 20 percent larger value of α.

C. *Trade Costs*

We discretize the world into $1° \times 1°$ grid cells, which means $180 \times 360 = 64{,}800$ grid cells in total. A location thus corresponds to a grid cell. To ship a good from location r to s, one has to follow a continuous and once-differentiable path $g(r, s)$ over the surface of the earth that connects the two locations. Passing through a location is costly. We assume that the cost of passing through location r is given (in logs) by

$$\begin{aligned}\log \varsigma(r) = {} & \log \varsigma_{\text{rail}}\text{rail}(r) + \log \varsigma_{\text{no_rail}}[1 - \text{rail}(r)] \\ & + \log \varsigma_{\text{major_road}}\text{major_road}(r) + \log \varsigma_{\text{other_road}}\text{other_road}(r) \\ & + \log \varsigma_{\text{no_road}}[1 - \text{major_road}(r) - \text{other_road}(r)] \\ & + \log \varsigma_{\text{water}}\text{water}(r) + \log \varsigma_{\text{no_water}}[1 - \text{water}(r)],\end{aligned}$$

where rail(r) equals one if there is a railroad passing through r and zero otherwise, major_road(r) equals one if there is a major road passing through r and zero otherwise, other_road(r) equals one if there is some other road (but no major road) passing through r and zero otherwise, and water(r) equals one if there is a major water route at r and zero otherwise. The coefficients ς_{rail}, $\varsigma_{\text{no_rail}}$, $\varsigma_{\text{major_road}}$, $\varsigma_{\text{other_road}}$, $\varsigma_{\text{no_road}}$, ς_{water}, and $\varsigma_{\text{no_water}}$ are positive constants, and their values are based on values in Allen and Arkolakis (2014).

We observe data on water, rail, and road networks at a finer spatial scale than the $1° \times 1°$ level. In particular, using data from http://www.naturalearthdata.com/, we can see whether there is a railroad, major road, and so forth passing through any cell of size $0.1° \times 0.1°$. We aggregate these data up to the $1° \times 1°$ grid cell level such that, for instance, rail(r) now corresponds to the fraction of smaller cells within cell r that have access to the rail network. We do the same aggregation for the road and water variables.[15]

Having $\varsigma(r)$, we use the Fast Marching Algorithm[16] to compute the lowest cost between any two cells $r \neq s$,

$$\varsigma(r, s) = \left[\inf_{g(r,s)} \int_{g(r,s)} \varsigma(u)\, du\right]^{\Upsilon},$$

where $\int_{g(r,s)} \varsigma(u)\, du$ denotes the line integral of $\varsigma(\cdot)$ along the path $g(r, s)$.[17]

[15] Clearly, building roads and rail is endogenous to local development. We abstract from this aspect but note that major roads and rail lines are in general constructed through geographically convenient locations, a feature of space that is, in fact, exogenous. In any case, the most important determinant of how costly it is to pass through a location is the presence of water.

[16] We apply Gabriel Peyre's Fast Marching Toolbox for Matlab to search for the lowest cost, taking into account that the earth is a sphere. In particular, we adjust the values of $\varsigma(r)$ on the basis of the distance required for crossing a cell, which varies with the position of the cell on the earth's surface. We perform the Fast Marching Algorithm and calculate these distances over a very fine triangular approximation of the surface.

[17] We choose the trade cost of the cell with itself, $\varsigma(r, r)$, to equal the average cost between points inside the cell,

$$\left[E\left(\int_{g(r_1,r_2)} \varsigma(r)\, du\right) | r_1, r_2 \in r\right]^{\Upsilon}.$$

Note that these assumptions do not guarantee that all bilateral trade costs are above one. However, least-cost paths across cells and average least-cost paths within cells are long enough such that this is not a concern in the numerical implementation.

We calibrate Υ to match the elasticity of bilateral trade flows across cells to distance in the data. In a meta-analysis of the empirical gravity literature, Head and Mayer (2014) find a mean value of this elasticity equal to -0.93. We run a standard gravity regression on trade data simulated by the model for the initial period and search for the value of Υ that matches this elasticity. This procedure identifies Υ uniquely since higher values of Υ must correspond to higher absolute values of the elasticity. It yields $\Upsilon = 0.393$, implying that, conditional on the mode of transportation, trade costs are concave in distance traveled. We also find that the resulting distance elasticity of within-country trade flows equals -1.32, which is very close to -1.29, the distance elasticity of trade flows across counties within the United States estimated by Monte et al. (2018).

Though we did not match it in the calibration, the simulation of the world economy we present in the next section yields a ratio of trade to GDP in the world that is identical to the one observed in the data. We perform robustness tests with respect to trade costs in Section D of appendix A.

D. *Local Amenities and Initial Productivity*

To simulate the model, we also need to know the spatial distribution of $\bar{a}(r)$ and $\tau_0(r)$. We use data at the grid cell level on land $H(r)$ from the National Oceanic and Atmospheric Administration, as well as population $\bar{L}_0(r)$ and wages $w_0(r)$ as measured by GDP per capita in 2000 (period 0) from G-Econ 4.0, to recover these distributions. Using equation (23) for $t = 0$,

$$w_0(r) = \bar{w}\left[\frac{\bar{a}(r)}{u_0(r)}\right]^{-\frac{\theta}{1+2\theta}} \tau_0(r)^{\frac{1}{1+2\theta}} H(r)^{-\frac{1}{1+2\theta}} \bar{L}_0(r)^{\frac{\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta}{1+2\theta}},$$

we obtain

$$\tau_0(r) = \bar{w}^{-(1+2\theta)}\left[\frac{\bar{a}(r)}{u_0(r)}\right]^{\theta} H(r) w_0(r)^{1+2\theta} \bar{L}_0(r)^{1-\alpha-\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta} \tag{31}$$

for any r. Plugging this into equation (24), we get

$$\begin{aligned} &w_0(r)^{-\theta}\bar{L}_0(r)^{\lambda\theta}\left[\frac{\bar{a}(r)}{u_0(r)}\right]^{-\theta} \\ &= \kappa_1 \bar{w}^{-(1+2\theta)} \int_S w_0(s)^{1+\theta}\bar{L}_0(s)^{1-\lambda\theta} H(s)\varsigma(r,s)^{-\theta}\left[\frac{\bar{a}(s)}{u_0(s)}\right]^{\theta} ds. \end{aligned} \tag{32}$$

Given $H(r)$, $\bar{L}_0(r)$, and $w_0(r)$, we solve equation (32) for $\bar{a}(r)/u_0(r)$. We can then use equation (31) to obtain $\tau_0(r)$. The following lemma shows

that the values of $\bar{a}(r)/u_0(r)$ and $\tau_0(r)$ that satisfy these equations are unique.

LEMMA 6. Given $\bar{w}$, the solution to equations (31) and (32) exists and is unique.

Proof. The existence and uniqueness of a solution to (31) directly follow from the existence and uniqueness of a solution to (32). To prove existence and uniqueness for (32), see theorem 2.19 in Zabreyko et al. (1975). QED

Lemma 6 guarantees the existence and uniqueness of the inversion of the model used to obtain $\bar{a}(r)/u_0(r)$ and $\tau_0(r)$. However, it does not guarantee that we can find a solution using an iterative procedure. In Section G of appendix B, we discuss the numerical algorithm we use to find a solution. Also, to solve for $\bar{a}(r)/u_0(r)$ and $\tau_0(r)$, we normalize $\bar{w}$ to the average wage in the world in 2000.

The system above identifies $\bar{a}(r)/u_0(r)$ but is unable to tell $\bar{a}(r)$ apart from $u_0(r)$. To disentangle a location's amenity from its initial utility, we need to obtain an estimate of $u_0(r)$. To do so we use data on subjective well-being from the Gallup World Poll. Subjective well-being is measured on a Cantril ladder from 0 to 10, where 0 represents the worst possible life and 10 the best possible life the individual can contemplate for herself. This measure is, of course, ordinal, not cardinal. Furthermore, it requires the individual to set her own comparison benchmark when determining what the best possible life, or the worst possible life, might mean. This benchmark might vary across individuals, regions, and countries. However, given that Deaton and Stone (2013) and Stevenson and Wolfers (2013) find a relationship between subjective well-being and the log of real income that is similar within the United States and across countries, we abstract from these potential differences in welfare benchmarks across the world. But we still need to transform subjective well-being into a cardinal measure of the level of well-being.

Ignoring migration costs, recall that in the model the flow utility of an individual i residing in location r is linear in her real income, namely,

$$u^i(r) = a(r)y(r)\varepsilon^i(r), \tag{33}$$

where real income is $y(r) = [w(r) + R(r)/\bar{L}(r)]/P(r)$. Since we are focusing on a given time period, we have dropped the time subscript in the previous expressions. Then, to make the "ladder" data from the Gallup World Poll comparable to the utility measure in the model, we need to transform subjective well-being into a measure that is linear in income.

Deaton and Stone (2013) find that the ladder measure of the subjective well-being of an individual i residing in location r is linearly related to the log of her real income (see also Kahneman and Deaton 2010). In particular, they estimate a relation

$$\check{u}^i(r) = \rho \ln y^i(r) + v(r) + \varepsilon^i_{DS}(r), \tag{34}$$

where the inverted hat refers to subjective well-being, as measured by the Cantril ladder, $v(r)$ is a location fixed effect, and $\varepsilon^i_{DS}(r)$ is a random variable with mean zero. Whereas the ladder measure is linear in the log of real income, our utility measure in (33) is linear in the level of real income. To make (34) consistent with our model, we can rewrite (33) as[18]

$$\rho \ln u^i(r) = \rho \ln y^i(r) + \rho \ln a(r) + \rho \ln \varepsilon^i(r). \tag{35}$$

Equations (34) and (35) imply the following relation between utility as defined in our model, $u^i(r)$, and utility as defined by subjective well-being, $\check{u}^i(r)$,

$$u^i(r) = e^{\psi \check{u}^i(r)}, \tag{36}$$

where $\psi = 1/\rho$. Given the structure of our model, one potential issue with estimating (34) if we were to use only cross-country or cross-regional data is endogeneity: a location with a higher utility attracts more people and therefore affects the amenity levels through $a(r) = \bar{a}(r)\bar{L}(r)^{-\lambda}$. However, if (34) is estimated for a given time period using the cross-sectional variation and including location fixed effects, this is less of a concern. Using individual-level data, Deaton and Stone (2013) estimate ρ to be around 0.55, which implies a value of ψ of 1.8. Our data on subjective well-being are at the country level, so we set $u_0(r) = e^{1.8\check{u}(c(r))}$, where $\check{u}(c(r))$ is the subjective well-being measure of the country c to which location r belongs.

Since the inversion has yielded estimates for $\bar{a}(r)/u_0(r)$, we can then use the estimates of $u_0(r)$ to get a separate estimate for $\bar{a}(r)$. Because the estimates of $u_0(r)$ vary only at the country level, the data on subjective well-being are correcting only for the average utility level in a country, but not for the relative utility levels across regions within a country.[19]

E. Migration Costs

With initial technology and amenities at each location we can use data on population levels in period 1 to estimate local migration costs. Equation (7) implies that

[18] For (34) and (35) to be completely consistent, we can rewrite $\ln a(r) + \ln \varepsilon^i(r)$ in (35) as $\ln a'(r) + \ln \varepsilon'^i(r)$, where $\ln \varepsilon'^i(r)$ has mean zero. For this not to affect the subsequent estimates of amenities, $\rho \ln a'(r)$ must (up to a constant) be equal to $\rho \ln a(r)$. To guarantee this, we assume that the distribution of taste shocks across individuals surveyed by the Gallup World Poll is identical across countries. Although theoretically these distributions might be different because of the selection of migrants, empirically this is unlikely to be an issue, given that only 3 percent of the world population lives outside their country of origin.

[19] Of course, in subsequent periods $u_t(r)$ is allowed to vary freely within and across countries.

$$u_1(r) = H(r)^{\Omega}\bar{L}_1(r)^{\Omega}\bar{L}^{-\Omega}\left[\int_S u_1(v)^{1/\Omega}m_2(v)^{-1/\Omega}dv\right]^{\Omega}m_2(r).$$

Plugging this into equation (24) that relates the period 1 population distribution to amenities, land, and period 1 productivity and utility, we get

$$\begin{aligned}
&\left[\frac{\bar{a}(r)}{\hat{m}_2(r)}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}}\tau_1(r)^{-\frac{\theta}{1+2\theta}}H(r)^{\frac{\theta+\theta(1+\theta)\Omega}{1+2\theta}}\\
&\quad\times\bar{L}_1(r)^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]+\frac{\theta(1+\theta)\Omega}{1+2\theta}}\\
&=\kappa_1\int_S\left[\frac{\bar{a}(s)}{\hat{m}_2(s)}\right]^{\frac{\theta^2}{1+2\theta}}\tau_1(s)^{\frac{1+\theta}{1+2\theta}}H(s)^{\frac{\theta-\theta^2\Omega}{1+2\theta}}\varsigma(r,s)^{-\theta}\\
&\quad\times\bar{L}_1(s)^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]-\frac{\theta^2\Omega}{1+2\theta}}ds,
\end{aligned}\tag{37}$$

where

$$\hat{m}_2(r) = \frac{m_2(r)}{\bar{L}^{\Omega}\left[\int_S u_1(v)^{1/\Omega}m_2(v)^{-1/\Omega}dv\right]^{-\Omega}}.\tag{38}$$

The following lemma shows that, for a given continuous distribution of $\bar{a}(\cdot)$, $\tau_1(\cdot)$, $H(\cdot)$, and $\bar{L}_1(\cdot)$, there exists a unique solution to equation (37).

LEMMA 7. The solution to equation (37), $\hat{m}_2(\cdot)$, exists, is unique, and can be found by iteration.

Proof. It follows from theorem 2.19 in Zabreyko et al. (1975) that the solution exists and is unique if the ratio of the exponents on the right-hand side and on the left-hand side is not larger than one in absolute value. It also follows from theorem 2.19 that iterating on the equation, we always converge to the solution if the ratio of the exponents is strictly smaller than one in absolute value. Both conditions are automatically satisfied as

$$\frac{\dfrac{\theta^2}{1+2\theta}}{\dfrac{\theta(1+\theta)}{1+2\theta}} = \frac{\theta}{1+\theta} < 1.$$

QED

The solution to equation (37) yields a unique $\hat{m}_2(r)$ for all $r \in S$. The implied values of $m_2(\cdot)$ are identified only up to a scale by equation (38). All equilibrium conditions depend on the distribution of $m_2(\cdot)$ only, not on its level, so we simply normalize the level of $m_2(\cdot)$ such that its minimum is equal to one. For interpretation purposes, entering any location is costly relative to staying in one's original location (and correspondingly,

leaving any location involves a gain due to assumption 1). Note that these formulation and quantification of migration costs imply that everyone is "endowed" with the value of the location where they are born (or where their family is located in the initial period). Of course, people born in a more desirable location obtain a larger "endowment" relative to people born in less desirable places.

To solve for $m_2(\cdot)$, we need to know period 1 productivity and population levels, $\tau_1(\cdot)$ and $\bar{L}_1(\cdot)$. We use the productivity evolution equation (8) to obtain $\tau_1(\cdot)$ from $\tau_0(\cdot)$ and $\bar{L}_0(\cdot)$, while we use data on the population distribution in 2005 to obtain $\bar{L}_1(\cdot)$.[20]

F. *Simulations and Counterfactual Migration Scenarios*

Once all parameters, as well as the functions $\bar{a}(\cdot)$, $\tau_0(\cdot)$, $m_2(\cdot)$, and $\varsigma(\cdot, \cdot)$, are known, we can simulate the model forward by solving the system of three equations in lemma 3 to obtain $u_t(\cdot)$, $\bar{L}_t(\cdot)$, and $w_t(\cdot)$ for every $t = 1, \ldots$. Every period we have to update the distribution of productivities $\tau_t(\cdot)$ using equation (8).

We also want to calculate a set of counterfactual migration scenarios in which we relax migration frictions in the world. To do so, we follow the procedure above but use counterfactual migration frictions given by $m_2(\cdot)^{\vartheta}$ for $\vartheta \in [0, 1]$. When $\vartheta = 1$, migration frictions are identical to the ones we estimate in the data, so this case corresponds to keeping migration frictions unchanged. When $\vartheta = 0$, the counterfactual migration restrictions imply that $m(s, r; \vartheta = 0) = 1$ for all $s, r \in S$, so that moving to any location in the world is free. We also calculate a variety of other migration scenarios in which we use values of ϑ strictly between zero and one. Note that, relative to $\vartheta = 1$, an intermediate value of ϑ implies a distribution of migration costs with smaller differences across locations. Since we normalize $m_2(\cdot)$ by the minimum entry cost, we know that $m_2(r) \geq 1$ for all $r \in S$. Hence, $\vartheta < 1$ implies that we are decreasing the cost of entering any destination. The value of ϑ can then be understood as an index of the severity of migration frictions in the world.

G. *Backcasting*

The procedure to simulate the model described in the previous section allows us to compute the distribution of economic activity over time starting from our base year, which we set to the year 2000. To gauge the performance of our model, we are also interested in simulating the model

[20] We adjust for the discrepancy between the 5-year evolution in the data (from 2000 to 2005) and the definition of a period as 1 year in our model by dividing population growth at each cell by 5. We also adjust all levels so that total population size remains unchanged.

backward in order to compute the implied distribution of population in the past. This allows us to compare, as an overidentification test, the correlation between the past population data and the past distribution of population predicted by the model.

