Quality of life in a dynamic spatial model*

Gabriel M. Ahlfeldt † Fabian Bald ‡ Duncan H.W. Roth § Tobias Seidel ¶

April 29, 2021

Abstract

We develop a dynamic spatial model in which heterogeneous workers are imperfectly mobile and forward-looking and yet all structural fundamentals can be inverted without assuming that the economy is in a stationary spatial equilibrium. Exploiting this novel feature of the model, we show that the canonical spatial equilibrium framework understates spatial quality-of-life differentials, the urban quality-of-life premium and the value of local non-marketed goods. Unlike the canonical spatial equilibrium framework, the model quantitatively accounts for local welfare effects that motivate many place-based policies seeking to improve quality of life.

Key words: Covid, dynamic, housing, migration, rents, pollution, productivity, spatial equilibrium, quality of life, wages, welfare

JEL: J2, J3, R2, R3, R5

*We thank seminar, conference, and workshop participants in Amsterdam (Tinbergen), Duisburg-Essen (Uni), EALE (online), EEA (online), EWMES (online), Essen (RWI), ifo CEMIR (online), Ispra (EC Joint Research Center, online), London (LSE), Lyon (GATE), Moscow (HSE, virtual), Paris (HEC, virtual), Potsdam (VfS), Rome (BOI), the virtual seminar in local public economics (online), and St. Gallen (Uni) for comments and suggestions, in particular Michael Amior, Thomas Bauer, Kirill Borusyak, Helge Braun, Martin Brown, Gilles Duranton, Christian Dustmann, Roland Fuess, Cecile Gaubert, Christoph Hanck, Stephan Heblich, Vernon Henderson, Christian Hilber, Philip Jung, Hans Koster, Joan Monras, Guy Morrell, Jos van Ommeren, Florian Oswald, Fernando Parro, Jacques Poot, Esteban Rossi-Hansberg, Olmo Silva, Daniel Sturm and Coen Teulings for comments and suggestions. The usual disclaimer applies.

†London School of Economics, CEPR & CESifo

‡University of Duisburg-Essen, RGS

§Institute for Employment Research (IAB)

¶University of Duisburg-Essen, CESifo & CRED

A Introduction

In economics, *quality of life* (QoL) is a location-specific utility shifter that can be used to value local public goods or bads such as clean or dirty air. From Ricardo (1817) via the neoclassical Rosen (1979)-Roback (1982) framework to quantitative spatial models (QSMs) summarised by Redding and Rossi-Hansberg (2017), economists have inferred QoL assuming a *competitive spatial equilibrium* (CSE) in which free mobility of homogeneous workers leads to perfect spatial arbitrage. Spatially invariant utility then ensures that spatial differences in amenity values are offset by differences in real wages, the so-called *compensating differential*. In reality, workers rarely move between local labour markets more than once or twice over their employment biography, owing to idiosyncratic tastes for locations and non-pecuniary migration costs that typically exceed the equivalent of an annual income (Koşar et al., 2021). Hence, spatial arbitrage is likely imperfect, raising a range of important questions. How should we measure QoL without imposing an exogenous reservation utility level? How should we value local non-marketed goods if real wage differences do not map directly to compensating differentials? How should we evaluate the aggregate and distributional consequences of QoL policies in a frictional world with *spatial incidence*, i.e. persistent localised utility effects?

To answer these questions, we develop a quantitative general equilibrium model that combines the strengths of two recent classes of spatial models. It inherits the complete invertibility from QSMs (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Monte et al., 2018) and the ability to account for frictional adjustments in the spatial economy from dynamic spatial models (DSMs) (Desmet et al., 2018; Caliendo et al., 2019a; Monras, 2020). Specifically, we propose the first DSM with heterogeneous, imperfectly mobile and forward-looking agents that can be fully quantified without assuming that the economy is observed in stationary spatial equilibrium. We exploit this novel feature for a threefold contribution. First, we propose a new approach to measuring QoL that allows for worker heterogeneity and costly migration and does not impose any restriction on the spatial distribution of worker utility. Second, we show theoretically and empirically that the canonical CSE framework severely understates spatial differentials in QoL, the urban QoL premium, and the value of local public goods. Third, we illustrate how the welfare effect of spatially targeted QoL policies critically depends on the social welfare function, owing to imperfect spatial arbitrage, relocation effects, and spatial incidence.

Our quantitative model incorporates an arbitrary number of worker groups and an arbitrary number of local labour markets that are interconnected through costly migration. Following the conventions in the literature, we treat QoL as a group-region-specific structural fundamental that shifts utility. Locations further differ in terms of exogenous housing productivity and land supply. Labour productivity is group-region-specific and consists of an exogenous component and an endogenous component that positively depends on density (Combes and Gobillon, 2015). Labour is the only factor of production used to produce one final good which is freely traded and consumed at a spatially invariant price. Housing is produced by developers who use capital and land from absentee owners as inputs. Workers spend their labour income on the tradable good and housing. All markets are competitive. Inelastic supply of land generates a dispersion force in the form of high rents in places in high demand (Combes et al., 2019).

Unlike in CSE models that assume perfect spatial arbitrage, spatial arbitrage is an endogenous mechanism in our model that operates through migration. Intuitively, migration into an attractive region congests the housing market, leading to subsequently reduced in-migration as long as the housing-market-related congestion force exceeds the labour-market-related agglomeration force. Concretely, we model migration as an investment decision in which workers choose destinations facing a trade-off between the present value of expected utility flows and a one-off relocation cost. Following the discrete choice literature in the tradition of McFadden (1974), workers receive bilateral amenity shocks with an idiosyncratic and a group-year-specific component. This stochastic formulation provides the microeconomic foundation for a migration gravity equation that has been found to be empirically successful (Kennan and Walker, 2011; Bryan and Morten, 2019; Tombe and Zhu, 2019). The dispersion of the idiosyncratic component is inversely related to the migration elasticity, which monitors how strongly bilateral migration probabilities respond to differences in expected indirect utility at migration destinations. If the migration elasticity approaches zero, shocks to labour and housing productivity or QoL will not trigger migration so that any localised utility effect remains persistent. If the migration elasticity approaches infinity, there is no taste heterogeneity so that migration will go on until a shock that has caused migration is fully offset by adjustments in wages and rents. Spatial arbitrage is then perfect. We show that for values of the migration elasticity found in our data and in previous research (Caliendo et al., 2019b), the marginal worker's willingness to accept high real living cost steeply decreases in the size of a local labour market. Therefore, our model rationalises real living cost differentials by much larger differences in group-specific average QoL than the canonical CSE framework, leading to a higher urban QoL premium and larger valuations of local public goods.