Equation (7) allows us to express utility as a function of population and moving costs,

$$u_t(r) = m_2(r)H(r)^{\Omega}\bar{L}_t(r)^{\Omega}\left[\frac{\int_S u_t(v)^{1/\Omega}m_2(v)^{-1/\Omega}dv}{\bar{L}}\right]^{\Omega}.$$

From equations (8), (12), and (13), productivity can be obtained as a function of population and next period's productivity, which yields

$$\tau_t(r) = \left[\frac{\mu + \gamma_1/\xi}{\gamma_1/\xi}\nu\right]^{\frac{\theta\gamma_1}{\xi\gamma_2}}\left[\int_S \eta\tau_t(s)ds\right]^{1-\frac{1}{\gamma_2}}\tau_{t+1}(r)^{\frac{1}{\gamma_2}}\bar{L}_t(r)^{-\frac{\theta\gamma_1}{\xi\gamma_2}}. \tag{39}$$

Substituting these two equations in equation (24), we obtain

$$\begin{aligned}
&\left[\frac{\bar{a}(r)}{m_2(r)}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}}H(r)^{\frac{\theta+\theta(1+\theta)\Omega}{1+2\theta}}\tau_{t+1}(r)^{-\frac{\theta}{\gamma_2(1+2\theta)}}\\
&\times \bar{L}_t(r)^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta\right]+\frac{\theta(1+\theta)\Omega}{1+2\theta}+\frac{\theta^2}{1+2\theta}\frac{\gamma_1}{\xi\gamma_2}}\\
&= \tilde{\kappa}_{Bt}\int_S\left[\frac{\bar{a}(s)}{m_2(s)}\right]^{\frac{\theta^2}{1+2\theta}}H(s)^{\frac{\theta-\theta^2\Omega}{1+2\theta}}\tau_{t+1}(s)^{\frac{1+\theta}{\gamma_2(1+2\theta)}}\\
&\times \bar{L}_t(s)^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta\right]-\frac{\theta^2\Omega}{1+2\theta}-\frac{\theta(1+\theta)}{1+2\theta}\frac{\gamma_1}{\xi\gamma_2}}\varsigma(r,s)^{-\theta}ds,
\end{aligned} \tag{40}$$

where $\tilde{\kappa}_{Bt}$ does not depend on location. Given a set of continuous functions $\bar{a}(\cdot)$, $H(\cdot)$, $m_2(\cdot)$, $\varsigma(\cdot,\cdot)$, $\tau_{t+1}(\cdot)$, and the values of structural parameters, we can solve this equation, together with world labor market clearing $\int_S \bar{L}_t(r)dr = \bar{L}$, to obtain the population distribution $\bar{L}_t(\cdot)$. We can then obtain $\tau_t(\cdot)$ from equation (39). Using theorem 2.19 in Zabreyko et al. (1975), as before, the solution to equation (40) exists and is unique if

$$\frac{\alpha}{\theta} - \frac{\gamma_1}{\xi}\left[\frac{1}{\gamma_2} - 1\right] \leq \lambda + 1 - \mu + \Omega.$$

If the inequality is strict, we can find a solution using the same iterative procedure as the one described in the proof of lemma 3. Note that this condition is strictly weaker than the condition guaranteeing the existence and uniqueness of an equilibrium in proposition 1, which is given by $\alpha/\theta + \gamma_1/\xi \leq \lambda + 1 - \mu + \Omega$, since $\gamma_2 < 1$. This result simply reflects that, since technology increases over time, dynamic agglomeration effects

are not an issue when running time backward. Hence, a backward prediction can be uniquely computed. We compare the outcome of this exercise with historical data below.

V. The Geographic Evolution of the World Economy

This section has three aims. First, we want to enhance our understanding of the relation between space and growth. Our model has predictions for the future evolution of the spatial distribution of population and productivity, as well as for the economy's aggregate growth rate. Second, we want to assess the accuracy of our model's predictions for the spatial distribution of population. To do so, we perform a backcasting exercise and compare its results with historical data. Finally, we want to understand the welfare impact of relaxing migration restrictions and how doing so changes the distribution of economic activity in the balanced-growth path.

A. Benchmark Calibration

As explained in Sections IV.D and IV.E, we use cell-level data on land, population at two dates, wages, and trade costs to recover amenities, productivity, and mobility frictions. Whenever time-variant, the data are mostly for 2000 (see app. C for more details). All outcomes are plotted in natural logarithms. Figure 1 presents the results from the inversion exercise with actual migration frictions to calculate the fundamental productivities and amenities. By "fundamental" we refer to the part of productivity and amenities that does not depend on population density. That is, it does not take into account the positive agglomeration economies that benefit productivity and the negative congestion effects that hurt amenities. We also present the values of utility-adjusted amenities that result directly from the inversion $\bar{a}(r)/u_0(r)$ and the values of the entry migration costs $(m_2(r))$ that we identify from population movements between years 2000 and 2005 as we described above.

Fundamental productivity exhibits the expected patterns (see fig. 1*A*). Productivity is generally high in North America, Europe, and Japan. When we look within countries, the main cities tend to display particularly high levels of productivity. Mexico, for example, exhibits relatively low productivities, except for Mexico City and Monterrey. In China, Beijing and Shanghai have clearly higher fundamental productivities than the rest of the country. This may reflect rules and regulations, as well as local educational institutions that lead to higher-skilled populations. It is also well known that cities often tend to attract more productive workers. In that sense it should not come as a surprise that urban areas stand out as places with high fundamental productivities. Clearly, since we are abstracting from local capital investments, fundamental productivities also reflect the local

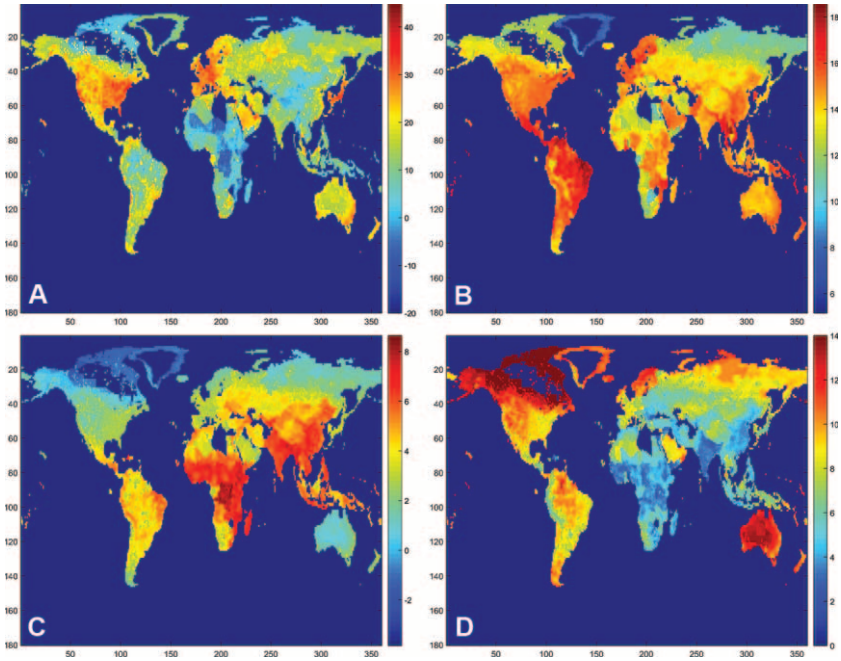

FIG. 1.—Benchmark calibration: results from the inversion and migration costs. *A*, Fundamental productivities: $\tau_1(r)$. *B*, Fundamental amenities: $\bar{a}(r)$. *C*, Amenities over utility: $\bar{a}(r)/u_0(r)$. *D*, Migration costs: $m_2(r)$.

stock of capital, which tends to be larger in cities and, more broadly, in developed economies.

As for fundamental amenities (see fig. 1*B*), the highest values can be found in South America and, in particular, Brazil. North America also enjoys high fundamental amenities, particularly in urban areas. In Europe, Scandinavia and the Netherlands exhibit high amenities, while Portugal and southern Europe fare worse. Thailand and southeastern Australia are also desirable places to live, while some of the lowest amenities are found in Africa and Siberia. In figure 1*C* we also present fundamental amenities divided by the initial utility level, $\bar{a}(r)/u_0(r)$. As we show in lemma 6, these values are the direct result of inverting the model using population and GDP per capita. Poor and densely populated areas have high values since the model rationalizes this combination of facts through either high amenities or low utility associated with migration restrictions. Central Africa as well as large regions in Asia exhibit particularly high ratios. Comparing figures 1*B* and 1*C*, one can ascertain the role of initial utility-level differences due to migration costs, which we measured using the subjective well-being data.

To validate our identification strategy for amenities, it is worthwhile to correlate our estimates of fundamental amenities with commonly used exogenous measures of quality of life. We therefore collect data on different measures of geography (distance to oceans, distance to water, elevation, and vegetation) and climate (precipitation and temperature). The correlations between our estimates of amenities and these measures are consistent with the literature. For example, we find that people like living close to water and prefer higher average temperatures and precipitation (Davis and Ortalo-Magné 2011; Albouy et al. 2016). Qualitatively, the results do not depend on whether we look at all cells of the world, at only cells in the United States, or at randomly drawn cells across countries. This suggests that our methodology of using data on subjective well-being to identify amenities performs well. To further reinforce this point, placebo correlations based on alternative identification strategies of amenities yield correlations that are no longer the same across countries and within countries. Section B of appendix A presents these correlations and provides further details.

In figure 1*D* we present the measured entry migration costs $m_2(r)$. The actual migration cost between s and r is given by $m_1(s)\,m_2(r)$, which is equal to $m_2(r)/m_2(s)$ as a result of assumption 1. Since we obtain these costs so as to exactly match population changes between two subsequent years in the data, they should be interpreted as the total flow costs of entering a region, including information and psychological costs, the cost of actually getting there and settling there, and of course any legal migration restrictions. The figure shows very large log values of $m_2(r)$ in many parts of the world, suggesting that a very large share of the migration costs likely reflects legal impediments. It is hard to go to parts of Europe, particularly Scandinavia, parts of Canada and Alaska, Australia, New Zealand, and parts of the United States. One can distinguish the border between Mexico and the United States, Mexico and Central America, as well as Finland and Russia among others. Entering northern regions that are cold and hard to get to and settle is costly. Entering Africa, India, and China, as well as parts of Russia, is cheap. Overall these costs appear reasonable, but more importantly, they make the model match exactly observed population flows. In the next section we show that they also imply past population distributions that compare well with historical data.

Figure 2 (different panels) maps population, productivity, utility and real income per capita. Following the theory, average productivity in each cell is defined as $[\tau_t(r)\bar{L}_t(r)^{\alpha}]^{1/\theta}$. This productivity includes the positive local agglomeration effect and takes into account that each location draws productivities from a Fréchet distribution. In contrast to figure 1, these are equilibrium outcomes in the first period of the model. Note that since $u_0(r)$ is based on our measure of subjective well-being and hence is con-

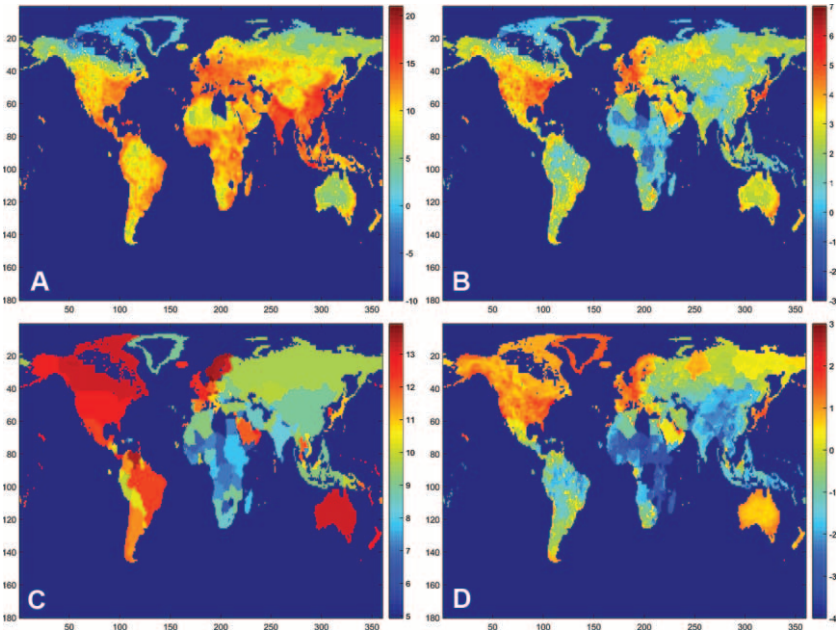

FIG. 2.—Equilibrium in benchmark calibration (period 1). *A*, Population density. *B*, Productivity: $[\tau_t(r)\bar{L}_t(r)^{\alpha}]^{1/\theta}$. *C*, Utility: $u_t(r)$. *D*, Real income per capita: $y_t(r)$.

stant within countries, $u_1(r)$ also varies little within countries since the adjustment takes place over a single year.

Throughout this section we use $u_t(r)$ as our measure of utility since it is a sufficient statistic of the characteristics of a location—real consumption and amenities—that individuals value. Actual experienced utility also includes the idiosyncratic preferences of individuals and possibly migration costs (if not rebated). At the end of this section we compare the behavior of these alternative measures of welfare.

A first observation is that the correlation between population density and productivity across countries is not that strong: there are some densely populated countries with high productivity levels, such as many of the European countries, but there are also some densely populated countries with low productivity levels, such as some of those located in sub-Saharan Africa. Countries that are densely populated in spite of having low productivity must have either low levels of utility (e.g., sub-Saharan Africa) or high levels of amenities (e.g., Latin America).[21] A second observation is that the same relation between population density and productivity is much stronger across locations within countries, where the high-productivity areas typically tend to be the metropolitan areas. This is not surprising since mobil-

[21] See also the map of the subjective welfare measures in Sec. C of app. A.

ity frictions tend to be smaller within countries, and so utility differences across locations within countries are relatively small. Thus, the negative congestion effects from living in densely populated areas have to be compensated by either higher productivity or better amenities. In line with this reasoning, we find that metropolitan areas have substantially higher productivity than the less dense neighboring locations, as well as somewhat higher, although less marked, real income per capita. The relatively high real income in cities suggests that cities must enjoy lower amenity levels due to the negative effect of density on amenities. Diamond (2016) provides empirical evidence on the importance of this effect.

Now consider the evolution of this economy over time, when we keep migratory barriers, as measured by $m_2(\cdot)$, unchanged. Figure 3 (different panels) maps the predicted distributions of population, productivity, amenities, and real income per capita in period 600, at which point the economy has mostly converged to its balanced-growth path. Videos that show the evolution of these variables over time and over space are available online. To visualize the changes over time, we should compare figure 3, which represents the year 2600, with figure 2, which represents the year 2000.

Clearly, over time the correlation between population and productivity across countries becomes much stronger. As predicted by the theory, in

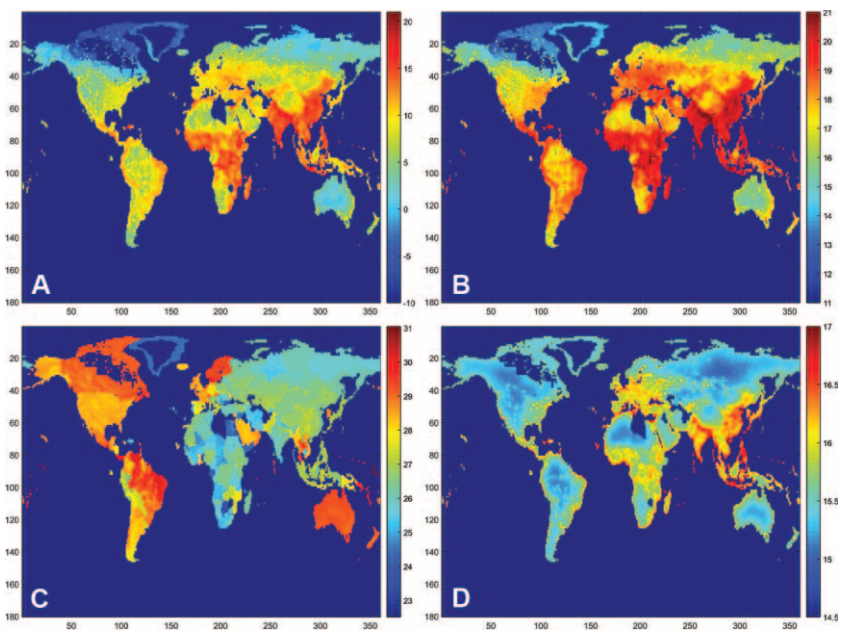

FIG. 3.—Equilibrium keeping migratory restrictions unchanged (period 600). *A*, Population density. *B*, Productivity: $[\tau_t(r)\bar{L}_t(r)^{\alpha}]^{1/\theta}$. *C*, Utility: $u_t(r)$. *D*, Real income per capita: $y_t(r)$.

the long run, high-density locations correspond to high-productivity locations. This can easily be seen when comparing the maps for period 1 and period 600. While in period 1 the population and the productivity maps look quite different, by period 600 they look very similar. As shown by the solid curve in panel B of figure 4, which presents the correlation of population density and productivity over time, the correlation increases from around .38 in period 0 to 1.0 by period 600. Note, however, that the correlation between population density and real GDP, which is driven not only by productivity but also by local prices, and therefore transport costs and geography, also grows but is never higher than .7. The solid curve in panel A of figure 4 shows this. Since the correlation between productivity and population density grows faster than the correlation between output and population density, panel C shows that the correlation between productivity and output exhibits a U-shaped pattern. Population responds slowly to increases in productivity because of the presence of migration restrictions.

The model's prediction that the correlation between population density and income becomes stronger as the world becomes richer is present in the data today. Although across all cells of the world the correlation today is negative, it is larger in wealthy regions. In 2000 the correlation between population density and real income per capita was −.11 in Africa, .33 in western Europe, and .50 in North America. Similarly, in the United States the correlation is greater in high-income zip codes. In particular, in 2000 the correlation between population density and income per capita was .36 in zip codes with an income per capita above the median, compared to −.06 in zip codes below the median. Section A of appendix A provides further details.

Another remark is that the high-productivity, high-density locations 600 years from now correspond to today's low-productivity, high-density locations, mostly countries located in sub-Saharan Africa, South Asia, and East Asia. In comparison, most of today's high-productivity, high-density locations in North America, Europe, Japan, and Australia fall behind in terms of both productivity and population. Clearly, this is not the case in terms of welfare as measured by $u_t(r)$ (namely, amenity-adjusted real income not including either migration costs or idiosyncratic shocks), where America, Europe, and Australia remain strong throughout.

This productivity reversal can be understood in the following way. The high population density in some of today's poor countries implies high future rates of innovation in those countries. Low inward migration costs and high outward ones imply that population in those countries increases, leading to greater congestion costs and worse amenities. As a result, today's high-density, low-productivity countries end up becoming high-density, high-productivity, high-congestion, and low-amenity countries, whereas today's high-density, high-productivity countries end up becoming

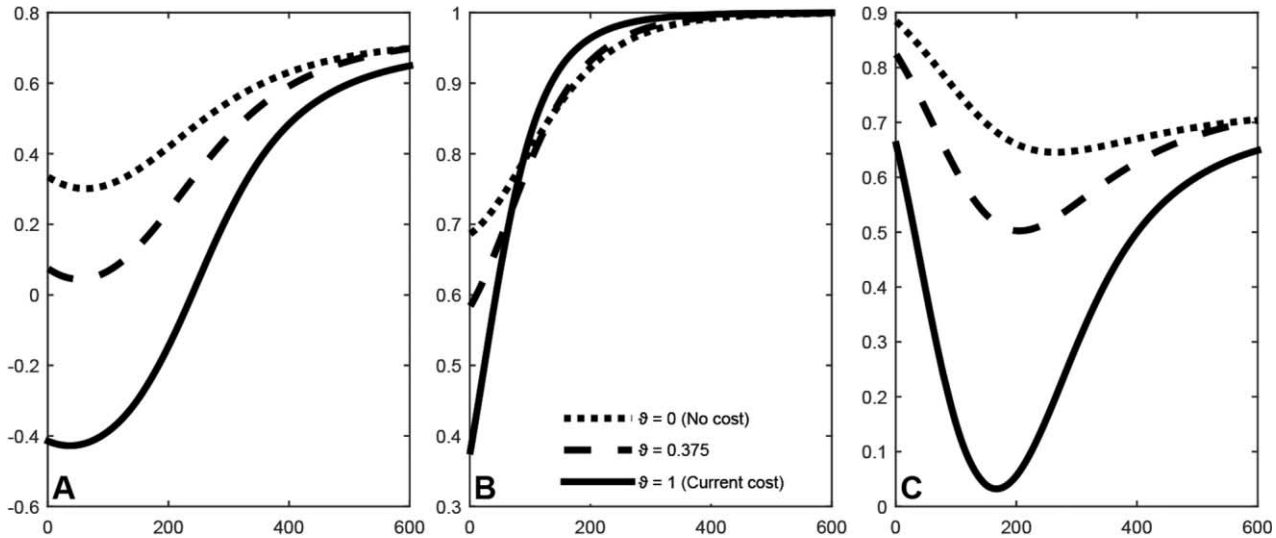

FIG. 4.—Correlations with different migration restrictions. *A*, Correlation of ln $y(r)$ and ln $L(r)$. *B*, Correlation of ln(productivity) and ln $L(r)$. *C*, Correlation of ln(productivity) and ln $y(r)$.

medium-density, medium-productivity, low-congestion, and high-amenity countries; the United States is among them. Australia's case is somewhat different since its low density and high inflow barriers imply that it becomes a low-productivity, high-amenity country. As can be seen in figure 4, given that the correlation between population density, productivity, and real income per capita becomes relatively high, it must be that low-density, high-utility countries have high amenities.

These dynamics imply a reallocation of population from high-utility to low-utility countries, as measured by $u_t(r)$. Since the differences in amenity-adjusted real GDP are now smaller, idiosyncratic preferences compensate more individuals to stay in areas with low $u_t(r)$. In principle this could be driven by decreased migration from low- to high-utility countries or by increased migration from high- to low-utility countries. As increased innovation pushes up relative utility in low-income countries, fewer people want to migrate out, and more people want to migrate in, keeping relative utility levels about, although not completely, constant. Over time, this reallocation of population peters out. As predicted by lemma 4, decreasing returns to innovation, together with congestion costs, imply that, in the long run, the distribution of population reaches a steady state in which all locations innovate at the same constant rate.

Most of our discussion so far has focused on the changing differences across countries, but there are also interesting differences within countries. When focusing on the population distribution within countries, we observe that as the population share of North America and Europe declines, the locations that better withstand this decline are the coastal areas, which benefit from lower transport costs and thus higher real income. In the countries whose population shares are increasing, such as China, India, and parts of sub-Saharan Africa, there is less evidence of coastal areas gaining. This movement toward the coasts is evident in all countries when focusing on real GDP per capita, since coastal locations are high-density, low-amenity locations (due to congestion) that have high market access and therefore more innovation.

Figure 5 presents the average growth rates and levels of productivity, real output, and welfare for different scenarios.[22] The benchmark calibration, which leaves mobility restrictions unchanged, corresponds to $\vartheta = 1$ (solid curve).[23] Consistent with the argument above, figure 5 shows that the average growth rates of productivity, real GDP per capita, and utility converge to constants after about 400 years. Note that the long-run growth

[22] World average productivity at time t is defined as $\int_S [H(r)\bar{L}_t(r)/\bar{L}][\tau_t(r)\bar{L}_t(r)^{\alpha}]^{1/\theta}dr$, i.e., the population-weighted average of locations' mean productivity levels. World average real GDP at t is defined analogously as $\int_S [H(r)\bar{L}_t(r)/\bar{L}]y_t(r)dr$. World average utility is measured as $\int_S [H(r)\bar{L}_t(r)/\bar{L}]u_t(r)dr$, and it is based on the measure u_t, which does not include migration restrictions or idiosyncratic preference shocks.

[23] We discuss the other curves in fig. 5. They represent alternative migration scenarios.

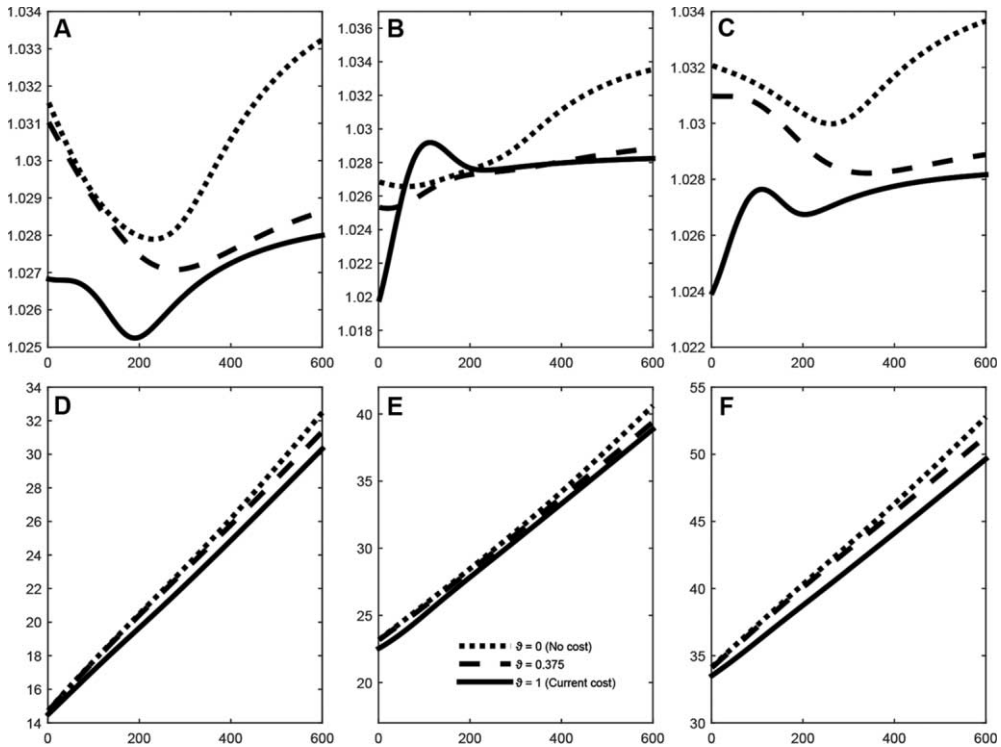

FIG. 5.—Growth rates and levels of productivity, real GDP, and utility under different migration scenarios. *A*, Growth rate of productivity. *B*, Growth rate of real GDP. *C*, Growth rate of utility (u). *D*, Ln world average productivity. *E*, Ln world average real GDP. *F*, Ln world average utility (u).

rate in real GDP per capita is higher than the initial growth rate, similar to the pattern of utility growth. The growth rate of real GDP increases from the calibrated 2 percent in 2000 to about 2.9 percent and then slowly decreases, to end slightly above 2.8 percent. Economic activity is more concentrated in the balanced-growth path, leading to more investment and faster growth. Productivity growth, in contrast, declines for the first 200 years, reflecting transitory losses from the initial population reallocation toward low-productivity regions. From then on the spatial distribution of population density, $\bar{L}_t(\cdot)$, changes little in this migration scenario. Finally, recall that the growth rate in utility is equal to the growth rate in real GDP per capita minus λ times the growth rate of population. Initially, the correlation between the growth in real GDP per capita and the growth in population is negative—many of the high-growth places in richer countries are losing population—so that the growth rate in utility is greater than the growth rate in GDP per capita. In the long run, when the steady state is reached, there is no more reallocation of population across space, so that real GDP per capita, productivity, and utility grow at the same rates. The bottom panels of figure 5 show the levels of world average productivity, real GDP, and utility as measured by u_t.

One of the forces that determine market size in the model is the ability to trade with other regions subject to transport costs. The level of transport costs determines the geographic scope of the area that firms consider when deciding how much to invest in innovation. In our framework, each cell trades with a group of other cells, within and across countries. Recall that we parameterize trade costs such that we match the mean empirical estimate of the elasticity of bilateral cell-level trade flows to distance (Head and Mayer 2014). We can also aggregate trade at the country level in order to calculate the overall level of international trade. The world trade share is calculated as

$$\text{World Trade Share}_t = 1 - \frac{\int_S \int_{c(s)} \pi_t(s, r) y_t(s) H(s) \bar{L}_t(s)\, dr ds}{\int_S y_t(s) H(s) \bar{L}_t(s)\, ds},$$

where $c(s)$ denotes the set of cells in the country of cell s. The second term denotes the share of domestic consumption, where the numerator includes the expenditure on domestic goods in the world and the denominator is total world GDP. In the simulation above, this trade share equals 14.1 percent in 2000. Though not targeted by our model, this matches the actual ratio of trade to GDP in the world economy.[24] Over time, the

[24] According to the World Trade Organization, the ratio of exports plus imports to GDP in the world was 24 percent in 2000, increasing rapidly to 32 percent by 2008. One needs to divide these numbers by 2 to make them comparable to our statistic. This gives an average share of 14 percent over the 2000s, identical to the share in our benchmark calibration.

model predicts the trade share to increase slightly, reaching 15.1 percent by 2600. This upward trend is due to the emergence of high-productivity clusters in Africa, South Asia, and East Asia, all of which belong to different countries.

Our model incorporates both national and international trade barriers and can be used to evaluate the overall effect of changes in trade costs. In particular, it can be used to measure both the static and dynamic gains from trade. This is in contrast with most of the trade literature, which focuses on static trade models without internal geography.[25] Consider a counterfactual scenario that increases trade costs by 40 percent in the first period. Such a change reduces the trade share to 3.90 percent in period 1 and to 4.27 percent in the long run. The real GDP loss is 30.1 percent and the welfare loss, measured by u_t, is 34.2 percent. These losses are an order of magnitude higher than the ones reported for static, one-sector, models in the survey by Costinot and Rodríguez-Clare (2014), who compute a real income loss from a similar change in trade costs of 2.28 percent. Incorporating dynamic and within-country effects seems to change the impact of trade frictions dramatically.

B. *Predicting the Past*

In this section we use the algorithm described in Section IV.G to compute the model's implied population distribution in the past. Starting with the calibration described in Section V.A for the year 2000, we run the simulation backward and compute the distribution of population in all cells of the world all the way back to the year 1870. We then compare the model-implied distribution to data from the Penn World Tables 8.1 for every decade until 1950 and to data from Maddison (2001) for 1913 and 1870. These historic data are easily available at the country level, so we aggregate the cells in the model to the country level.

This exercise is intended as an overidentification check for the model, given that none of these historic population data were used in the quantification. Of course, there is a variety of world events and shocks that the model does not incorporate, so it would be unreasonable to ask the model to perfectly match the data. The model's implications should be understood as the distribution of population that would result in the absence of any forces and shocks not directly modeled. Examples of such unmodeled shocks abound: the two world wars, trade liberalization, decolonization, and so forth. With that caveat in mind, it is interesting to see how much of the population changes the model can account for. Given that our framework attempts to model long-run forces driving population changes and growth, rather than short-term fluctuations and shocks, we

[25] See, e.g., the survey in Costinot and Rodríguez-Clare (2014).

would expect the model to perform best over the medium run, when short-term shocks are less important, but the economy has not changed in fundamental ways not modeled here.

Table 2 reports the results. The model does very well: in levels, the correlations back to 1950 are all larger than .96. Clearly, this is in large part due to the fact that we match population levels in 2000 by construction, and population changes are not that large (about 3 percent per year). Hence, we also present the correlations of population growth rates. The correlations for growth rates are clearly lower, but altogether still relatively high. The correlation between the population growth rate from 1950 to 2000 in the model and the data is .74. We view this as quite a success of the model. If we go all the way to 1870, the predictive power of the simulation is smaller, but the correlation of growth rates is still .34. In spite of the disruption of two world wars and many other major world events not accounted for here, the model preserves substantial predictive power.

C. *Evaluating Mobility Restrictions*

We now analyze the effect of completely or partially relaxing existing migration restrictions. We start with the free mobility scenario and then present some calculations for scenarios with partial mobility in which ϑ is between one (current restrictions) and zero (free mobility).

1. Free Migration

In this exercise we evaluate the effects of completely liberalizing migration restrictions. We start off with the benchmark calibration in period 0 and then reallocate population such that $m_2(r) = 1$ for all $r \in S$, as corresponds to the case with $\vartheta = 0$. For now, we take amenity-adjusted real GDP, $u_t(r)$, as our measure of a location's utility. With free migration people move from locations with low u_t to locations with high u_t. This reduces utility differences across locations because of greater land congestion and lower amenities in high-utility locations (and the opposite in low-utility locations). This reallocation of population not only has a static ef-

TABLE 2
Country-Level Population Correlations, 1870–2000: Model vs. Data

	Penn World Tables 8.1					Maddison	
Year t	1990	1980	1970	1960	1950	1913	1870
Correlation log population in t	.993	.991	.982	.974	.965	.843	.681
Correlation population growth from t to 2000	.414	.535	.504	.671	.742	.461	.344
Number of countries	152	131	131	102	53	76	76

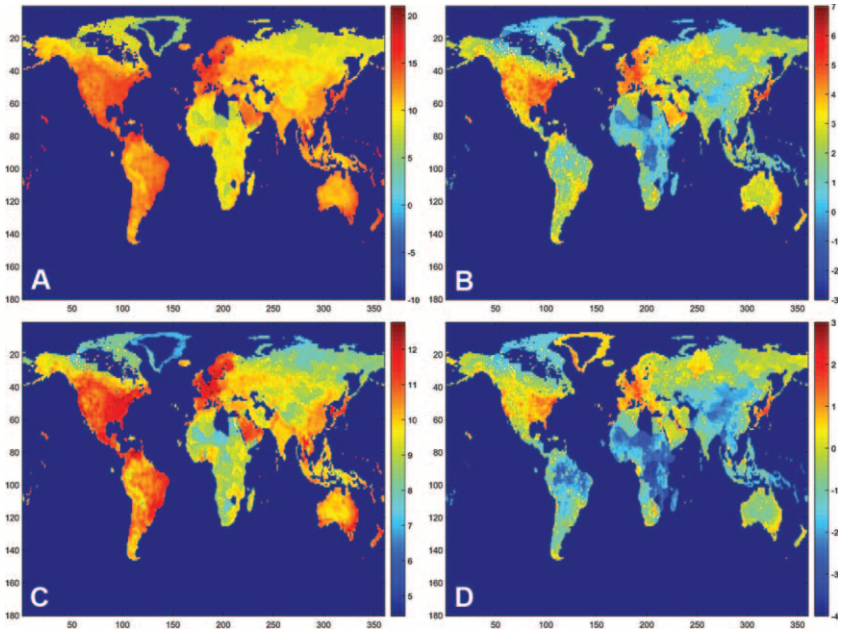

FIG. 6.—Equilibrium with free migration (period 1). *A*, Population density for $\vartheta = 0$. *B*, Productivity for $\vartheta = 0$: $[\tau_t(r)\bar{L}_t(r)^{\alpha}]^{1/\theta}$. *C*, Utility for $\vartheta = 0$: $u_t(r)$. *D*, Real income per capita for $\vartheta = 0$: $y_t(r)$.

fect but also has a dynamic effect by putting the economy on a different dynamic path. As we have emphasized repeatedly, even though $m_2(r) = 1$ for all $r \in S$, differences in $u_t(r)$ are not eliminated, since idiosyncratic taste shocks imply that the elasticity of population to $u_t(r)$ is positive but finite.