When switching between labour markets workers pay an origin-destination-group-specific migration cost in the form of foregone utility in the relocation period. Workers remaining at their origin incur no migration cost. Larger bilateral migration costs map to smaller migration flows between local labour markets, leading to a lower speed of spatial arbitrage. More generally, positive migration costs imply that spatial adjustments are non-instantaneous, giving rise to the dynamic structure of the model and distinct notions of spatial equilibria. In the absence of a consensus, we take the liberty of naming a *transitory spatial equilibrium* (TSE) and a *stationary spatial equilibrium* (SSE) that prevail in the nascent DSM literature. The role of the TSE is to rationalise observed data assuming that goods and factor markets clear without imposing any restriction on trends in prices and quantities on labour and housing markets. In the SSE, goods and factor markets clear *and* all prices and quantities are stationary. Intuitively, the SSE is a counterfactual situation to which a spatial economy would mean-revert in the absence of further shocks to labour productivity, housing productivity and QoL (the structural fundamentals). Since imperfectly mobile workers likely form sophisticated expectations about the economic prospects at destinations, we assume that workers anticipate all model-endogenous adjustments in wages and prices that occur over the transition from the

TSE into the SSE.

Our main methodological contribution is to develop a DSM with forward-looking agents that can be fully quantified from the TSE. The quantification follows the basic steps known from the QSM literature (Redding and Rossi-Hansberg, 2017). First, we use observed data and the structure of the model to estimate the key structural parameters. Second, we use observed data, the structure of the model, and the structural parameters to invert the structural fundamentals. For the quantification, we leverage on a matched employer-employee data set covering about 30M German workers contributing to social insurances, who we track over space and time. In particular, we observe the local labour market in which they work (Kosfeld and Werner, 2012), the nominal wage, and a range of characteristics including age, gender, and education for all years from 1993 to 2017. Aggregation of these micro data yields total employment and bilateral migration by region, year and 18 worker groups based on age, gender, and skills. To these data, we merge a regional mix-adjusted property price index starting in 2007, which we generate from property micro data containing about 17M observations.

We derive all empirical specifications used in the estimation of the structural parameters directly from the structure of the model. The identification strategies we use are close to what we consider the current best-practice examples in the respective literature. Our contribution is to exploit the richness of our data to provide parameter estimates for 18 gender-skill-age groups. We estimate the density elasticity of productivity from between-labour market movers controlling for individual fixed effects (Combes et al., 2008) using a 100-year lag of population density as an instrument (Ciccone and Hall, 1996). Depending on the group, our elasticity estimates range from near zero to 0.042, with relatively large estimates for female, skilled, and middle-aged workers. The weighted average of 0.024 is close to the consensus in the literature (Combes and Gobillon, 2015). Our strategy to estimating the share of land in housing is closest to Combes et al. (2019). We estimate a value of 0.19 which is within the typical range in the literature (Ahlfeldt and Pietrostefani, 2019). For the migration elasticity, we use a log-linearised and spatially differenced version of a migration gravity equation in which leading migration probabilities control for future utility flows following Artuç et al. (2010). Our group-specific estimates range from 0.12 to 0.58 which compares to an estimate of 0.5 for the average worker in the US (Caliendo et al., 2019b). To obtain group-origin-destination-specific estimates of bilateral migration costs, we use our estimates of the migration elasticity, the restriction that internal migration is costless, and a non-parametric version of a conventional migration gravity equation (Head and Mayer, 2014). Based on our estimates, we monetise the average moving cost at €170K which is towards the higher end of the survey-based estimates provided by Koşar et al. (2021). Controlling for distance and instrumenting with historic dialect similarity (Falck et al., 2012), social connectedness as measured by Bailey et al. (2018) has a large and positive effect on our estimated migration costs, suggesting a role for social capital (Glaeser et al., 2002).

Conditional on these estimates, the inversion of fundamental housing and labour productivity is straightforward as there is a one-to-one mapping from wages and rents for given structural parameters and observed density. In contrast, the inversion of QoL from the TSE in a DSM with forward-looking agents is challenging. While QoL is straightforward to invert for given expected wages and rents, the model requires QoL as an input to forecast the transition paths of wages and rents to the SSE. The DSM literature has not yet found an elegant solution to this circularity problem. Desmet et al. (2018) and Conte et al. (2020) avoid the problem by assuming that workers have static expectations.[1] Monras (2020) avoids the problem by assuming that the economy is observed in a long-run equilibrium. Caliendo et al. (2019b), Caliendo et al. (2019a) and Balboni (2019) use "dynamic hat algebra" to quantify the model in differences and do not invert QoL.[2] Our contribution is to develop a new procedure that inverts QoL and solves for the SSE simultaneously. To this end, we exploit that there is a one-to-one mapping from employment to wages and rents for given structural fundamentals and parameters. Therefore, we can conclude the quantification of the model by treating the identification of the unknown group-region-specific QoL *and* the unknown vector of group-region-specific employment for all future periods as a fixed point problem that is solved numerically. With this approach, we find that about 65% of the spatial convergence from the TSE to the SSE are completed within 30 years.