Figure 6 and figure 7 map the distributions of population, productivity, utility, and real income per capita for period 1 and period 600, under the assumption that people are freely mobile across locations.[26] Compared to the exercise in which we kept migratory restrictions unchanged, several observations stand out. First, migration increases the initial correlation between population density and productivity across countries, so that we see fewer countries in which high density and low productivity coexist. In panel B of figure 4 we see that in the case of complete liberalization ($\vartheta = 0$) the initial correlation between density and productivity increases from around .38 to nearly .69. Second, this initial effect has important dynamic consequences. Because today's poor countries lose population through migration, they innovate less. As a result, and in contrast to the

[26] For videos that show the gradual evolution over time and over space, see the online edition.

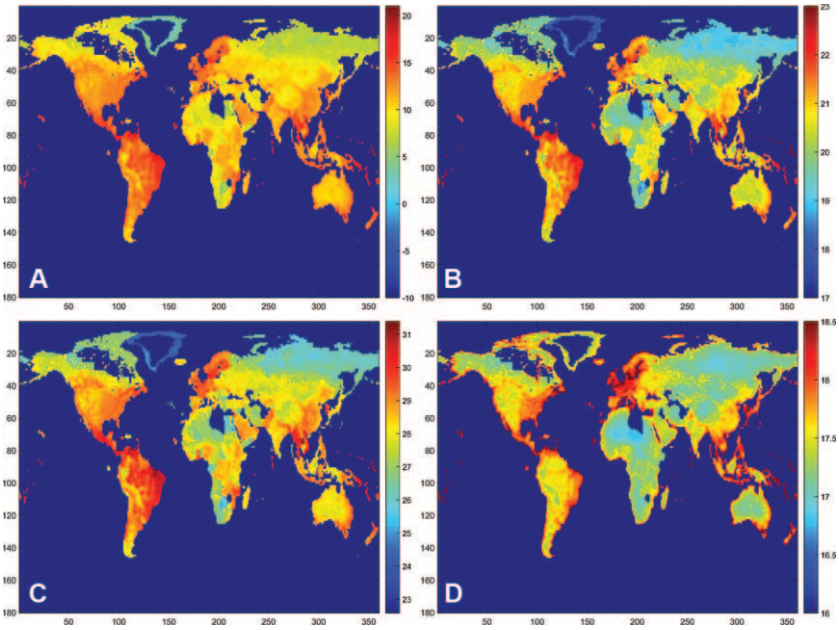

FIG. 7.—Equilibrium with free migration (period 600). *A*, Population density for $\vartheta = 0$. *B*, Productivity for $\vartheta = 0$: $[\tau_t(r)\bar{L}_t(r)^{\alpha}]^{1/\theta}$. *C*, Utility for $\vartheta = 0$: $u_t(r)$. *D*, Real income per capita for $\vartheta = 0$: $y_t(r)$.

previous exercise, no productivity reversal occurs between the United States, India, China, and sub-Saharan Africa. Third, this absence of a large-scale productivity reversal does not mean that relative productivities across countries remain unchanged. Some countries, such as Venezuela, Brazil, and Mexico, start off with relatively high utility levels but relatively low productivity levels. This means they must have high amenities. Because of migration, they end up becoming some of the world's densest and most productive countries, together with parts of Australia, Europe, and the United States. Fourth, migration changes the determination of the development path of the world and thereby increases the growth rate in the balanced-growth path, as is clear from figure 5. In the balanced-growth path, real GDP growth increases by about 0.5 percentage points. Fifth, within countries, there is stronger evidence of an increasing concentration of the population in coastal areas, compared to the benchmark case. The greater concentration of population within countries, which was already apparent in the benchmark case, is now reinforced by greater migration across countries. Sixth, the importance of coastal areas becomes even more apparent when we look at real income per capita. The fact that locations close to coastal areas have lower real income per capita but high population suggests that those

locations have higher amenities than the coastal areas themselves because of congestion.

When analyzing the average growth in real income per capita and utility, two differences become immediately apparent in figure 5 when comparing the case of free migration ($\vartheta = 0$) to the case of no liberalization ($\vartheta = 1$). First, mobility increases the long-run growth rate of the economy as well as the average level of welfare, output, and productivity in the world. Second, growth in utility drops substantially in the short run as many people move to areas with high real GDP; hence these areas become more congested and become worse places to live (lower amenities). This initial loss in growth is, however, compensated in the long run by a large surge in productivity growth after year 2200.

2. Partial Liberalization

We now present the changes in discounted real income and welfare, as well as migration flows, that result from a variety of scenarios in which we partially relax migration frictions. Figures 4 and 5 in the previous section already presented correlations as well as growth rates and levels of the different variables for an example scenario of partial liberalization, where we set $\vartheta = 0.375$.[27] As can be seen in table 3, although this value of ϑ is around the middle between keeping existing migration restrictions unchanged and free migration, it leads to reallocations of population that are closer to the free migration case. When we liberalize migration fully, 70.3 percent of the population changes country immediately; the corresponding number when $\vartheta = 0.375$ is 52.7 percent. As expected, the share of population that moves decreases monotonically with the magnitude of migration frictions as measured by ϑ.

In terms of welfare, migration frictions are tremendously important, particularly when we move closer to free migration.[28] As shown in table 3, a liberalization that implies that about a quarter of the world population moves at impact ($\vartheta = 0.75$) yields an increase in real output of 30.6 percent. In terms of welfare, the increase depends on the measure we use. If, as we have done so far, we measure welfare as the present discounted value of the population-weighted average of $u(r)$, the resulting welfare increase when we set $\vartheta = 0.75$ is 60 percent. This measure includes neither the mobility costs nor the idiosyncratic preference shocks. As we have noted before, given that a large share of the migration costs are likely

[27] We chose $\vartheta = 0.375$ to maximize visibility in the figures. Table 3 presents statistics for many other intermediate values of ϑ.

[28] These calculations assume that the economy is on its balanced-growth path after period 600. Note in fig. 5 that in all cases the balanced-growth path growth rate remains strictly below 3.5 percent, and so, since $\beta = 0.965$, the discounted present values of income and utility are well defined.

TABLE 3
THE IMPACT OF MOBILITY FRICTIONS

Mobility ϑ	Real Income[a] % Δ w.r.t. $\vartheta = 1$ (1)	Welfare (u)[b] % Δ w.r.t. $\vartheta = 1$ (2)	Expected Welfare $E[u\varepsilon^i]$[c] % Δ w.r.t. $\vartheta = 1$ (3)	Migration Flows[d] % Δ from $t = 0$ to 1 (4)
1[e]	0	0	0	.30
.875	13.8	26.4	9.1	10.0
.750	30.6	59.8	19.2	21.2
.500	69.2	144.3	40.5	43.2
.375	87.1	188.8	50.3	52.7
.250	101.6	228.8	58.9	60.2
.125	113.2	264.5	67.3	66.0
0[f]	125.8	305.9	79.0	70.3

[a] Change in present discounted value (PDV).
[b] Change in PDV of population-weighted average of cells' discounted utility levels $u(r)$.
[c] Change of PDV of expected discounted utility levels as measured in eq. (4).
[d] Net population changes in countries that grow divided by $\bar{L}$ immediately after the change in ϑ.
[e] Observed restrictions.
[f] Free mobility.

to derive from legal restrictions, accounting for the direct benefits from lowering migration restrictions would tend to grossly overestimate the gains from liberalization. However, it might make sense to incorporate idiosyncratic locational preferences into our welfare calculation. If we do so and calculate the value of the expected utility of agents as measured by equation (4), we obtain a welfare gain of 19.2 percent (as can be observed in col. 3 table 3 with heading $E[u\varepsilon^i]$). The second measure is naturally smaller since agents are less selected when migration costs drop.

Full liberalization that sets $\vartheta = 0$ leads to slightly less than three-quarters of the world population migrating at impact and gives welfare gains, as measured by amenity-adjusted real GDP, of 305.9 percent and an increase in real income of 125.8 percent.[29] Furthermore, if we take into account idiosyncratic preference shocks, the gain is 79 percent. As explained before, the welfare gains when taking into account locational preferences are smaller because there is less selection under free mobility. These idiosyncratic preferences have sometimes been interpreted as capturing another type of mobility frictions. This is reflected in the elasticity of population to u_t being finite. When assessing the gains from free migration, we are not relaxing those implicit frictions.

[29] To the best of our knowledge, this is the first paper to analyze the global gains from mobility in an endogenous growth model with spatial heterogeneity, costly trade, and amenities. Previous work that focuses on models with capital accumulation estimates long-run gains in income per capita of around 100 percent (Klein and Ventura 2007; Kennan 2013). Though not directly comparable to the present discounted numbers reported in table 3, we find somewhat larger effects. The main reason is that in our model mobility generates both a level and a long-run growth effect, whereas in these other models there is only a level effect as the economy moves to its new steady state.

Two comments are in order when analyzing the numbers in table 3. First, the mobility numbers presented in table 3 include agents who move across countries at impact in period 1 only. Of course, as a result of migratory liberalization, the entire development path of the world economy changes significantly as shown in figures 4–7. Second, in general the gains in utility as measured by amenity-adjusted real GDP are much larger than the gains in real income, especially when migration becomes freer. Relaxing migration restrictions tends to relocate people to high-amenity regions. In the case of free migration, this reallocation accounts for nearly 60 percent of the total welfare effect. Although the planner's problem is intractable and unsolvable, this finding suggests what a planner might want to do. Given that our model implies that in the long run the high-productivity places will be the high-density places, ideally one would like to see these high-density places in high-amenity places. Hence, a planner interested in maximizing long-run welfare would probably want to implement policies that facilitate people moving to high-amenity places. Liberalizing migration restrictions pushes the economy in that direction.

Figure 8 presents growth rates of welfare using the two alternative measures. The figures in the first column simply reproduce the ones we presented in figure 5; the ones in the second column present growth rates

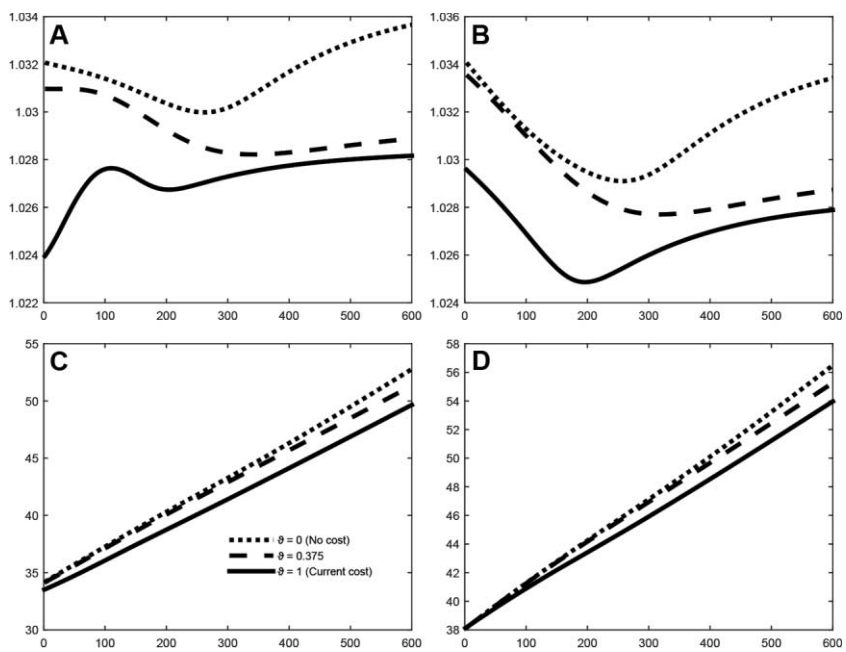

FIG. 8.—Growth rates and levels with different migration restrictions. *A*, Growth rate of utility (u). *B*, Growth rate of $E[u\varepsilon^i]$. *C*, Ln world average utility (u). *D*, Ln $E[u\varepsilon^i]$.

and levels of welfare as measured by $E[u\varepsilon^i]$ as in equation (4). With both measures the growth rate of welfare is in general higher with freer migration. Fluctuations over time in the growth rate of $E[u\varepsilon^i]$ are a bit larger. When we compare the levels of welfare, the two measures exhibit a similar behavior.

So far we have ignored the interaction between migration liberalization and trade liberalization. Of course, the gains from lowering migratory barriers are likely to depend on the level of trade costs. To explore this question, we compare the gains from free migration under different trade costs. In particular, we analyze how those gains change if we increase trade barriers by 40 percent. Completely liberalizing migration increases real income by 125.8 percent if we leave trade restrictions unchanged, but those gains rise to 156.9 percent in a world with trade costs that are 40 percent higher. The corresponding figures in terms of welfare, as measured by amenity-adjusted real GDP, are 305.9 percent and 382.9 percent, respectively. The larger gains from free migration when trade costs are higher suggest that trade and migration are substitutes: when trade is less free, the impact of liberalizing migration is larger.

VI. Conclusion

The complex interaction between geography and economic development is at the core of a wide variety of important economic phenomena. In our analysis we have underscored the fact that the world is interconnected through trade, technology diffusion, and migration. We have conducted our analysis at a fine geographic detail, enabling us to incorporate the significant heterogeneity across regions in amenities, productivity, geography, and transport costs, as well as migration frictions. This framework and its quantification have allowed us to forecast the future evolution of the world economy and to evaluate the impact of migration frictions. Our confidence in this exercise was enhanced by the model's ability to backcast population changes and levels, quite accurately, for many decades.

Our results highlight the complexity of the impact of geographic characteristics as well as the importance of their interaction with factor mobility. We found that relaxing migration restrictions can lead to very large welfare gains but that the world economy will concentrate in very different sets of regions and nations depending on migratory frictions. This inequity in the cross-country economic implications of relaxing migration frictions will (and does) undoubtedly lead to political disagreements about their implementation. For example, developed economies today will guarantee their future economic superiority only in scenarios in which the world relaxes migration restrictions. We have abstracted from these political economy considerations, but they are clearly extremely important.

Even though our framework incorporates a large set of forces in a rather large general equilibrium exercise, we had to make the choice of leaving some important forces aside. Perhaps the most relevant one is the ability of the world economy to accumulate a factor of production like capital. In our framework, only technology can be accumulated over time, not capital. In our view, the ability of regions to invest in technology substitutes partially for the lack of capital accumulation, but clearly not fully. We also abstracted from local investments in amenities. In our framework, amenities vary only through changes in congestion as a result of in- or out-migration. Finally, we also abstracted from individual heterogeneity in education or skills. Given our long-term focus, we view educational heterogeneity as part of the local technology of regions. It is obviously the case that individuals can migrate with their human capital, and we ignore this. However, the human capital embedded in migrants lasts mostly for only one generation. Ultimately, migrants and, more important, their descendants will obtain an education that is commensurate with the local technology in a region.

The framework we have proposed can be enhanced in the future to incorporate some of these other forces. However, as it already stands, it is a useful tool to understand the dynamic impact of any spatial friction. We have illustrated this using migration frictions, but clearly one could do a large variety of other exercises. Two important ones that we hope we, or other researchers, will do in the future relate to climate change and trade liberalizations. Both the international trade and the environmental literatures require a quantitative framework to understand the dynamic impact of environmental and trade policy that takes into account, and measures, local effects. The quantitative framework in this paper is ready to take on these new challenges.

Appendix A

Density-Income Correlation and Subjective Well-Being

A. Correlation of Density and Income

In our theory, the presence of both static and dynamic agglomeration economies, together with the role played by amenities, implies that the correlation between density and income per capita is relatively low (and possibly negative) when income per capita is low and relatively high when income per capita is high. Two forces drive this result. On the one hand, the standard positive correlation between density and per capita income, due to static agglomeration economies, is stronger in high-productivity places that benefit from greater dynamic agglomeration economies. This explains why the correlation between density and per capita income is increasing in per capita income. On the other hand, mobility means that high-amenity places tend to have both high population density and low per capita income. This explains why in relatively low per capita income locations, where dynamic agglomeration economies are weak, the correlation be-

tween density and per capita income might be negative. To see whether these findings hold in the data, we provide evidence for the United States and the globe.

Evidence from the United States.—For the United States we do two exercises. First, we compute the correlation between population density and income per worker across US zip codes. We split zip codes into different groups: those with income per capita (worker) below the median and those with income per capita (worker) above the median. The theory predicts a higher correlation between density and income per capita for the latter group and possibly a negative correlation for the former group. We also consider a finer split-up into four different groups by income per capita quartile. In that case, we would expect the correlation to increase when we go from lower income per capita quartiles to higher income per capita quartiles.

Table A1 reports the results. Data on population, mean income per worker, and geographic area come from the 2000 US census and from the American Community Survey 5-year estimates. The geographic units of observation are ZIP Code Tabulation Areas (ZCTAs). We use two different definitions of income: "earnings" correspond essentially to labor income, whereas "total income" also includes capital income. As predicted by the theory, panel A shows that the correlation between population density and earnings per full-time worker, both measured in logs, increases from .12 for ZCTAs with below-median earnings per full-time worker to .31 for ZCTAs above the median. When we focus on total income in panel B, the results are similar, but now the correlation between population density and income per full-time worker is negative for ZCTAs below the median. In particular, the correlation increases from −.06 for those ZCTAs below the median to .36 for those above the median. All these correlations are statistically different from zero at the 1 percent level, and in both cases, the increase in the correlation is statistically significant at the 1 percent level. When we compare different quartiles, rather than below and above the median, the results continue to be consistent with the theory: the correlations between density and income are greater for higher-income quartiles.

Second, we explore whether the relation between density and income is stronger in richer areas by analyzing how that relation at the zip code level changes with the relative income of the Core Based Statistical Area (CBSA) the zip code pertains to. To analyze this, we start by running separate regressions for each CBSA of the log of payroll per employee on the log of employee density at the zip code level. Using data of 2010 from the ZIP Code Business Patterns, this yields 653 coefficients of employee density, one for each CBSA. Figure A1 then plots these 653 coefficients against the relative income of the CBSA. As expected, as the relative income of the CBSA increases, the effect of employee density on payroll per employee increases.

TABLE A1
DENSITY AND INCOME

	Percentiles Based on Mean Earnings					
Year	<25th	25th–50th	50th–75th	>75th	<Median	≥Median
	A. Correlation Population Density (Logs) and Mean Earnings per Full-Time Worker (Logs)					
2000	.0549***	.0884***	.1291***	.2199***	.1216***	.3128***
2007–11	−.0237*	.0428***	.1048***	.2727***	.0222***	.3546***
	B. Correlation Population Density (Logs) and Per Capita Income (Logs)					
2000	−.1001***	.0495***	.1499***	.2248***	−.0609***	.3589***
2007–11	−.0930***	.0175	.0733***	.2420***	−.0781***	.3234***

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

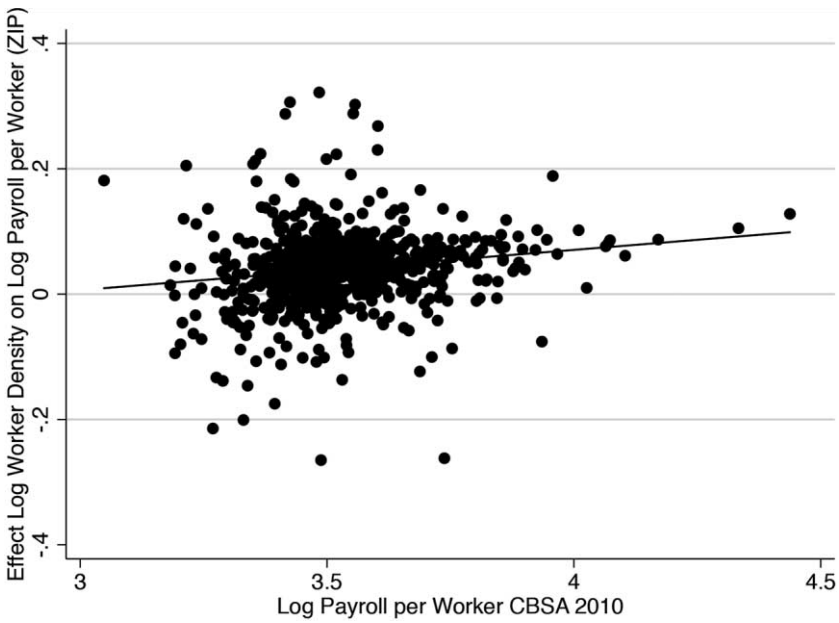

FIG. A1.—Payroll per employee and employee density

Rather than running separate regressions for each CBSA, we can also run a single regression with an interaction term of density at the zip code level with payroll per employee at the CBSA level:

$$y(\text{zip}) = a_0 + a_1 d(\text{zip}) + a_2 d(\text{zip}) y(\text{cbsa}) + \varepsilon(\text{zip}), \tag{A1}$$

where y(zip) is the log of payroll per employee at the zip code level, y(cbsa) is the log of payroll per employee at the CBSA level, and d(zip) is the log of employee density at the zip code level. The theory predicts that a_1 may be negative and that a_2 is positive. Consistent with this, when using data from 2010, we find $a_1 = -0.241$ and $a_2 = 0.079$, and both are statistically significant at the 1 percent level. Using data from 2000 yields similar results: $a_1 = -0.194$ and $a_2 = 0.075$, and once again both are statistically significant at the 1 percent level.

Comparing different countries and regions.—Next we analyze the correlation between real income per capita and population density across cells. Table A2 reports the results for different geographic subsets of the data. Across all cells of the world, the correlation between population density and real income per capita is −.41. Not surprisingly, within countries in which population mobility is higher, that correlation is positive and equal to .17. Of more interest is the comparison between richer and poorer cells. We start by splitting up the world into the 50 percent poorest cells and the 50 percent richest cells. As expected, the correlation between population density and income per capita is lower for the poorer cells (−.06) than for the richer cells (.11). A similar ranking emerges when contrasting the 50 percent poorest cells and the 50 percent richest cells within countries: the correlation is .16 for the poorer cells and .48 for the richer cells. Finally, we compare different continents and regions across the world. Once again, we would expect the correlation between density and income per capita to be higher in more developed regions. With the exception of Latin America, the results confirm this picture: we see the lowest correlation in Africa (−.11) and the highest correlations in North America (.50) and in Australia and New Zealand (.70).

TABLE A2
Correlation Population Density and Real Income per Capita across Cells

1. Across all cells of the world	−.41
2. Weighted within-country average	.17
3. Richest vs. poorest cells of the world:	
50% poorest cells	−.06
50% richest cells	.11
4. Richest vs. poorest cells within countries (weighted by number of cells per country):	
50% poorest cells	.16
50% richest cells	.48
5. Cell average within continents and regions:	
Africa	−.11
Asia	.06
Latin America and Caribbean	.18
Europe	.15
Of which western Europe	.33
North America	.50
Australia and New Zealand	.70

B. *Correlations of Amenities and Quality of Life*

Our goal is to compute correlations between our estimated measures of fundamental amenities and different exogenous measures of a location's quality of

life. The different variables on quality of life, related to distance to water, elevation, precipitation, temperature, and vegetation, come from G-Econ 4.0.[30] The results are reported in table A3.

TABLE A3
CORRELATIONS BETWEEN ESTIMATED AMENITIES AND DIFFERENT MEASURES OF QUALITY OF LIFE

	Correlations with Estimated Amenities (Logs)				
	All Cells (1)	United States (2)	One Cell per Country (3)	Placebo of Col. 1 (4)	Placebo of Col. 3 (5)
A. Water:					
Distance to ocean	−.3288***	.0544**	−.1102*	−.1277***	.3181***
Distance to water	−.4894***	−.3045***	−.2054***	−.3675***	.1638***
Water < 50 km	.2653***	.1731***	.1123*	.1442***	−.1428**
B. Elevation (logs):					
Level	−.4536***	−.2116***	−.2491***	−.3151***	.2694***
Standard deviation	−.4912***	−.3018***	−.2781***	−.3597***	.2023**
C. Precipitation:					
Average	.4301***	.1350***	.3860***	.3350***	.1016**
Maximum	.4462***	.1733***	.2383***	.4443***	.2992***
Minimum	.2279***	.2653***	.2150***	.0064	−.2613***
Standard deviation	.4126***	.0833***	.1969***	.4824***	.3954***
D. Temperature:					
Average	.5920***	.7836***	.1123*	.6832***	.4652***
Maximum	.5045***	.8141***	−.0498	.6449***	.4034***
Minimum	.5867***	.7029***	.1621***	.6529***	.4131***
Standard deviation	−.5455***	−.3953***	−.2438***	−.5576***	−.2983**
E. Vegetation:					
Desert, ice, or tundra	−.3553***	−.4412***	−.1771***	−.2775***	−.0919

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Column 1 presents correlations based on all 1° × 1° cells of the world. These correlations suggest that people prefer to live close to water, dislike high elevations and rough terrain, like precipitation but not constantly, prefer warm and stable temperatures, and dislike deserts, tundras, and ice-covered areas. It is reassuring that these correlations are largely consistent with those found in the literature. For example, Albouy et al. (2016) provide evidence that Americans have a mild preference for precipitation and a strong preference to live close to water, and Davis and Ortalo-Magné (2011) find that there is a positive correlation between quality of life and average temperature and precipitation. One thing that may come as a surprise is that people not only prefer higher average temperatures but also like higher maximum temperatures. This result is driven by the im-

[30] All variables come directly from G-Econ 4.0, but a couple require some further manipulation. In particular, the "distance to water" measure is defined as the minimum distance to a major navigable lake, a navigable river, or an ice-free ocean; the different vegetation categories in G-Econ can be found at http://data.giss.nasa.gov/landuse/vegeem.html.

posed linearity: once we allow for a quadratic relation between our measure of amenities and maximum temperatures, we find an optimal maximum temperature of 25.9° Celsius.

The fact that our correlations are overall in line with what we would expect suggests that our methodology of identifying amenities by using data on subjective well-being is reasonable. To further confirm this, we compare our results to what these correlations would look like within the United States. Because utility is the same across all cells within a country, the correlations within the United States do not depend on our use of data on subjective well-being. Hence, if the results for the United States are similar to those of the world, this is further evidence in favor of our methodology. Column 2 confirms that this is largely the case.

Another possible worry is that the correlations across all cells of the world are driven by a few large countries. If so, what we observe in column 1 may mostly reflect within-country variation, and this might explain why columns 1 and 2 look similar. To address this concern, we choose a random sample of 176 cells, one for each country, and compute the correlations. We repeat this procedure 5,000 times and report the resulting cross-country correlations in column 3. They look similar to those in columns 1 and 2.

As a further robustness check, we compute some placebo correlations to be compared with those in columns 1 and 3. To do so, we take a different value of ψ when transforming subjective well-being into our measure of utility. In particular, we assume that all countries in the world have the same utility when identifying fundamental amenities. The results are reported in columns 4 and 5. Note that there is no need to run a similar placebo test on column 2 since the correlations within countries are independent of our measure of utility. Columns 4 and 1 are similar. This is not surprising since most of the variation in those two columns is across cells within countries, and that variation is unaffected by the particular value of ψ. In contrast, when we focus on the variation across countries, using a different value of ψ should change the results. This is indeed what we find: column 5 yields correlations very different from those in column 3. All the signs on proximity to water and elevation have switched signs, and the correlation on minimum precipitation changes from positive to negative.

If the fundamental amenities that we have estimated are reasonable, then columns 1, 2, and 3 should yield similar results, columns 4 and 1 should not differ much, and columns 5 and 3 should be quite different. Our results in table A3 confirm these priors. Together with the fact that the correlations in column 1 are in line with those in the literature, this allows us to conclude that our particular methodology of using subjective well-being to identify fundamental amenities performs well.

C. *Subjective Well-Being*

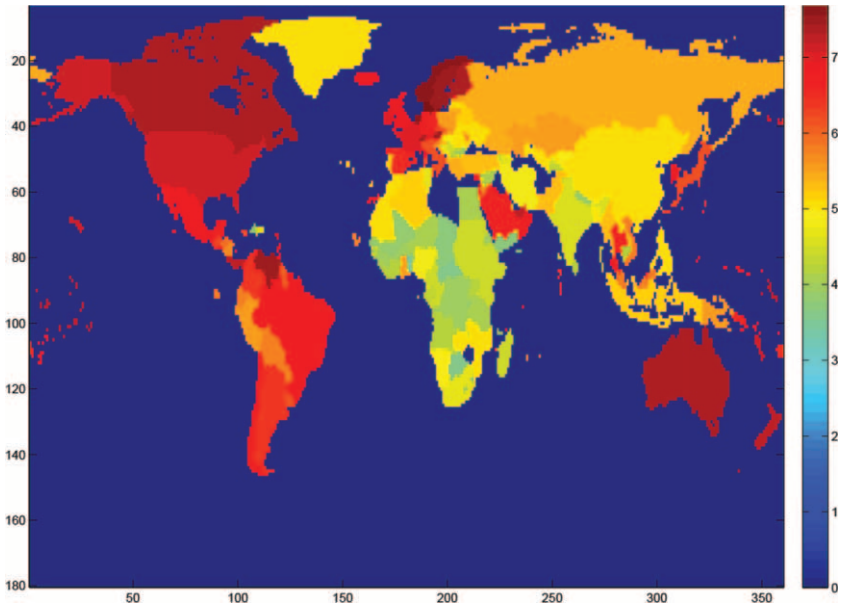

FIG. A2.—Cantril ladder measure of subjective well-being from the Gallup World Poll (max = 10, min = 0).

D. *Robustness Exercises*

We now conduct a number of robustness checks, related to different parameters. We study the effect of 20 percent changes in the parameters driving taste heterogeneity (Ω), static agglomeration economies (α), congestion in amenities (λ), the degree of technological diffusion ($1 - \gamma_2$), the spatial span of technological diffusion (η), and the level of trade costs (Υ). All exercises keep the values of initial amenities, productivities, and migration costs, as well as other parameters, as in the benchmark calibration. In addition to providing a quantitative assessment of the sensitivity of our findings to these different parameters, these robustness checks also contribute to our understanding of the workings of the model.

Preference heterogeneity.—In this exercise we increase Ω, the parameter that determines preference heterogeneity, from 0.5 to 0.6. The elasticity of migrant flows to real income drops from 2 to 1.67. As a result, people are less likely to move to high-income or high-amenity places. This lowers the degree of spatial concentration over time, so that growth rates go down in the long run (fig. A3). Relaxing mobility restrictions also has a smaller effect if people are less willing to move because of stronger locational preferences. Consistent with this, table A4 shows a welfare gain of 243.0 percent from free mobility, compared to 305.9 percent in the benchmark.

TABLE A4
THE IMPACT OF MOBILITY FRICTIONS

	Real Income (%)		Welfare (u) (%)[a]		Welfare $E[u\varepsilon^i]$ (%)[b]	
	$\vartheta = 1$	$\vartheta = 0$	$\vartheta = 1$	$\vartheta = 0$	$\vartheta = 1$	$\vartheta = 0$
1. Benchmark	0	125.8	0	305.9	0	79.0
2. Higher Ω: .6	.1	109.6	4.6	258.7	499.3	703.7
		109.4		243.0		34.1
3. Higher α: .072	2.8	132.9	2.6	319.9	2.3	84.4
		126.6		309.4		80.4
4. Lower λ: .256	1.0	176.6	142.8	1,476	115.6	459.3
		174.0		549.1		159.4
5. Higher $1 - \gamma_2$: $1 - .991$	16.9	156.9	18.7	396.7	21.9	128.2
		119.7		318.5		87.2
6. Local spatial diffusion	−6.1	117.7	−6.5	203.8	−7.8	32.5
		131.8		224.9		43.7
7. Trade costs + 20%	−17.3	97.0	−20.0	251.0	−23.7	47.2
		138.3		338.9		93.0
8. Trade costs + 40%	−30.1	79.5	−34.2	217.8	−39.3	26.1
		156.9		382.9		107.8

NOTE.—For each robustness exercise, the first line gives changes relative to the benchmark with $\vartheta = 1$, while the second line gives gains from fully liberalizing migration.

[a] Change in PDV of population-weighted average of cells' discounted utility levels, $u(r)$.

[b] Change of PDV of expected discounted utility levels as measured in eq. (4).

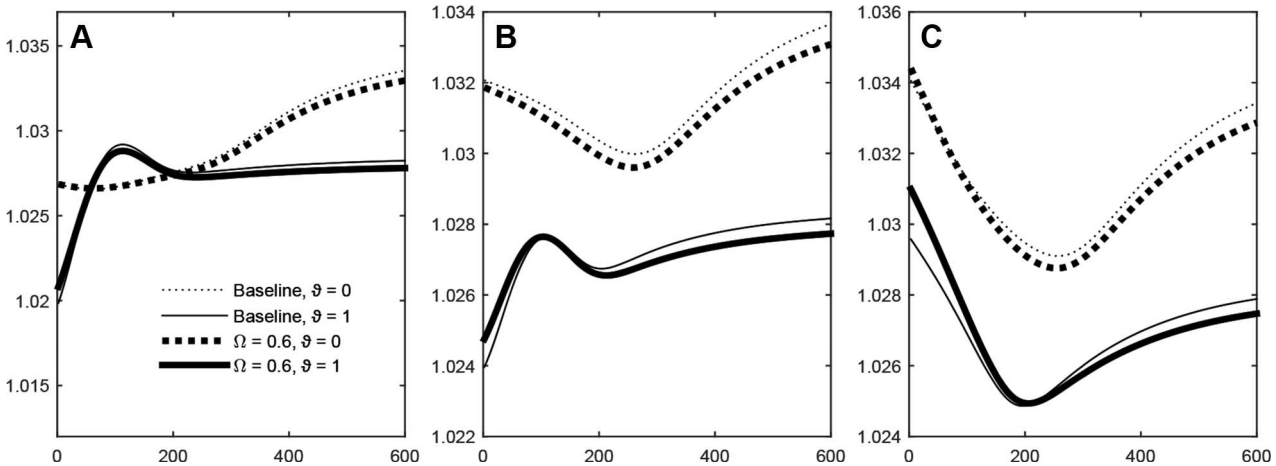

FIG. A3.—Growth rates of real GDP and utility for different values of Ω. *A*, Growth rate of real GDP. *B*, Growth rate of utility (u). *C*, Growth rate of $E[u\varepsilon^i]$.

Agglomeration economies.—Static agglomeration economies in our model are given by α/θ. In the benchmark, with $\alpha = 0.06$ and $\theta = 6.5$, the static elasticity of productivity to density was around 0.01. We now increase α by 20 percent to 0.072. As expected, compared to the benchmark case, the present discounted values of real income and welfare increase. We would expect this effect to be larger under free mobility because it is easier to benefit from increased agglomeration economies when people can freely move. This is indeed the case: table A4 shows that the gains in real income increase from 125.8 percent in the benchmark to 126.6 percent, and the welfare gains increase from 305.9 percent in the benchmark to 309.4 percent. Given that we increased the magnitude of static agglomeration economies by 20 percent, the effects may seem modest. Note, however, that dynamic agglomeration economies are more important in the long run than static ones.

Congestion in amenities.—We now reduce the elasticity of amenities to density by 20 percent. This drastically reduces congestion and hence makes it less costly to agglomerate. As a result, real income increases compared to the benchmark. The effects are larger under free mobility, when the possibility to concentrate in the most attractive places is enhanced. Greater concentration leads to faster growth, explaining why the increase in real GDP from free mobility rises from 125.8 percent in the benchmark to 174.0 percent (table A4). Not surprisingly, because the negative congestion effect on amenities is much lower, the welfare gains are in relative terms much greater. Free mobility now increases welfare by 549.1 percent.