In the first application of our quantified model, we establish that our novel QoL index (DSM-QoL) is much more dispersed than the canonical Rosen-Roback measure (RR-QoL). In log terms, the within-group standard deviation of the DSM-QoL exceeds that of the RR-QoL by a factor of three. This is a striking result that has major implications for the literatures on the origins of QoL (e.g. Roback, 1982; Blomquist et al., 1988; Albouy, 2011) and the value of local public goods (e.g. Chay and Greenstone, 2005; Linden and Rockoff, 2008; Cellini et al., 2010). We estimate that the city size elasticity of the DSM-QoL, at about 0.45, is about four times as large as for the RR-QoL. Hence, the extant literature may have dismissed an urban QoL premium too soon (see Albouy, 2011, for a summary). We find that consumption benefits contribute more to the spatial concentration of workers in cities than productivity advantages, which helps rationalizing why high-skilled workers tend to live in cities although many of their jobs could be done remotely (Althoff et al., 2020). The relatively low dispersion of the RR-QoL is also consequential for the valuation of local public goods. As an example, a decrease in air pollution is associated with a more than twice as large increase in DSM-QoL than in RR-QoL.[3] This result helps reconciling the puzzling finding that the monetised effect of dirty air on self-reported well-being is larger than the willingness to pay for clean air inferred from property prices under the CSE assumption (Luechinger, 2009). Quantifying the model under alternative values of the migration elasticity, we find that the elasticity of the RR-QoL with respect to the DSM-QoL increases from less than 0.3 to 0.8 if we increase the migration elasticity from 0.3 to 3.0. Hence, the CSE remains a useful and convenient framework for settings where the idiosyncrasy of tastes plays a subordinate role. Since a simple count measure of geo-tagged photos shared online (Ahlfeldt, 2013) explains almost 60% of the variation in DSM-QoL, social

[1] In a static model, quality of life can be interpreted as a component of the destination fixed effect in a gravity regression (Lin et al., 2019).

[2] See Table A1 for a summary classification of the related literature.

[3] This finding echos Bayer et al. (2009) who extend a hedonic model to account for moving cost when estimating the marginal willingness to pay for clean air.

media represents an alternative avenue to proxy for QoL differentials, similar to use of lights at night as a proxy for GDP (Henderson et al., 2012).

In the second application of our quantified model, we illustrate how the tractability of our DSM makes it a powerful tool for spatial policy analysis. We introduce a procedure suitable for the evaluation of any spatial policy that has an effect on any of the structural fundamentals in general equilibrium. Since returns to agglomeration and amenity preferences are group-specific, the model delivers rich sorting effects and group-specific welfare effects (Diamond, 2016; Almagro and Domínguez-Iino, 2020). Because the model accounts for imperfect spatial arbitrage and does not impose any restriction on the spatial distribution of expected worker utility, spatial policies have spatial welfare effects. This is an important contribution to a literature on place-based policy evaluation in which the incidence on non-marginal workers is well understood theoretically (Moretti, 2011; Kline and Moretti, 2014), but ruled out in the extant quantitative frameworks based on the CSE (Blouri and Ehrlich, 2020; Fajgelbaum and Gaubert, 2020).[4] We illustrate our procedure for a hypothetical policy that reduces air pollution in the most polluted areas, similar to the US Clean Air Act (Chay and Greenstone, 2005). To this end, we establish the group-specific causal link between the inverted DSM-QoL and observed air pollution (PM^{10}) exploiting wind-induced exogenous variation (Deryugina et al., 2019; Heblich et al., 2020a). Starting from the SSE, we use these estimates to update QoL to reflect the policy change and let the model converge to a counterfactual SSE. Comparing the initial to the counterfactual SSE, we obtain group-region-specific changes in expected utility alongside group-region-specific wage, region-specific rent and rich sorting effects. This SSE-to-SSE comparison provides causal estimates of the place-based policy that are unconfounded by the mean-reversion tendency of the economy and account for displacement effects that are a challenge in the reduced-form estimation of spatial policy effects. In a nutshell, we find that workers move from the untreated to the (positively) treated regions. Due to sorting and agglomeration effects, the policy effect on GDP is somewhat larger than on population. Since only about one fourth of the QoL increase capitalises into rents, expected utility in the treated areas increases. Expected utility also increases in the untreated areas since the relocation of workers reduces congestion on the housing market. In our example, spatial incidence increases spatial inequality in welfare. Applying a lower-bound penalty for inequality aversion following Atkinson (1970) reduces the social welfare effect by 13%. This is an important insight for the literature in the tradition of Rosen (1979)-Roback (1982) which has abstracted from a potential efficiency-equity trade-off by assuming perfect spatial arbitrage.

The remainder of the paper is structured as follows. Section B presents stylised evidence that guides our modelling choices. Section C outlines the model. Section D describes the quantification of the model. Section E compares our new QoL index to the canonical measure in the literature. Section F shows how to use the model for policy analysis. Section G concludes.

[4]Much of the place-based policy focuses on reduced-form methods to provide causal evidence (Kline and Moretti, 2013, 2014; Criscuolo et al., 2019). See Neumark and Simpson (2015) for a recent summary.

B Stylised facts

To motivate the structure of the model developed in Section C, we present some stylised facts of a spatial economy in Figure 1 using data that we describe in Section D.1. The upper panels show how spatial concentration is associated with benefits due to agglomeration economies on labour markets (a) and costs due to congestion on housing markets (b). Intuitively, the strengths of these agglomeration and dispersion forces determine the spatial concentration of economic activity.

In the middle panels, we turn to causes and consequences of migration. There is a positive association between the wage a local labour market offers and the number of workers it attracts (c). At the same time there is a positive association between net in-migration into labour markets and changes in local housing cost (d). This descriptive evidence supports some important assumptions that are implicit to the notion of a spatial equilibrium and the idea of spatial arbitrage. First, workers are at least imperfectly mobile and respond to economic incentives when making location decisions. Second, due to inelastic supply of land, migration into attractive destinations leads to rising house prices and mean reversion in the attractiveness of locations.

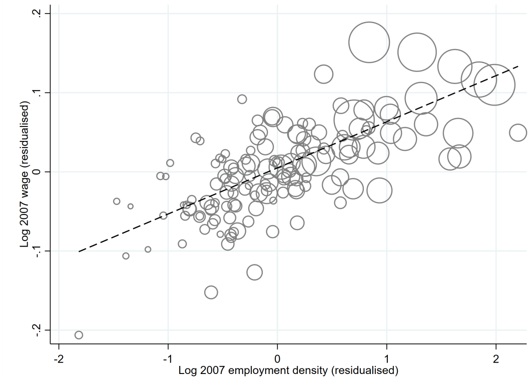
Yet, the bottom panels of Figure 1 reveal that workers are not perfectly mobile. The average worker changes the labour market region about once (1.08) over the employment biography, although there is some variation across groups (e). Conditional on migrating, the propensity of a location becoming a migration destination declines rapidly in space, which points to spatially variant migration costs (f).