Strength of technology diffusion.—In the model the degree of spatial diffusion of technology is given by $1 - \gamma_2$. In this exercise, we decrease γ_2 from 0.993 in the benchmark to 0.991. This amounts to an increase in the degree of spatial diffusion of 20 percent.[31] If best-practice technology diffuses faster, we would expect this to increase growth. As can be seen in figure A4, this is indeed the case, but only under current migration restrictions. In the case in which we eliminate moving costs, after some decades of faster growth, more diffusion actually lowers growth. The reason is that greater diffusion of technology lowers the incentive to concentrate in space. In the short run, the positive effect of more diffusion dominates, but in the long run the negative effect of less concentration and therefore less innovation dominates. This reversal underscores the importance of dynamics when thinking through the effect of more technological diffusion. In present discounted value (PDV) terms, the effect of greater diffusion is positive on real income and welfare gains. Since technological diffusion acts as a dispersion force, there is less congestion, so amenities improve. Stronger diffusion also implies that the economy takes less time to reach its BGP, as can be seen from figure A4.

[31] In order to eliminate the direct initial effect of the change in the level of technology diffusion, we adjust the value of the constant η to match the initial growth rate of real GDP of 2 percent that we use in the benchmark calibration. Including this direct effect decreases the visibility of the graphs but does not lead to large changes in the gains from mobility, which are 109.9 percent in terms of real GDP, 256.8 percent in terms of utility, and 61.5 percent in terms of expected utility.

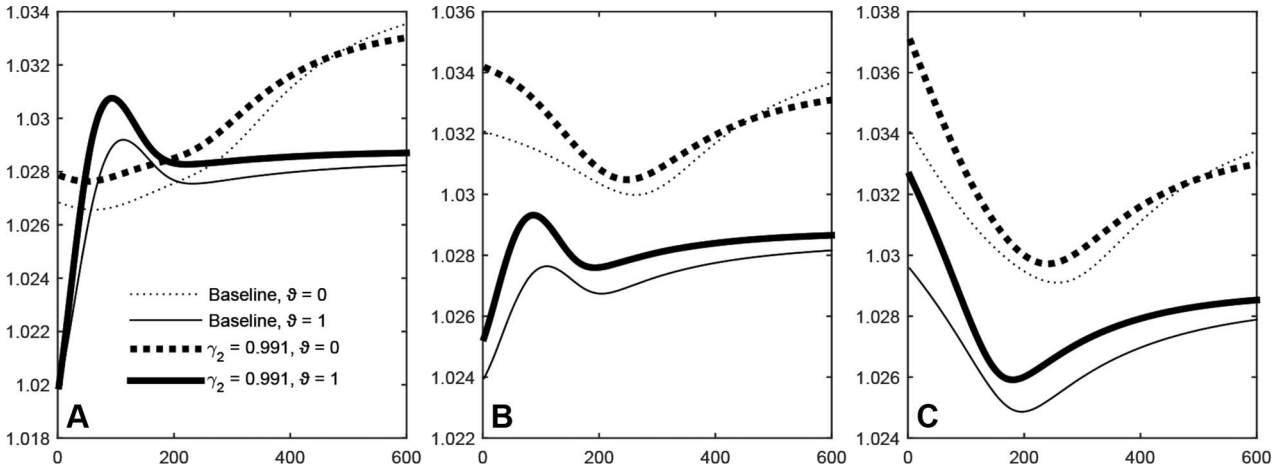

FIG. A4.—Growth rates of real GDP and utility for different values of γ_2. *A*, Growth rate of real GDP. *B*, Growth rate of utility (u). *C*, Growth rate of $E[u\varepsilon^t]$.

Spatial scope of technology diffusion.—Though technology diffusion in the model is slow, it diffuses to the whole world simultaneously. In this exercise, we explore what happens when we let technology diffuse locally rather than globally. In equation (8), we no longer take η to be constant, but instead assume that $\eta(r, s) = \bar{\eta} \exp(-\varkappa\delta(r, s))$, with $\varkappa > 0$ and $\delta(r, s)$ denoting the distance between locations r and s measured in kilometers. That is, location r obtains more technology spillovers from nearby locations. On the basis of the median of the distance decay parameters across different technologies in Comin, Dmitriev, and Rossi-Hansberg (2012), we use $\varkappa = 0.0015$. We choose the value of $\bar{\eta}$ such that the initial growth rate of the economy is 2 percent as in the benchmark calibration; in other words, we change the geographic distribution of technology diffusion, but not its overall level. We present the results in figure A5. In the short run less spatial diffusion lowers growth rates, but in the long run it increases growth rates. To understand figure A5 note that there are two effects at work. On the one hand, less spatial diffusion hurts some places, lowering growth. On the other hand, less spatial diffusion keeps economic activity more spatially concentrated, increasing innovation. In the short run, the first effect dominates, whereas in the long run the second one does. Overall, going from global to local diffusion decreases real output and welfare by about 7 percent as shown in table A4.

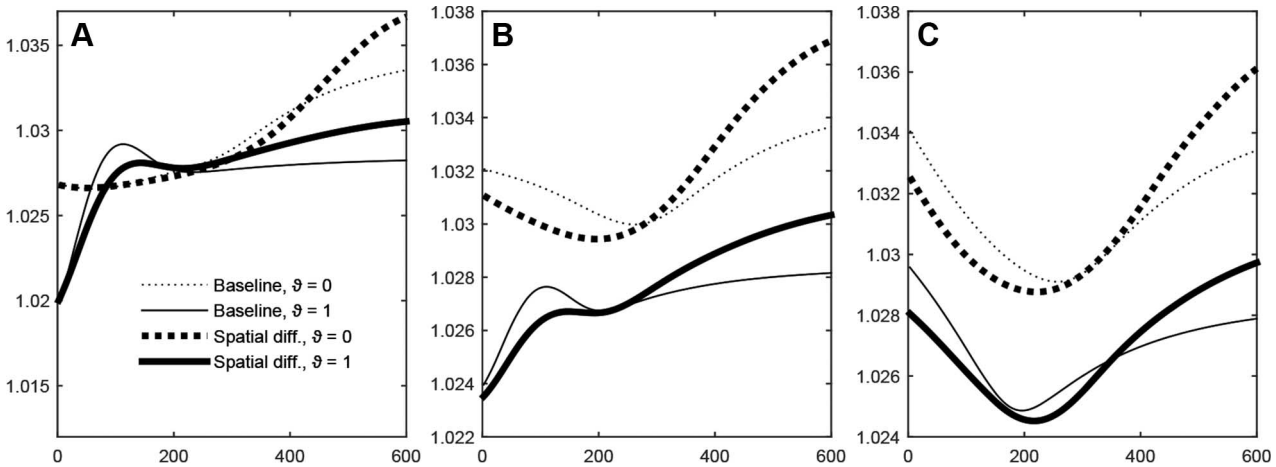

FIG. A5.—Growth rates of real GDP and utility with distance-dependent diffusion. *A*, Growth rate of real GDP. *B*, Growth rate of utility (u). *C*, Growth rate of $E[u\varepsilon^i]$.

In terms of the gains from liberalization, the effect on real income increases from 125.8 percent in the benchmark to 131.8 percent in the case with local diffusion. Greater movement of people makes it easier for the high-productivity places to attract more people, and this is more important when technology is more concentrated. Focusing on the welfare measures yields the opposite results; namely, the gains are smaller than in the benchmark. This result is intuitive: with less diffusion of technology, you get more people living in places they do not particularly like.[32]

Transportation costs.—As a last robustness check, we increase transportation costs, first by 20 percent and then by 40 percent. We discuss some of these results in the main text. As expected, the rise in transport costs reduces real income and welfare. Table A4 shows that real output declines by 17.3 percent with an increase in trade costs of 20 percent: a fairly large effect. When transport costs go up by 20 percent, the gains from liberalizing migration are higher than in the benchmark case. The reason is as follows: greater transportation costs increase the incentives to spatially concentrate. Lower migration costs facilitate this geographic concentration, hence giving a boost to real income. So trade and migration are substitutes. The results are similar, but even larger in magnitude, in the case in which transport costs go up by 40 percent.

Appendix B

Proofs, Derivations, and Other Details

A. *Derivation of Expected Utility*

To determine the expected utility of agents at r, $E[u(r)\varepsilon^i(r)|i \text{ lives at } r]$, we first derive the expected utility of agents who lived at u in period 0 and live at r in period t.[33] The cumulative density function of utility levels $u^i(u, r)$ conditional on originating from u and choosing to live at r is

[32] The PDVs for spatial diffusion are all calculated with a discount factor of $\beta = 0.9625$, since the long-run growth rate exceeds 3.5 percent in the case of no mobility restrictions (and so the PDV would not be well defined with $\beta = 0.965$).

[33] Throughout the proof, we omit the time index t for simplicity.

$$\begin{aligned}
&\Pr[u^i(u,r) \le \bar{u} | u^i(u,r) \ge u^i(u,s) \ \ \forall\ s \ne r] \\
&= \frac{\Pr[u^i(u,r) \le \bar{u},\ u^i(u,r) \ge u^i(u,s) \ \ \forall\ s \ne r]}{\Pr[u^i(u,r) \ge u^i(u,s) \ \ \forall\ s \ne r]} \\
&= \left[-\int_0^{\bar{u}} \prod_{S\setminus\{r\}} e^{-u(s)^{1/\Omega} m(u,s)^{-1/\Omega} z^{-1/\Omega} ds} \frac{1}{\Omega} u(r)^{1/\Omega} m(u,r)^{-1/\Omega} \right. \\
&\qquad \left. \times z^{-(1/\Omega+1)} e^{-u(r)^{1/\Omega} m(u,r)^{-1/\Omega} z^{-1/\Omega}} dz \right] \\
&\div \left\{ u(r)^{1/\Omega} m(u,r)^{-1/\Omega} \Big/ \left[\int_S u(s)^{1/\Omega} m(u,s)^{-1/\Omega} ds \right] \right\} \\
&= \int_0^{\bar{u}} -\frac{1}{\Omega} \left[\int_S u(s)^{1/\Omega} m(u,s)^{-1/\Omega} ds \right] \\
&\qquad \times z^{-(1/\Omega+1)} e^{-\left[\int_S u(s)^{1/\Omega} m(u,s)^{-1/\Omega} ds \right] z^{-1/\Omega}} dz \\
&= e^{-\left[\int_S u(s)^{1/\Omega} m(u,s)^{-1/\Omega} ds \right] z^{-1/\Omega}};
\end{aligned}$$

that is, $u^i(u, r)$ is distributed Fréchet, and therefore its mean is given by

$$\begin{aligned}
&E[u^i(u,r) | i \text{ originates from } u, \text{lives at } r] \\
&= \Gamma(1-\Omega) \left[\int_S u(s)^{1/\Omega} m(u,s)^{-1/\Omega} ds \right]^{\Omega}.
\end{aligned}$$

By equation (1), the expected value of $u(r)\varepsilon^i(r)$ equals this expression divided by the moving costs the agent has to pay,

$$\begin{aligned}
&E[u(r)\varepsilon^i(r) | i \text{ originates from } u, \text{lives at } r] \\
&= \frac{\Gamma(1-\Omega) \left[\int_S u(s)^{1/\Omega} m(u,s)^{-1/\Omega} ds \right]^{\Omega}}{\prod_{s=1}^{t} m(r_{s-1}, r_s)},
\end{aligned}$$

from which, using assumption 1, we get

$$\begin{aligned}
&E[u(r)\varepsilon^i(r) | i \text{ originates from } u, \text{ lives at } r] \\
&= \Gamma(1-\Omega) m_2(r) \left[\int_S u(s)^{1/\Omega} m_2(s)^{-1/\Omega} ds \right]^{\Omega};
\end{aligned}$$

since this expression does not depend on u, we also have

$$E[u(r)\varepsilon^i(r) | i \text{ lives at } r] = \Gamma(1-\Omega) m_2(r) \left[\int_S u(s)^{1/\Omega} m_2(s)^{-1/\Omega} ds \right]^{\Omega},$$

which is equation (4).

B. Derivation of Trade Shares and the Price Index

We derive here in detail first the probability that a given good produced in an area r is bought in s. The area $B(r, \delta)$ (a ball of radius δ centered at r) offers dif-

ferent goods in location s. The distribution of prices $B(r, \delta)$ offered in s is given by

$$G_t(p, s, B(r, \delta)) = 1 - e^{-\int_{B(r,\delta)} T_t(r)\left[\frac{mc_t(r)\varsigma(s,r)}{p}\right]^{-\theta} dr}.$$

The distribution of the prices of the goods s actually buys is then

$$G_t(p, s) = 1 - e^{-\int_S T_t(u)\left[\frac{mc_t(u)\varsigma(s,u)}{p}\right]^{-\theta} du} = 1 - e^{-\chi_t(s)p^{\theta}},$$

where

$$\chi_t(s) = \int_S T_t(u)[mc_t(u)\varsigma(s, u)]^{-\theta} du. \tag{B1}$$

Now we calculate the probability that a given good produced in an area $B(r, \delta)$ is bought in s. Start by computing the probability density that the price of the good produced in $B(r, \delta)$ and offered in s is equal to p and that this is the lowest price offered in s. This is

$$\frac{dG_t(p, s, B(r, \delta))}{dp} e^{-\int_{S\setminus B(r,\delta)} T_t(u)\left[\frac{mc_t(u)\varsigma(u,s)}{p}\right]^{-\theta} du}.$$

Rewrite

$$\begin{aligned} dG_t(p, s, B(r, \delta)) = {} & e^{-\int_{B(r,\delta)} T_t(r)\left[\frac{mc_t(r)\varsigma(s,r)}{p}\right]^{-\theta} dr} \\ & \times \int_{B(r,\delta)} T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta} dr \theta p^{\theta-1} dp. \end{aligned}$$

Replace this into the previous expression and integrate over all possible prices,

$$\int_0^{\infty} e^{-\int_S T_t(u)\left[\frac{mc_t(u)\varsigma(s,u)}{p}\right]^{-\theta} du} \int_{B(r,\delta)} T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta} dr \theta p^{\theta-1} dp.$$

Solving this integral yields

$$\left[-e^{-\chi_t(s)p^{\theta}} \frac{\int_{B(r,\delta)} T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta} dr}{\chi_t(s)} \right]_0^{\infty},$$

which gives

$$\begin{aligned} \pi_t(s, B(r, \delta)) &= \frac{\int_{B(r,\delta)} T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta} dr}{\int_S T_t(u)[mc_t(u)\varsigma(u, s)]^{-\theta} du} \\ &= \frac{\int_{B(r,\delta)} T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta} dr}{\chi_t(s)} \end{aligned}$$

for all $r, s \in S$. Since in any small interval with positive measure there will be many firms producing many goods, the above expressions can be interpreted as the fraction of goods location s buys from $B(r, \delta)$. In the limit, as $\delta \to 0$, this expression can be rewritten as

$$\pi_t(s, r) = \frac{T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta}}{\int_S T_t(u)[mc_t(u)\varsigma(u, s)]^{-\theta} du}$$
$$= \frac{T_t(r)[mc_t(r)\varsigma(r, s)]^{-\theta}}{\chi_t(s)} \quad \text{all } r, s \in S,$$

which is the expression in the text.

To finish this subsection, we derive the price index of a location s. We know that

$$P_t(s)^{\frac{-\rho}{1-\rho}} = \int_0^1 p_t^{\omega}(s)^{\frac{-\rho}{1-\rho}} d\omega.$$

This can be interpreted as the average price computed over the different goods that are being sold in s. This is the same as the expected value of the price of any good sold in s, so that

$$P_t(s)^{\frac{-\rho}{1-\rho}} = \int_0^{\infty} p^{\frac{-\rho}{1-\rho}} \frac{dG_t(s)}{dp} dp.$$

Since

$$\frac{dG_t(s)}{dp} = \chi_t(s)\theta p^{\theta-1} e^{-\chi_t(s)p^{\theta}},$$

we can write the previous expression as

$$P_t(s)^{\frac{-\rho}{1-\rho}} = \int_0^{\infty} p^{\frac{-\rho}{1-\rho}} \chi_t(s)\theta p^{\theta-1} e^{-\chi_t(s)p^{\theta}} dp.$$

By defining $x = \chi_t(s)p^{\theta}$, we can write $dx = \theta p^{\theta-1}\chi_t(s)$. Substituting yields

$$\int_0^{\infty} \left[\frac{x}{\chi_t(s)}\right]^{\frac{-\rho}{(1-\rho)\theta}} e^{-x} dx,$$

which is equal to

$$\chi_t(s)^{\frac{\rho}{(1-\rho)\theta}} \int_0^{\infty} x^{\frac{-\rho}{(1-\rho)\theta}} e^{-x} dx.$$

The Γ function is defined as $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy$, so that we can rewrite the above expression as

$$\chi_t(s)^{\frac{\rho}{(1-\rho)\theta}} \Gamma\left(\frac{-\rho}{(1-\rho)\theta} + 1\right).$$

Therefore,

$$P_t(s)^{\frac{-\rho}{1-\rho}} = \chi_t(s)^{\frac{\rho}{(1-\rho)\theta}} \Gamma\left(\frac{-\rho}{(1-\rho)\theta} + 1\right),$$

so that

$$P_t(s) = \chi_t(s)^{\frac{-1}{\theta}}\left[\Gamma\left(\frac{-\rho}{(1-\rho)\theta}+1\right)\right]^{-\frac{1-\rho}{\rho}},$$

which is the expression in the text.

C. *Proof of Lemma 2*

Substituting (20) and (B1) into (22), we obtain

$$u_t(r) = \bar{a}(r)\bar{L}_t(r)^{-\lambda}\frac{\xi}{\mu\xi+\gamma_1}\frac{w_t(r)}{\left[\int_S T_t(s)[mc_t(s)\varsigma(r,s)]^{-\theta}ds\right]^{\frac{-1}{\theta}}\bar{p}}$$

for any location r, where

$$\bar{p} = \left[\Gamma\left(\frac{-\rho}{(1-\rho)\theta}+1\right)\right]^{-\frac{1-\rho}{\rho}}.$$

We can rewrite this as

$$\bar{a}(r)^{-\theta}w_t(r)^{-\theta}\bar{L}_t(r)^{\lambda\theta} = \left[\frac{\mu\xi+\gamma_1}{\xi}\right]^{-\theta}(u_t(r)\bar{p})^{-\theta}\int_S T_t(s)[mc_t(s)\varsigma(r,s)]^{-\theta}ds,$$

from which

$$\begin{aligned}&\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\theta}w_t(r)^{-\theta}\bar{L}_t(r)^{\lambda\theta}\\ &= \kappa_1\int_S \tau_t(s)\varsigma(r,s)^{-\theta}w_t(s)^{-\theta}\bar{L}_t(s)^{\alpha-\left(1-\mu-\frac{\gamma_1}{\xi}\right)\theta}ds,\end{aligned} \tag{B2}$$

where

$$\kappa_1 = \left[\frac{\mu\xi+\gamma_1}{\xi}\right]^{-\left[\mu+\frac{\gamma_1}{\xi}\right]\theta}\mu^{\mu\theta}\left[\frac{\xi\nu}{\gamma_1}\right]^{-\frac{\gamma_1\theta}{\xi}}\bar{p}^{-\theta}.$$

This yields the first set of equations that $u_t(\cdot)$, $\bar{L}_t(\cdot)$, and $w_t(\cdot)$ have to solve.

Inserting (19) and (20) into the balanced trade condition, we get

$$w_t(r)H(r)\bar{L}_t(r) = \bar{p}^{-\theta}\int_S T_t(r)[mc_t(r)\varsigma(r,s)]^{-\theta}P_t(s)^{\theta}w_t(s)H(s)\bar{L}_t(s)ds.$$

Substituting (22) and $T_t(r) = \tau_t(r)\bar{L}_t(r)^{\alpha}$ into the previous equation yields

$$\begin{aligned}&\tau_t(r)^{-1}H(r)w_t(r)^{1+\theta}\bar{L}_t(r)^{1-\alpha+\left(1-\mu-\frac{\gamma_1}{\xi}\right)\theta}\\ &= \kappa_1\int_S\left[\frac{\bar{a}(s)}{u_t(s)}\right]^{\theta}H(s)\varsigma(s,r)^{-\theta}w_t(s)^{1+\theta}\bar{L}_t(s)^{1-\lambda\theta}ds.\end{aligned} \tag{B3}$$

This gives the second set of equations that $u_t(\cdot)$, $\bar{L}_t(\cdot)$, and $w_t(\cdot)$ have to solve. The third set of equations that $u_t(\cdot)$, $\bar{L}_t(\cdot)$, and $w_t(\cdot)$ have to solve is given by (7). Clearly, $\tau_t(\cdot)$ is obtained directly from (8) and $\bar{L}_{t-1}(\cdot)$.

We can use (B2) and (B3) to analyze the equilibrium allocation further under symmetric trade costs. In doing so, we follow the proof of theorem 2 in Allen and Arkolakis (2014). Assume trade costs are symmetric so that $\varsigma(r, s) = \varsigma(s, r)$. Introduce the following function, which is the ratio of the left-hand sides of (B1) and (B2):

$$f_1(r) = \frac{\tau_t(r)^{-1} H(r) w_t(r)^{1+\theta} \bar{L}_t(r)^{1-\alpha+\left(1-\mu-\frac{\gamma_1}{\xi}\right)\theta}}{\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\theta} w_t(r)^{-\theta} \bar{L}_t(r)^{\lambda\theta}}.$$

Obviously, $f_1(r)$ equals the ratio of the right-hand sides, that is,

$$f_1(r) = \frac{\int_S \left[\frac{\bar{a}(s)}{u_t(s)}\right]^{\theta} H(s) w_t(s)^{1+\theta} \bar{L}_t(s)^{1-\lambda\theta} \varsigma(s, r)^{-\theta} ds}{\int_S \tau_t(s) w_t(s)^{-\theta} \bar{L}_t(s)^{\alpha-\left(1-\mu-\frac{\gamma_1}{\xi}\right)\theta} \varsigma(r, s)^{-\theta} ds},$$

from which, using $\varsigma(r, s) = \varsigma(s, r)$, we obtain

$$f_1(r) = \frac{\int_S f_1(s)^{-\lambda} f_2(s, r) ds}{\int_S f_1(s)^{-(1+\lambda)} f_2(s, r) ds}, \tag{B4}$$

where

$$f_2(s, r) = \frac{\tau_t(s)^{-\lambda}}{\varsigma(s, r)^{\theta}} \left[\frac{\bar{a}(s)}{u_t(s)}\right]^{(1+\lambda)\theta} H(s)^{1+\lambda} w_t(s)^{1+\theta+(1+2\theta)\lambda} \times \bar{L}_t(s)^{1-\lambda\theta-\lambda\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta\right]}.$$

Write (B4) as

$$f_3(r) = \frac{f_1(r)^{-\lambda}}{\int_S f_1(s)^{-\lambda} f_2(s, r) ds} = \frac{f_1(r)^{-(1+\lambda)}}{\int_S f_1(s)^{-(1+\lambda)} f_2(s, r) ds}. \tag{B5}$$

Then, using the notation

$$g_1(r) = f_1(r)^{-\lambda}$$

and

$$g_2(r) = f_1(r)^{-(1+\lambda)},$$

one can write this last equation as

$$g_1(r) = \int_S f_3(r) f_2(s, r) g_1(s) ds \tag{B6}$$

and

$$g_2(r) = \int_S f_3(r) f_2(s, r) g_2(s) ds. \tag{B7}$$

Define $K(s, r)$ as the value of $f_3(r)f_2(s, r)$ evaluated at the solution g_1. By (B5), the value of $f_3(r)f_2(s, r)$ evaluated at g_2 is the same for any pair of locations (r, s). Therefore, g_1 and g_2 are both solutions to the integral equation

$$x(r) = \int_S K(s, r)x(s)ds. \tag{B8}$$

Note that $K(s, r)$ is nonnegative, continuous, and square-integrable. Nonnegativity of $K(\cdot, \cdot)$ immediately follows from the nonnegativity of f_2 and f_3. To see measurability, recall that $\bar{a}(r)$, $H(r)$, and $\tau_0(r)$ are assumed to be continuous functions. We also need to assume that $u_0(r)$, $w_0(r)$, and $\bar{L}_0(r)$ are continuous; otherwise, the integrals on the right-hand sides of (B2) and (B3) would not be well defined. Once this is the case, $\tau_1(r)$ is also continuous by (8); hence so are $u_1(r)$, $w_1(r)$, and $\bar{L}_1(r)$. Using this logic, one can show that $\tau_1(r)$, $u_t(r)$, $w_t(r)$, and $\bar{L}_t(r)$ are continuous for any t. Thus, f_1, f_2, and f_3 are all continuous, which implies that $K(\cdot, \cdot)$ is continuous.[34] Square-integrability, which means

$$\int_S \int_S K(s, r)^2 dsdr < \infty,$$

is due to the fact that S is bounded by assumption, but so is $K(\cdot, \cdot)$. To see why $K(\cdot, \cdot)$ is bounded, note that population at a location cannot shrink to zero unless the location offers zero nominal wages that compensate for its infinitely good amenities. Also, population at a location cannot be larger than $\bar{L}$. These upper and lower bounds on population translate into upper and lower bounds on $f_2(\cdot, \cdot)$ and $f_3(\cdot)$ and hence on $K(\cdot, \cdot)$.

Given the above properties of $K(\cdot, \cdot)$, theorem 2.1 in Zabreyko et al. (1975) guarantees that the solution to (B8) exists and is unique up to a scalar multiple. Hence

$$g_1(r) = \varpi g_2(r),$$

where ϖ is a constant. Therefore, we have

$$f_1(r)^{-\lambda} = \varpi f_1(r)^{-(1+\lambda)},$$

from which

$$f_1(r) = \varpi.$$

That is,

$$\frac{\tau_t(r)^{-1} H(r) w_t(r)^{1+\theta} \bar{L}_t(r)^{1-\alpha+\left(1-\mu-\frac{\gamma_1}{\xi}\right)\theta}}{\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\theta} w_t(r)^{-\theta} \bar{L}_t(r)^{\lambda\theta}} = \varpi;$$

thus

[34] A rigorous proof of the measurability of $K(\cdot, \cdot)$ requires some further steps that we have not included in this draft but are available on request. We acknowledge the help of Áron Tóbiás in formulating this more rigorous proof.

$$w_t(r) = \varpi^{\frac{1}{1+2\theta}}\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\frac{\theta}{1+2\theta}}\tau_t(r)^{\frac{1}{1+2\theta}}H(r)^{-\frac{1}{1+2\theta}}\bar{L}_t(r)^{\frac{\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta}{1+2\theta}},$$

which is the same as equation (23), with $\bar{w}$ defined as $\bar{w} = \varpi^{1/(1+2\theta)}$.

Substituting this into (B2) yields the second equation in lemma 2:

$$\begin{aligned}\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}}&\tau_t(r)^{-\frac{\theta}{1+2\theta}}H(r)^{\frac{\theta}{1+2\theta}}\bar{L}_t(r)^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta\right]}\\ &= \kappa_1\int_S\left[\frac{\bar{a}(s)}{u_t(s)}\right]^{\frac{\theta^2}{1+2\theta}}\tau_t(s)^{\frac{1+\theta}{1+2\theta}}H(s)^{\frac{\theta}{1+2\theta}}\varsigma(r,s)^{-\theta}\\ &\quad\times\bar{L}_t(s)^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta\right]}ds.\end{aligned} \tag{B9}$$

QED

D. Proof of Lemma 3

Substituting (7) in (24) yields

$$\begin{aligned}&B_{1t}(r)\hat{u}_t(r)^{\frac{1}{\Omega}\left[\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]+\frac{\theta(1+\theta)}{1+2\theta}}\\ &= \kappa_1\int_S\hat{u}_t(s)^{\frac{1}{\Omega}\left[1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]-\frac{\theta^2}{1+2\theta}}B_{2t}(s)\varsigma(r,s)^{-\theta}ds,\end{aligned} \tag{B10}$$

where

$$\begin{aligned}B_{1t}(r) &= \bar{a}(r)^{-\frac{\theta(1+\theta)}{1+2\theta}}\tau_t(r)^{-\frac{\theta}{1+2\theta}}H(r)^{\frac{\theta}{1+2\theta}\left[\alpha+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]-\lambda\theta}\\ &\quad\times m_2(r)^{-\frac{1}{\Omega}\left[\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]}\end{aligned}$$

and

$$\begin{aligned}B_{2t}(r,s) &= \bar{a}(s)^{\frac{\theta^2}{1+2\theta}}\tau_t(s)^{\frac{1+\theta}{1+2\theta}}H(s)^{\frac{\theta}{1+2\theta}-1+\lambda\theta-\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]}\\ &\quad\times m_2(s)^{-\frac{1}{\Omega}\left[1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]}\varsigma(r,s)^{-\theta}\end{aligned}$$

are exogenously given functions, and

$$\hat{u}_t(r) = u_t(r)\left[\frac{\bar{L}}{\int_S u_t(v)^{1/\Omega}m_2(v)^{-1/\Omega}dv}\right]^{\Omega\left[1-\frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda+(1-\mu)-\frac{\gamma_1}{\xi}\right]\theta-\alpha\right]+\theta}\right]}. \tag{B11}$$

First we prove that a solution to (B10) exists and is unique if $\alpha/\theta + \gamma_1/\xi \leq \Omega + \lambda + 1 - \mu$. Equations (B10) constitute a system of equations that pin down $\hat{u}_t(r)$. It follows from theorem 2.19 in Zabreyko et al. (1975) that the solution $f(\cdot)$ to equation

$$B_1(r)f(r)^{\tilde{\gamma}_1} = \kappa_1\int_S B_2(r,s)f(s)^{\tilde{\gamma}_2}ds$$

exists and is unique if (i) the function $\kappa_1 B_1(r)^{-1} B_2(r, s)$ is strictly positive and continuous, which is a direct consequence of assumption 3, and (ii) $|\tilde{\gamma}_2/\tilde{\gamma}_1| \leq 1$; that is, the ratio of exponents on the right-hand side and on the left-hand side is not larger than one in absolute value. In the case of equation (B10), this condition implies

$$\frac{\frac{1}{\Omega}\left[1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta\right]\right] - \frac{\theta^2}{1+2\theta}}{\frac{1}{\Omega}\left[\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta\right]\right] + \frac{\theta(1+\theta)}{1+2\theta}} \leq 1,$$

from which, by arranging, we get

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} \leq \lambda + 1 - \mu + \Omega.$$

Theorem 2.19 in Zabreyko et al. (1975) also implies that we can solve equation (B10) using the following iterative procedure. Guess an initial distribution, $\hat{u}_t^0(\cdot)$. Plug it into the right-hand side of equation (B10), calculate the left-hand side, and solve for the distribution of utility; call this $\hat{u}_t^1(\cdot)$. Then compute the distance between $\hat{u}_t^1(\cdot)$ and $\hat{u}_t^0(\cdot)$, defined as

$$\text{dist}_t^1 = \int_S \left[\hat{u}_t^1(r) - \hat{u}_t^0(r)\right]^2 dr.$$

If $\text{dist}_t^1 < \varepsilon$, where ε is an exogenously given tolerance level, stop. Otherwise, plug $\hat{u}_t^1(\cdot)$ into the right-hand side of (B10), obtain the updated distribution $\hat{u}_t^2(r)$, and compute dist_t^2, defined analogously to dist_t^1. Continue the procedure until $\text{dist}_t^i < \varepsilon$ for some i.

To obtain $u_t(\cdot)$ from $\hat{u}_t(\cdot)$, we write equation (B11) as

$$u_t(r) = \frac{\hat{u}_t(r)}{\hat{U}_t^F}, \tag{B12}$$

where

$$\hat{U}_t = \frac{\bar{L}}{\int_S m_2(v)^{-1/\Omega} u_t(v)^{1/\Omega} dv} \tag{B13}$$

is independent of r and

$$\frac{F}{\Omega} = 1 - \frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda + (1-\mu) - \frac{\gamma_1}{\xi}\right]\theta - \alpha\right] + \theta}.$$

Plugging (B12) into (B13) and rearranging yields

$$\hat{U}_t^{1-\frac{F}{\Omega}} = \frac{\bar{L}}{\int_S m_2(v)^{-1/\Omega} \hat{u}_t(v)^{1/\Omega} dv}.$$

Therefore, the value of $\hat{U}_t$ can be uniquely expressed as

$$\hat{U}_t = \left[\frac{\bar{L}}{\int_S m_2(v)^{-1/\Omega} u_t(v)^{1/\Omega} dv}\right]^{\frac{1}{1-\frac{F}{\Omega}}},$$

as long as $F/\Omega \neq 1$ which is guaranteed by $\theta > 0$. Hence, under the stated parameter restriction the value of $u_t(r)$ can be uniquely expressed from equation (B12). With unique values of $u_t(r)$ in hand, we can simply use equations (7) to obtain unique population levels $\bar{L}_t(s)$ and equations (23) to obtain unique wages, $w_t(r)$ for all $r \in S$. QED

E. *Proof of Lemma 4*

Given the evolution of technology in (8), the growth rate of $\tau_t(r)$ is given by

$$\frac{\tau_{t+1}(r)}{\tau_t(r)} = \phi_t(r)^{\theta\gamma_1}\left[\int_S \eta \frac{\tau_t(s)}{\tau_t(r)} ds\right]^{1-\gamma_2}.$$

Divide both sides of the equation by the corresponding equation for location s, and rearrange to get

$$\frac{\tau_t(s)}{\tau_t(r)} = \left[\frac{\phi(s)}{\phi(r)}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[\frac{\bar{L}(s)}{\bar{L}(r)}\right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}},$$

where the second equality follows from (12), and where we drop the time subscript to indicate that we refer to a variable that remains constant in the BGP. Use this relationship to obtain that

$$\bar{L}(s) = \left[\frac{\tau_t(s)}{\tau_t(r)}\right]^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} \bar{L}(r),$$

and so, after integrating over s and using the labor market clearing condition, that

$$\int_S H(s)\bar{L}(s)ds = \bar{L} = \bar{L}(r)\tau_t(r)^{-\frac{(1-\gamma_2)\xi}{\theta\gamma_1}}\int_S H(s)\tau_t(s)^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} ds$$

or

$$\tau_t(r) = \kappa_t^2 L(r)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}, \tag{B14}$$

where κ_t^2 depends on time but not on location. Substituting (B14) into (24) implies that

$$\begin{aligned}\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} & H(r)^{\frac{\theta}{1+2\theta}}\bar{L}(r)^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta+\frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} \\ &= \kappa_1\kappa_t^2\int_S\left[\frac{\bar{a}(s)}{u_t(s)}\right]^{\frac{\theta^2}{1+2\theta}} H(s)^{\frac{\theta}{1+2\theta}}\varsigma(r,s)^{-\theta} \\ &\quad\times \bar{L}(s)^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta+\frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} ds.\end{aligned} \tag{B15}$$