Motivated by these stylised facts, we develop a model in which imperfectly mobile workers trade off expected utility at migration destinations against migration costs. In-migration reduces incentives to migrate into a region since the cost of agglomeration exceeds the benefit, so that in the absence of shocks, the spatial economy tends to revert to a stationary spatial equilibrium.

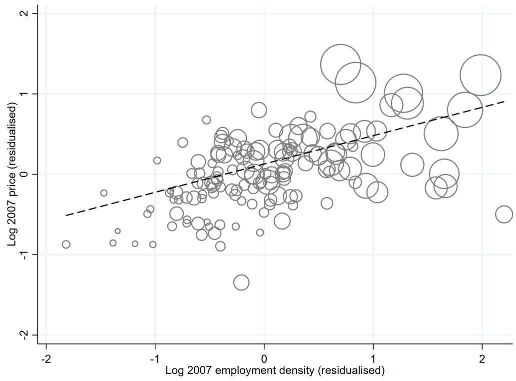
C Model

Consider an economy that is populated by $\bar{L} = \sum_\theta \bar{L}^\theta$ workers who we categorise into groups $\theta \in \Theta$ (e.g. according to age, gender, skill) and who supply one unit of labour inelastically. Individuals choose their place of residence and work among $i, j \in J$ local labour markets to which we refer as regions. Workers in i have idiosyncratic tastes for living in j and incur a cost when migrating from i to j. Each region is endowed with a measure $\bar{T}_i$ of land used for housing.

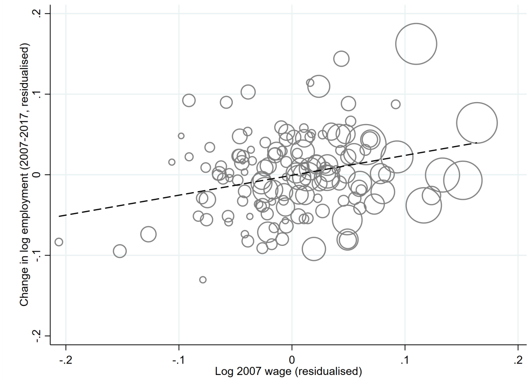
Figure 1: Stylized facts of the spatial economy



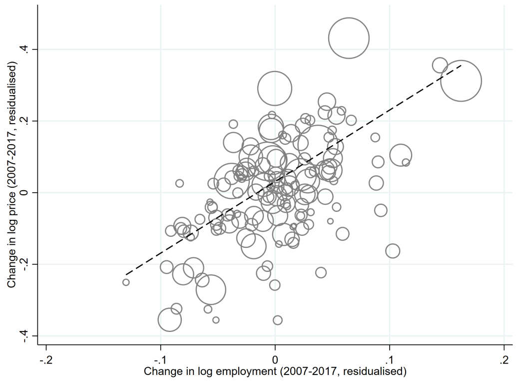
(a) Agglomeration benefits



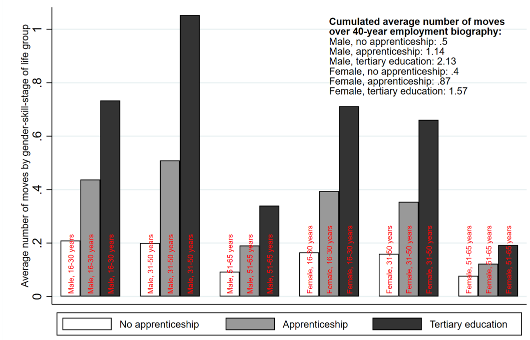
(b) Agglomeration costs



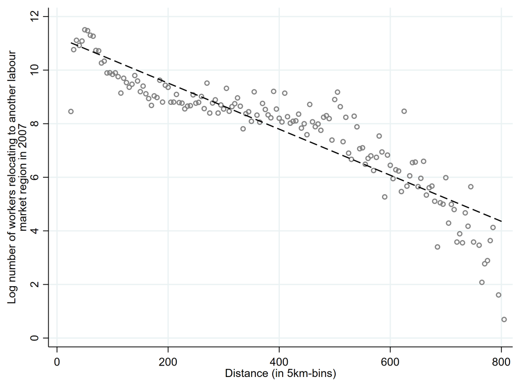
(c) Wages and migration



(d) Migration and housing costs



(e) Number of moves in employment biography



(f) Spatial decay in migration flows

Note: Unit of observation in panels (a)-(d) is 141 labour market areas as defined by Kosfeld and Werner (2012). Residualised variables have been purged of mean differences between East and West German regions. Panel (e) uses all workers observed in at least 35 years over at least 40 years starting in 1975 (in West Germany). A "move" is a change in local labour market. Outcomes in panels (a-d) are residualised in regressions against zone (for former East and West Germany) fixed effects. Panel (f) is based on 5km-bins of bilateral distance between labour market regions. Sub-section D.1 provides a description of the underlying variables.

C.1 Workers

Individual ω belonging to group θ, living in region i at time period t, and previously living in region k derives utility from the consumption of a freely-tradable homogeneous good ($x^{\theta}_{i,t}(\omega)$), housing ($h^{\theta}_{i,t}(\omega)$) and amenities ($A^{\theta}_{i,t}$, $\exp[a^{\theta}_{ki,t}(\omega)]$) according to

$$U^{\theta}_{i|k,t}(\omega) = \left(\frac{x^{\theta}_{i,t}(\omega)}{\alpha}\right)^{\alpha} \left(\frac{h^{\theta}_{i,t}(\omega)}{1-\alpha}\right)^{1-\alpha} A^{\theta}_{i,t} \exp\left[a^{\theta}_{ki,t}(\omega) - \tau^{\theta}_{ki}\right]. \tag{1}$$

The Cobb-Douglas structure implies that individuals spend constant shares α and $1-\alpha$ of their income on the tradable good and housing. Normalising the price of the homogeneous good to unity, $p_{i,t}$ represents the relative price of housing in region i. We then obtain the demand functions

$$\begin{aligned} x^{\theta}_{i,t}(\omega) &= \alpha(1-\iota)w^{\theta}_{i,t}(\omega) \\ h^{\theta}_{i,t}(\omega) &= \frac{(1-\alpha)(1-\iota)w^{\theta}_{i,t}(\omega)}{p_{i,t}}, \end{aligned} \tag{2}$$

where ι denotes the federal income tax rate and $w^{\theta}_{i,t}(\omega)$ are gross wages for an individual ω in group θ in region i.