Substituting equation (7) in (B15) yields

$$\left[\frac{\bar{a}(r)}{u_t(r)}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} H(r)^{\frac{\theta}{1+2\theta}} \left[\frac{u_t(r)^{1/\Omega} m_2(r)^{-1/\Omega}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv} \frac{\bar{L}}{H(r)}\right]^{\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]}$$

$$= \kappa_1 \kappa_t^2 \int_S \left[\frac{\bar{a}(s)}{u_t(s)}\right]^{\frac{\theta^2}{1+2\theta}} H(s)^{\frac{\theta}{1+2\theta}} \varsigma(r, s)^{-\theta} \tag{B16}$$

$$\times \left[\frac{u_t(s)^{1/\Omega} m_2(s)^{-1/\Omega}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv} \frac{\bar{L}}{H(s)}\right]^{1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} ds,$$

where time subscripts have been dropped for variables that do not change in the BGP. Rearranging (B16) yields

$$\begin{aligned} & B_1(r)\hat{u}_t(r)^{\frac{1}{\Omega}\left[\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] + \frac{\theta(1+\theta)}{1+2\theta}} \\ = \; & \kappa_1 \kappa_t^2 \int_S \hat{u}_t(s)^{\frac{1}{\Omega}\left[1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] - \frac{\theta^2}{1+2\theta}} B_2(s)\varsigma(r, s)^{-\theta} ds, \end{aligned} \tag{B17}$$

where

$$\begin{aligned} B_1(r) = \; & \bar{a}(r)^{-\frac{\theta(1+\theta)}{1+2\theta}} H(r)^{\frac{\theta}{1+2\theta}\left[\alpha + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right] - \lambda\theta} \\ & \times m_2(r)^{-\frac{1}{\Omega}\left[\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right]} \end{aligned}$$

and

$$\begin{aligned} B_2(s) = \; & \bar{a}(s)^{\frac{\theta^2}{1+2\theta}} H(s)^{\frac{\theta}{1+2\theta} - 1 + \lambda\theta - \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} \\ & \times m_2(s)^{-\frac{1}{\Omega}\left[1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right]} \end{aligned}$$

are exogenously given functions, and

$$\hat{u}_t(r) = u_t(r)\tilde{U}_t, \tag{B18}$$

where

$$\tilde{U}_t = \left[\frac{\bar{L}}{\int_S u_t(v)^{1/\Omega} m_2(v)^{-1/\Omega} dv}\right]^{\Omega\left[1 - \frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda + (1-\mu) - \frac{\gamma_1}{\xi}\right]\theta - \alpha - \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right] + \theta}\right]}.$$

Equations (B17) constitute a system of equations that pin down $\hat{u}_t(r)$. It follows from theorem 2.19 in Zabreyko et al. (1975) that the solution $f(\cdot)$ to equation

$$B_1(r)f(r)^{\tilde{\gamma}_1} = B_3\int_S B_2(s)f(s)^{\tilde{\gamma}_2}ds$$

exists and is unique if $|\tilde{\gamma}_2/\tilde{\gamma}_1| \leq 1$; that is, the ratio of exponents on the right-hand side and on the left-hand side is not larger than one in absolute value. In the case of equation (B17), this condition implies

$$\frac{\frac{1}{\Omega}\left[1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] - \frac{\theta^2}{1+2\theta}}{\frac{1}{\Omega}\left[\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] + \frac{\theta(1+\theta)}{1+2\theta}} \leq 1,$$

from which, by arranging, we get

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} + \frac{\gamma_1}{(1-\gamma_2)\xi} \leq \lambda + 1 - \mu + \Omega.$$

Once we have found a solution $\hat{u}_t(r)$, the rest of the equilibrium calculation proceeds as in lemma 3. QED

F. *Proof of Lemma 5*

Lemma 4 guarantees that if the economy is in its BGP in periods t and $t+1$, then

$$\frac{u_{t+1}(r)}{u_t(r)} = \left[\frac{\kappa_{t+1}^2}{\kappa_t^2}\right]^{\frac{1}{\theta}} = \left[\frac{\tau_{t+1}(r)}{\tau_t(r)}\right]^{\frac{1}{\theta}}$$

for all r; so since

$$\frac{\tau_{t+1}(r)}{\tau_t(r)} = \phi(r)^{\theta\gamma_1}\left[\int_S \eta\frac{\tau_t(s)}{\tau_t(r)}ds\right]^{1-\gamma_2} = \phi(r)^{\theta\gamma_1}\left[\int_S \eta\left[\frac{\phi(s)}{\phi(r)}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}}ds\right]^{1-\gamma_2}$$
$$= \eta^{1-\gamma_2}\left[\frac{\gamma_1/\nu}{\gamma_1 + \mu\xi}\right]^{\frac{\theta\gamma_1}{\xi}}\left[\int_S \bar{L}(s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}}ds\right]^{1-\gamma_2},$$

we have that

$$\frac{u_{t+1}(r)}{u_t(r)} = \eta^{\frac{1-\gamma_2}{\theta}}\left[\frac{\gamma_1/\nu}{\gamma_1 + \mu\xi}\right]^{\frac{\gamma_1}{\xi}}\left[\int_S \bar{L}(s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}}ds\right]^{\frac{1-\gamma_2}{\theta}}.$$

QED

G. *Procedure to Find a Solution to Equation (32)*

We use the following procedure to solve equation (32). We approximate (32) by

$$w_0(r)^{-\theta}\bar{L}_0(r)^{\lambda\theta}\left[\frac{\bar{a}(r)}{u_0(r)}\right]^{-\theta} = \kappa_1\bar{w}^{-[1+2\theta]}\int_S w_0(s)^{1+\theta}\bar{L}_0(s)^{1-\lambda\theta}H(s)\varsigma(r,s)^{-\theta}\left[\frac{\bar{a}(s)}{u_0(s)}\right]^{\theta-\epsilon^0}ds, \tag{B19}$$

where $\epsilon^0 > 0$ is a constant. For any positive ϵ^0, theorem 2.19 in Zabreyko et al. (1975) guarantees that (B19) can be solved by the following simple iterative procedure. Guess some initial distribution $[\bar{a}(r)/u_0(r)]^0_{\epsilon^0}$, plug it into the right-hand side of equation (B19), calculate the left-hand side, and solve for $\bar{a}(r)/u_0(r)$; call this $[\bar{a}(r)/u_0(r)]^1_{\epsilon^0}$. Compute the distance between $[\bar{a}(r)/u_0(r)]^1_{\epsilon^0}$ and $[\bar{a}(r)/u_0(r)]^0_{\epsilon^0}$, defined as

$$\text{dist}^1 = \int_S \left[\left[\frac{\bar{a}(r)}{u_0(r)} \right]^1_{\epsilon^0} - \left[\frac{\bar{a}(r)}{u_0(r)} \right]^0_{\epsilon^0} \right]^2 dr.$$

If $\text{dist}^1 < \varepsilon$, where ε is an exogenously given tolerance level, stop. Otherwise, plug $[\bar{a}(r)/u_0(r)]^1_{\epsilon^0}$ into the right-hand side of (B19), express $\bar{a}(r)/u_0(r)$ from the left-hand side, that is, obtain $[\bar{a}(r)/u_0(r)]^2_{\epsilon^0}$, and compute dist^2, defined analogously to dist^1. Continue the procedure until $\text{dist}^i < \varepsilon$ for some i.

Having the solution to (B19), $[\bar{a}(r)/u_0(r)]_{\epsilon^0}$, one needs to check whether it is sufficiently close to the solution of (32). Given that the system of (24) and (23) is equivalent to the system of (32) and (31), we can check this in the following way. First, we solve for $\tau_0(r)_{\epsilon^0}$ using $[\bar{a}(r)/u_0(r)]_{\epsilon^0}$ and equation (31). Next, we plug $[\bar{a}(r)/u_0(r)]_{\epsilon^0}$ and $\tau_0(r)_{\epsilon^0}$ into (24) and solve for population levels $\bar{L}(r)_{\epsilon^0}$.[35] Finally, we check whether $\bar{L}_0(r)_{\epsilon^0}$ are sufficiently close to $\bar{L}_0(r)$, the population levels seen in the data.[36] If they are not close enough, we take $\epsilon^1 = \epsilon^0/2$ and redo the whole procedure with ϵ^1 instead of ϵ^0. Now, if $\bar{L}_0(r)_{\epsilon^1}$ are sufficiently close to $\bar{L}_0(r)$, we stop. Otherwise, we proceed with $\epsilon^2 = \epsilon^1/2$; and so on. This procedure results in a pair of distributions $[\bar{a}(r)/u_0(r)]_{\epsilon^i}$ and $\tau_0(r)_{\epsilon^i}$, which are sufficiently close to the solution of (32) and (31).

Appendix C

Data Appendix

The data description and sources follow approximately the order in which they appear in the paper. For the numerical exercise, all data are essentially for the time period 2000–2010, whereas for the calibration of the parameters we sometimes use data from a longer time period.

Population and amenities of US metropolitan statistical areas.—Data on population of US MSAs come from the Bureau of Economic Analysis, Regional Economic Accounts, and are for the years 2005 and 2008. Corresponding data on amenities are estimated by the structural model in Desmet and Rossi-Hansberg (2013).

Population, GDP, and wages at 1° × 1° resolution.—Data on population and GDP for 1° × 1° cells for the entire world come from the G-Econ 4.0 research project at Yale University. For the calibration of the technology parameters, we use data of

[35] This is done using the procedure explained in Sec. IV.F.

[36] The criterion used here is

$$\widetilde{\text{dist}}^0 = \int_S [\bar{L}_0(r)_{\epsilon^0} - \bar{L}_0(r)]^2 dr < \tilde{\varepsilon},$$

where $\tilde{\varepsilon}$ is a positive tolerance level.

1990, 1995, 2000, and 2005 and take GDP measured in purchasing power parity (PPP) terms. For the benchmark model, we use data of 2000. Wages are taken to be GDP in PPP terms divided by population.

Railroads, major roads, other roads, and waterways at 1° × 1° resolution.—Data on railroads, major roads, and other roads come from http://www.naturalearthdata.com/, a public-domain map data set built through the collaborative effort of many volunteers and supported by the North American Cartographic Information Society. For each 1° × 1° cell, we define railroads as the share of 0.1° × 0.1° cells through which a railroad passes. Major roads, other roads, and waterways are defined analogously. Major roads refer to a major highway, a beltway, or a bypass; other roads refer to any other type of road; and waterways refer to either a river or an ocean.

Land at 1° × 1° and 30″ × 30″ resolution.—Data on land come from the Global Land One-km Base Elevation (GLOBE) digital elevation model (DEM), a raster elevation data set from the National Oceanic and Atmospheric Administration covering all 30″ × 30″ (arcsecond) cells that are located on land. Using this information, we compute for each 1° × 1° cell on the globe the share of the 30″ × 30″ cells that are on land.

Subjective well-being at country level.—Subjective well-being is measured on a Cantril ladder from 0 to 10, where 0 represents the worst possible life and 10 the best possible life the individual can contemplate for himself. The main data source is the Gallup World Poll, and we take the mean for the period 2007–11 as reported in the Human Development Report 2013. This gives us data on 151 countries. To increase the coverage of countries, we proceed in three ways. First, we use the 2013 World Database of Happiness by R. Veenhoven (http://worldatabaseofhappiness.eur.nl), who reports data for a similar time period on the same evaluative measure of subjective well-being from Gallup, the Pew Research Center, and Latinobarómetro. This gives an additional seven countries. Second, Abdallah, Thompson, and Marks (2008) propose a model to estimate subjective measures of well-being for countries that are typically not surveyed. To make the data comparable, we normalize the Abdallah et al. measure so that the countries that are common with those surveyed by Gallup have the same means and standard deviations. This increases the coverage to 184 countries. Third, for the rest of the world—mostly small islands and territories—we assign subjective well-being measures in an ad hoc manner. For example, for Andorra we take the average of France and Spain. For another example, in the case of Nauru, a small island in the Pacific, we take the average of the Solomon Islands and Vanuatu.

Population at 30″ × 30″ resolution.—Population data at the same resolution of 30″ × 30″ come from LandScan 2005 at the Oak Ridge National Laboratory.

US zip code and CBSA data.—Data on area, population, mean earnings per worker, and income per capita come from the US census. Data referred to as 2000 are from the 2000 census, and data referred to as 2007–11 come from the 2007–11 American Community Survey 5-year estimates. The geographic unit of observation is the ZIP Code Tabulation Areas (ZCTAs).

Data on payroll and number of employees for US zip codes in 2010 come from the ZIP Code Business Patterns from the US Census Bureau. Note that zip codes and ZCTAs are not exactly the same. In particular, zip codes do not have areas. ZCTAs should be understood as areal representations of US Postal Service zip codes. When calculating employee density for zip codes, we therefore use the

corresponding ZCTAs. To match ZCTAs to CBSA, we rely on tabulation files from the US Census Bureau.

Historical population data.—Country-level historical population data for the period 1950–2000 come from the Penn World Tables 8.1. Similar data for 1920 and 1870 come from Maddison (2001).

References

Abdallah, S., S. Thompson, and N. Marks. 2008. "Estimating Worldwide Life Satisfaction." *Ecological Econ.* 65 (1): 35–47.

Albouy, D., W. Graf, R. Kellogg, and H. Wolff. 2016. "Climate Amenities, Climate Change, and American Quality of Life." *J. Assoc. Environmental and Resource Econ.* 3 (1): 205–46.

Allen, T., and C. Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Q.J.E.* 129 (3): 1085–1140.

Behrens, K., G. Mion, Y. Murata, and J. Suedekum. 2017. "Spatial Frictions." *J. Urban Econ.* 97 (C): 40–70.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. "Plants and Productivity in International Trade." *A.E.R.* 93 (4): 1268–90.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P. Sarte. 2018. "The Impact of Regional and Sectoral Productivity Changes in the U.S. Economy." *Rev. Econ. Studies*, forthcoming.

Carlino, G., S. Chatterjee, and R. Hunt. 2007. "Urban Density and the Rate of Invention." *J. Urban Econ.* 61:389–419.

Combes, P., G. Duranton, L. Gobillon, D. Puga, and S. Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica* 80:2543–94.

Comin, D. A., M. Dmitriev, and E. Rossi-Hansberg. 2012. "The Spatial Diffusion of Technology." Working Paper no. 18534, NBER, Cambridge, MA.

Costinot, A., and A. Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 197–261. Amsterdam: Elsevier.

Davis, M. A., and F. Ortalo-Magné. 2011. "Household Expenditures, Wages, Rents." *Rev. Econ. Dynamics* 14 (2): 248–61.

Deaton, A. 2008. "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll." *J. Econ. Perspectives* 22 (2): 53–72.

Deaton, A., and A. Stone. 2013. "Two Happiness Puzzles." *A.E.R.* 103 (3): 591–97.

Desmet, K., and J. Rappaport. 2017. "The Settlement of the United States, 1800–2000: The Long Transition towards Gibrat's Law." *J. Urban Econ.* 98 (March): 50–68.

Desmet, K., and E. Rossi-Hansberg. 2013. "Urban Accounting and Welfare." *A.E.R.* 103 (6): 2296–2327.

———. 2014. "Spatial Development." *A.E.R.* 104 (4): 1211–43.

———. 2015. "On the Spatial Economic Impact of Global Warming." *J. Urban Econ.* 88:16–37.

Diamond, R. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *A.E.R.* 106 (3): 479–524.

Eaton, J., and S. Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–79.

Fajgelbaum, P., E. Morales, J. Suarez Serrato, and O. Zidar. 2016. "State Taxes and Spatial Misallocation." Manuscript, Princeton Univ.

Fajgelbaum, P., and S. Redding. 2014. "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870–1914." Manuscript, Princeton Univ.

Glaeser, E. L., J. D. Gottlieb, and O. Ziv. 2016. "Unhappy Cities." *J. Labor Econ.* 34 (S2): S129–S182.

Greenwood, J., Z. Hercowitz, and P. Krusell. 1997. "Long-Run Implications of Investment-Specific Technological Change." *A.E.R.* 87 (3): 342–62.

Head, K., and T. Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 131–95. Amsterdam: Elsevier.

Jones, C. I. 1995. "R&D-Based Models of Economic Growth." *J.P.E.* 103 (4): 759–84.

Kahneman, D., and A. Deaton. 2010. "High Income Improves Evaluation of Life but Not Emotional Well-Being." *Proc. Nat. Acad. Sci.* 107 (38): 16489–93.

Kennan, J. 2013. "Open Borders." *Rev. Econ. Dynamics* 16 (2): L1–L13.

Klein, P., and G. Ventura. 2007. "TFP Differences and the Aggregate Effects of Labor Mobility in the Long Run." *B.E.J. Macroeconomics* 7 (1): 1–36.

Kline, P., and E. Moretti. 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *Q.J.E.* 129 (1): 275–331.

Maddison, A. 2001. *The World Economy: A Millennial Perspective*. Paris: Development Centre of the OECD.

Monte, F., S. Redding, and E. Rossi-Hansberg. 2018. "Commuting, Migration and Local Employment Elasticities." *A.E.R.*, forthcoming.

Ortega, F., and G. Peri. 2013. "The Effect of Income and Immigration Policies on International Migration." *Migration Studies* 1 (1): 47–74.

Roback, J. 1982. "Wages, Rents, and the Quality of Life." *J.P.E.* 90 (6): 1257–78.

Rosen, S. 1979. "Wage-Based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics*, edited by Peter M. Mieszowski and Mahlon R. Straszheim, 74–104. Baltimore: Johns Hopkins Univ. Press.

Simonovska, I., and W. E. Waugh. 2014. "Trade Models, Trade Elasticities, and the Gains from Trade." Working Paper no. 20495, NBER, Cambridge, MA.

Stevenson, B., and J. Wolfers. 2013. "Subjective Well-Being and Income: Is There Any Evidence of Satiation?" *A.E.R.* 103 (3): 598–604.

Zabreyko, P., A. Koshelev, M. Krasnoselskii, S. Mikhlin, L. Rakovshchik, and V. Stet'senko. 1975. *Integral Equations: A Reference Text*. Leyden: Noordhoff.