Migration from k to i comes at a time-invariant cost that depreciates utility in the moving period to $\exp\left[-\tau^{\theta}_{ki}\right]$, with $\tau^{\theta}_{ki} \geq 0$ and $\tau^{\theta}_{k,i=k} = 0$. Since we allow for arbitrary group-origin-destination-specific migration costs, we can remain agnostic about the exact nature of this cost. An intuitive interpretation is the cost of rebuilding social capital (Glaeser et al., 2002) which may depend on how closely two regions are connected geographically, culturally (Falck et al., 2012), or socially (Bailey et al., 2018).

The composite amenity consists of two components. The first component is QoL, an exogenous group-region utility shifter that collects the group-specific effects of region-specific (dis)amenities:

$$A^{\theta}_{i,t} = \zeta^{\theta}_{t} \bar{A}^{\theta}_{i,t}, \tag{3}$$

where ζ^{θ}_{t} is a group-period-specific constant and $\bar{A}^{\theta}_{i,t}$ is a relative QoL measure with a within-group mean of one. The second component $\exp[a^{\theta}_{ki,t}(\omega)]$ is a stochastic bilateral amenity shock, with $a^{\theta}_{ki,t}(\omega)$ being drawn from a type-I-extreme value (Gumbel) distribution

$$F^{\theta}_{ki,t}(a) = \exp\left(-\tilde{B}^{\theta}_{ki,t} \exp\{-\left[\gamma^{\theta} a + \Gamma\right]\}\right) \quad \forall \;\; \theta \;\; \text{and} \;\; \gamma^{\theta} > 0, \tag{4}$$

where $\tilde{B}^{\theta}_{ki,t} \equiv \left(B^{\theta}_{ki,t}\right)^{\gamma^{\theta}}$. With this formulation, we follow the multinomial logit model of discrete response (McFadden and Train, 2000) and allow for a group-specific mean and a group-specific variance of the amenity shock. $\ln(B^{\theta}_{ki,t})$ is the time-varying, group-specific mean of the amenity shock and Γ is the Euler-Mascheroni constant.[5] γ^θ governs the group-specific dispersion of individual amenity shocks.

Amenity shocks are conceptually important and essential for the tractability of the model. The bilateral group-year component $B^\theta_{ki,t}$ captures common trends such as downtown gentrification that make specific pairs of locations closer substitutes for certain groups in certain periods. Since we view migration cost as time-invariant, this is important to rationalise migration flows that vary over time within groups and bilateral region pairs even if wages, rents, and QoL remain constant. The heterogeneity of shocks within groups allows for some idiosyncrasy in tastes for being in i among workers of group θ from k. Unless we are in the limit case $\gamma^\theta \to \infty$ and tastes are homogeneous, there will be some workers within a group who will have decided to migrate from k to i for given wages, rents, QoL, and migration costs, while others did not. Hence, spatial arbitrage is imperfect in the real world and in our model.

C.2 Production

Tradable good. Firms produce the tradable good under perfect competition using labour as their only input. Following the conventions in urban economics (Combes and Gobillon, 2015) we model the productivity of individuals, $\varphi^\theta_{i,t}(\omega)$, as dependent on location factors that are exogenous to our model (e.g. access to navigable rivers), endogenous agglomeration (employment density), and an individual effect that consists of time-invariant (innate skill) and time-varying (e.g. employment status) factors:

$$\varphi^\theta_{i,t}(\omega) = \psi^\theta_{i,t} \left(\frac{L_{i,t}}{\bar{T}_i} \right)^{\kappa^\theta} \delta^\theta_{i,t}(\omega), \tag{5}$$

where $\delta^\theta_{i,t}(\omega)$ summarises idiosyncratic determinants of productivity and the group-region productivity $\varphi^\theta_{i,t} = \psi^\theta_{i,t}(\frac{L_{i,t}}{\bar{T}_i})^{\kappa^\theta}$ depends on an exogenous component $\psi^\theta_{i,t}$ and on density $L_{i,t}/\bar{T}_i$. Prompted by evidence on skill-biased returns to agglomeration (Baum-Snow and Pavan, 2013), we allow the density elasticity of productivity $\kappa^\theta \geq 0$ to vary across groups. Similarly, each group is equipped with a location-specific exogenous productivity $\psi^\theta_{i,t}$ to capture any complementarity between skills and exogenous location factors, such as an airport that allows high-skilled workers to quickly travel to business meetings.

We assume that firms only observe the average productivity per group, so we impose $\delta^\theta_{i,t}(\omega)$ to be a log-normally distributed error term of mean zero for the sake of simplicity. As the price serves as the numeraire, the first-order condition of labour demand implies that group-region productivity $\varphi^\theta_{i,t}$ directly maps into wages:

$$w^\theta_{i,t} = \psi^\theta_{i,t} \left(\frac{L_{i,t}}{\bar{T}_i} \right)^{\kappa^\theta}. \tag{6}$$

[5] This implies that shocks are i.i.d across locations, individuals, and time. This approach is established in the literature and has been applied to describe productivity distributions, e.g. as in Eaton and Kortum (2002), or individual preferences, e.g. as in Ahlfeldt et al. (2015).

Total output (equal to revenues and nominal income) in i is then given by $X_{i,t} = \sum_{\theta} L^{\theta}_{i,t}\varphi^{\theta}_{i,t}$.

Housing Profit-maximizing developers supply housing under perfect competition according to a Cobb-Douglas production function combining a share of the globally available capital stock with location-specific land:

$$H^S_{i,t} = \eta_{i,t}\left(\frac{\bar{T}_i}{\beta}\right)^{\beta}\left(\frac{K_{i,t}}{1-\beta}\right)^{1-\beta}, \tag{7}$$

where $K_{i,t}$ is the capital used in region i and $\eta_{i,t}$ denotes total factor housing productivity, capturing the role of regulatory (e.g. height regulations) and physical (e.g. a rugged surface) constraints (Saiz, 2010). Owners of employed capital and land are absent so their income is irrelevant for local demand. Normalising the world price of capital to unity and assuming that developers make zero profits and housing markets clear, we obtain

$$p_{i,t} = \left(\frac{(1-\alpha)\beta(1-\iota)X_{i,t}}{\eta_{i,t}^{\frac{1}{\beta}}\bar{T}_i}\right)^{\beta}. \tag{8}$$

This formulation implies that both capital input and housing prices are increasing in housing expenditure, and that $p_{i,t}$ is lower in locations with more land supply and higher housing productivity, ceteris paribus. The larger the share of land in housing β, the smaller the housing supply elasticity $(1-\beta)/\beta$, and the greater the congestion force the housing market generates (see Appendix J.1 for details).

C.3 Migration and timing

Our approach to modelling migration decisions draws from financial economics. Intuitively, we model migration as an investment decision in which expected returns in the form of utility flows are traded against a migration cost, e.g. for rebuilding social capital at a potential destination. The timing is as follows. Throughout period t, workers living in i realise their k-i-worker-specific utility. At the end of period t, workers receive i-j-worker-specific amenity shocks introduced in Section C.1. At the beginning of period $t+1$ workers choose location $j \in J$ such that they maximize their expected utility. Then, the procedure starts over again.

In line with the conventions in the emerging DSM literature (Caliendo et al., 2019b), we assume that workers have logarithmic preferences. This gives a migration net present value (NPV) for a worker of type θ who was in region k in period $t-1$, is in region i in period t, considers moving to region j in period $t+1$ and expects to be in regions $\{m, ..., n \in J\}$ in the subsequent periods and in $h \in J$ from $(t+1)+T$ onward (see Appendix J.2 for derivations):

$$\begin{aligned}\ln NPV^{\theta}_{i|k,t}(\omega) = &\ln\left(v^{\theta}_{i|k,t}(\omega)\right) + \max_{\{j,m,\dots,n,h\}^{J}_{j,m,\dots,n,h=1}} \Bigg\{\frac{1}{1+\rho}\ln\left(v^{\theta}_{j|i,t+1}(\omega)\right) \\ &+\left(\frac{1}{1+\rho}\right)^{2} E\left[\ln\left(v^{\theta}_{m|j,t+2}(\omega)\right)\right] + \ \dots + \left(\frac{1}{1+\rho}\right)^{T+1}\Bigg[E\left[\ln\left(v^{\theta}_{h|n,(t+1)+T}(\omega)\right)\right] \\ &+ \sum_{s=(t+2)+T}^{\infty}\left(\frac{1}{1+\rho}\right)^{s-(t+1+T)} E\left[\ln\left(v^{\theta}_{h|h,(t+2)+T}(\omega)\right)\right]\Bigg]\Bigg\},\end{aligned} \tag{9}$$

where ρ is a discount rate monitoring the time preference, and $v^{\theta}_{i|k,t}(\omega)$ is the per-period indirect utility defined by Eqs. (1) and (2). Intuitively, the NPV depends on future wages, rents, and QoL in the next and all subsequent migration destinations, the probabilities at which subsequent migration destinations will be accessed, and the migration cost at which the next and the subsequent destinations can be accessed.

Given the distributional assumption regarding the idiosyncratic amenity component, we obtain the following conditional probability that a worker from group θ migrates from i to j (see Appendix J.3 for derivations):

$$\chi^{\theta}_{ij|i,t} = \frac{\left(m^{\theta}_{ij}B^{\theta}_{ij,t+1}\mathcal{V}^{\theta}_{j,t+1}\right)^{\gamma^{\theta}}}{\sum_{n\in J}\left(m^{\theta}_{in}B^{\theta}_{in,t+1}\mathcal{V}^{\theta}_{n,t+1}\right)^{\gamma^{\theta}}}, \tag{10}$$

where $m^{\theta}_{ij} = \exp\left[-\tau^{\theta}_{ij}\right]$ and $\mathcal{V}^{\theta}_{j,t+1} = \exp\left[\ln\left(\frac{(1-\iota)w^{\theta}_{j,t+1}A^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,t+1}}\right) + \mathcal{O}^{\theta}_{j,t+2}\right]$ is a function of next period's utility as well as the migration option value $\mathcal{O}^{\theta}_{t+2}$.

This migration option value at j in $t+1$ is an expected NPV of utility a worker can attain at any location $m \in J$ in period $t+2$, weighted by the expected probability to migrate from j to m. The expected utility at m in $t+2$ itself depends on the migration option value $\mathcal{O}^{\theta}_{m,t+3}$ a worker expects at that destination. Similarly, all future migration option values up until the next-to-last location $n \in J$ reached in period $t+T$ enter the equation. Intuitively, the migration option value $\mathcal{O}^{\theta}_{j+2}$ captures the ease at which migration destinations can be accessed in the future should there be idiosyncratic reasons to relocate, e.g. a new romance:

$$\mathcal{O}^{\theta}_{j,t+2} = \frac{1}{1+\rho}\ln\left[\sum_{m\in J}\left(\exp\left\{E\left[v^{\theta}_{m|j,t+2}(\omega)\right] + \mathcal{O}^{\theta}_{m,t+3}\left[\mathcal{O}^{\theta}_{l,t+4}\dots\mathcal{O}^{\theta}_{n,(t+1)+T}\right]\right\}\right)^{\gamma^{\theta}}\right]^{\frac{1}{\gamma^{\theta}}} \tag{11}$$

Migration flows from i to j are simply given by $M^{\theta}_{ij,t} = \chi^{\theta}_{ij|i,t}L^{\theta}_{i,t}$. Since all workers migrate to a destination in period t (which can be the origin), aggregate employment in region i in $t+1$ equates to the sum of inflows $M^{\theta}_{ji,t}$ from all locations j:

$$L^{\theta}_{i,t+1} = \sum_{j\in J} M^{\theta}_{ji,t} = \sum_{j\in J}\chi^{\theta}_{ji|j,t}L^{\theta}_{j,t}. \tag{12}$$

Eq. (10) provides the micro-foundations for a migration gravity equation with a destination-group-specific present value of future utilities $\mathcal{V}^{\theta}_{j,t+1}$, origin-destination-group-specific migration costs τ^{θ}_{ij} and bilateral amenity shocks $B^{\theta}_{ij,t+1}$, and an origin-group-specific component akin to the multilateral resistance known from trade models (the denominator). Via $\mathcal{V}^{\theta}_{j,t+1}$, higher wages, lower rents, a greater QoL and a larger migration option value at a potential destination increase the probability that workers migrate to j.

The amenity dispersion parameter γ^{θ} can be interpreted as a migration elasticity as it moderates how sensitive migration decisions are to economic incentives. At low values of γ^{θ}, the idiosyncrasy of tastes dominates and migration is inelastic whereas at high values difference in wages, rents, and QoL have large effects on migration flows. Migration costs τ^{θ}_{ij} are critical to rationalising why physically, culturally, or socially close region pairs generate larger migration flows. Since all $\tau^{\theta}_{i,j\neq i} \geq 0$ are defined relative to $\tau^{\theta}_{i,j=i} = 0$, migration costs critically determine the share of workers leaving a region in a period and, hence, the speed of spatial adjustments in our DSM. Empirically, the effects of migration costs and the migration elasticity are jointly determined by the origin-destination-group component $\tau^{\theta}_{ij} \times \gamma^{\theta}$, which we term *migration resistance*. Therefore, the typically observed distance decay in migration flows can be rationalised by a large difference in migration cost if tastes are heterogeneous (small γ^{θ}) or a small difference in migration costs if tastes are homogeneous (large γ^{θ}).

It is immediate from Eq. (10) that there are isomorphic model formulations in which bilateral amenity shocks $B^{\theta}_{ij,t}$ are subsumed into time-varying migration costs, or vice versa. We choose our parameterisation because we believe that differences in average migration flows observed over 25 years in our data are most likely driven by fundamental determinants of migration costs that hardly change over time, whereas deviations from the long-run average most likely reflect the short-run effects of random events that tend to cancel out over time.

C.4 Equilibrium

We take the *structural parameters* $\{\alpha, \beta, \rho, \iota, \gamma^{\theta}, \kappa^{\theta}, B^{\theta}_{ij,t}, \tau^{\theta}_{ij}\}$, *structural fundamentals* $\{\psi^{\theta}_{i,t}, \eta_{i,t}, A^{\theta}_{i,t}\}$, and labour and land endowments $\{\bar{L}^{\theta}_{t}, \bar{T}_{i}\}$ as exogenously given. We impose the following labour market clearing conditions:

$$\bar{L}^{\theta}_{t} = \sum_{i \in J} L^{\theta}_{i,t} \tag{13}$$

with the economy-wide labour endowment $\bar{L}_{t} = \sum_{\theta} \bar{L}^{\theta}_{t}$. Region-group specific labour supply determined by Eq. (12) aggregates to regional employment $L_{i,t} = \sum_{\theta} L^{\theta}_{i,t}$ which maps into wages $w^{\theta}_{i,t}$ via the first-order condition of labour demand, Eq. (6). Likewise, we impose housing market clearing so that regional employment $L^{\theta}_{i,t}$ maps into rents $p_{i,t}$ via output $X_{i,t}$ according to Eq. (8) (see Appendix Section J.1). Trade with the rest of the world clears the markets for tradable goods and capital inputs.

Transitory spatial equilibrium. Frictional migration implies that shocks to structural fundamentals lead to non-instantaneous adjustment in $L^{\theta}_{i,t}$. The role of the TSE is to rationalise unbalanced migration flows and non-stationary employment that are typically observed in data.

Stationary spatial equilibrium. Migration is spatially neutral if the sum of outflows equals the sum of inflows for each location:

$$\sum_{j \in J} \chi^{\theta}_{ij|i,t} L^{\theta}_{i,t} = \sum_{j \in J} \chi^{\theta}_{ji|j,t} L^{\theta}_{j,t} \qquad \forall \;\; j \in J, \theta \in \Theta. \tag{14}$$

This condition enforces that $L^{\theta}_{i,t}$ is stationary, but it does not rule out migration due to idiosyncratic taste shocks. We assume that the congestion force dominates the agglomeration force to ensure that all regions are populated. The latter is governed by κ^{θ} for each group according to Eq. (6). The former works through the price for housing as described by Eq. (8). The effect of changes in population on individual housing expenditure is given by $(1-\alpha)\partial p_{i,t}/\partial L^{\theta}_{i,t}$. We relegate details to Appendix J.4 where we also show that the economy converges to a unique SSE for given primitives. Since Eq. (14) is unlikely to hold in the data, we view the SSE as a counterfactual situation to which an economy observed in a TSE would converge in the absence of further shocks.

Dynamic equilibrium. For given structural parameters and structural fundamentals the dynamic equilibrium of the model is referenced by a $(J \times \Theta)\; \times Z_t$ vector of region-group-year-specific employment $\mathbf{L^{\theta}_{i,t}}$, where Z_t denotes the number of periods in the transition period from a TSE in t to the SSE reached in $t + Z_t$. Hence, the dynamic equilibrium nests the SSE and all TSEs up to the period where the spatial economy has converged to the SSE. For given structural fundamentals $\{\psi^{\theta}_{i_t}, \eta_{i,t}\}$, $\mathbf{L^{\theta}_{i,t}}$ maps to $(J \times \Theta)\; \times Z_t$ vectors of wages $\mathbf{w^{\theta}_{i,t}}$ and prices $\mathbf{p_{i,t}}$ via the first-order condition of labour demand, Eq. (6), and housing market clearing, Eq. (8).

Competitive spatial equilibrium. Characteristic for the CSE is the absence of spatial frictions. Within our framework, we can remove frictions by setting preference shocks and migration costs to zero ($a^{\theta}_{ki,t}(\omega) = 0, \tau^{\theta}_{ki} = 0$). Since workers optimally relocate across locations within any period, we can impose the standard spatial equilibrium condition that workers are indifferent between locations. To this end, we set the indirect utility equal to a group-time-specific reservation utility $\bar{U}^{\theta}_t$.

$$V^{\theta}_{i,t} = \left((1-\iota) w^{\theta}_{i,t}\right)^{\alpha} \left(\frac{(1-\iota) w^{\theta}_{i,t}}{p_{i,t}}\right)^{1-\alpha} \mathcal{A}^{\theta}_{i,t} = \bar{U}^{\theta}_t \tag{15}$$

Hence, observed wages and rents directly map to a Rosen-Roback (RR) QoL measure $\mathcal{A}^{\theta}_{i,t} = q^{\theta}_t p^{1-\alpha}_{i,t} / w^{\theta}_{i,t}$ (where q^{θ}_t collects all group-period-specific constants).

C.5 Worker expectations

In specifying how agents form expectations in DSMs, there is a trade-off between foresight and tractability. Desmet et al. (2018) develop a fully tractable DSM under static expectations, i.e. workers project current realizations of good and factor prices into the infinite future. In contrast, Caliendo et al. (2019b) exploit Bellman's principle to estimate model parameters and conduct counterfactual analyses under perfect foresight without pinning down all primitives. We marry both approaches with the aim of incorporating forward-looking expectations into a model where all structural parameters and fundamentals will be quantified.

Our choices are guided by the stylised fact that the average worker switches between local labour market only once over the entire employment biography (see Section B). For the quantification of the model, we therefore assume that workers deciding on a migration destination j do not expect to make a further move in the future. Workers who expect to migrate only once likely form sophisticated expectations with respect to the evolution of wages and rents. Therefore, we assume that workers correctly anticipate the dynamic equilibrium referenced by the employment vector $\mathbf{L}^{\theta}_{\mathbf{i,t}}$ and all model-endogenous adjustments in wages and prices summarised by $\mathbf{w}^{\theta}_{\mathbf{i,t}}$ and $\mathbf{p}_{\mathbf{i,t}}$. Shocks to exogenous structural fundamentals cannot be anticipated, so workers project observed realisations of QoL $A^{\theta}_{i,t+1}$ into the future. Consistent with the distributional assumptions in Eq. (4), workers expect a bilateral amenity $\mathbb{E}(B^{\theta}_{ij,t+s}) = 1$ for $s > 1$. In line with the conventions in DSMs, workers have an infinite time horizon and do not expect to age.

The assumption that workers do not expect to re-optimise their location choice in the future delivers a special case in which we can abstract from the migration option value $\mathcal{O}^{\theta}_{j,t+2}$ in Eq. (10) (see Appendix J.3 for details). To evaluate the role of expectations in a DSM, we replicate the main stages of our quantitative analyses under more and less restrictive special cases. In a less restrictive special case, we allow residents to anticipate one additional migration decision subsequent to an initial migration decision, which results in a migration option value $\mathcal{O}^{\theta}_{j,t+2}$ that does not depend on the option value $\mathcal{O}^{\theta}_{m,t+3}$. This special case covers nearly 90% of German workers who migrate twice or less over their employment biography. Extending expectations to additional future migration decisions results in a dimensionality problem in the inversion of structural fundamentals as highlighted by Caliendo et al. (2019b). In a more restrictive special case, we assume that workers are myopic and expect goods and factor prices to remain constant in real terms. We provide a detailed discussion of the motivations for and the consequences of our choices concerning expectations in Appendix N. The main takeaway is that the choice has moderate effects on the quantitative predictions of the model. This is an important insight as it suggests that even the seemingly restrictive assumption of myopic agents may represent a sensible approximation in high-dimensional settings such as in Desmet et al. (2018) or Conte et al. (2020). Among the considered special cases, we choose the one for the quantitative analyses in the main paper that delivers the most conservative quantitative results.

C.6 Spatial arbitrage

The CSE is the urban economics equivalent of the no-arbitrage condition in financial economics (Glaeser, 2008). Perfect spatial arbitrage is an assumption that leads to constant reservation utility as a building block of neoclassical urban economics models. In contrast, spatial arbitrage is an endogenous process in our DSM that moderates the transition from the TSE to a SSE.

Intuitively, shocks to structural fundamentals affect expected utility directly or indirectly. For example, a positive shock to labour productivity maps into higher wages $w^{\theta}_{i,t+s}$ according to Eq. (6) due to perfect competition on goods and labour markets and the choice of the tradable good as the numeraire. Likewise, a positive shock to housing productivity maps into lower housing costs $p_{i,t+s}$ according to Eq. (8) due to perfect competition among developers. Higher $w^{\theta}_{i,t+s}$ and lower $p_{i,t+s}$ affect bilateral migration probabilities $\chi^{\theta}_{ij|i,t}$ according to Eq. (10), leading to in-migration. Given Eq. (12), this results in endogenous changes in employment which in turn determine changes in wages according to Eq. (6) and housing costs according to Eq. (8). As long as agglomeration costs exceed agglomeration benefits at the margin, the consequence of migration is to reduce the differences in expected utility that cause migration. The pace at which this spatial arbitrage process takes place depends positively on the migration elasticity γ^{θ} and negatively on migration costs τ^{θ}_{ij}. Eqs. (10) and (12) establish how regions offering a greater indirect utility $\mathcal{V}^{\theta}_{j,t+1}$ will experience larger net-immigration the larger γ^{θ} and the smaller the migration resistance $\tau^{\theta}_{ij} \times \gamma^{\theta}$, ceteris paribus.

C.7 Quality-of-life premiums

The revealed-preference literature computes the value of amenities that jointly constitute QoL via spatial differences in *real living cost* $p_i^{1-\alpha}/w_i^{\theta}$, the inverse of the real wage (Rosen, 1979; Roback, 1982). Using the structural parameters and fundamentals quantified in Section D, Figure 2 provides a graphical illustration of the simulated model to show how QoL premiums are determined. Our case in point is the *urban QoL premium* which captures how QoL depends on city size, a question that is controversially debated in the literature (Albouy, 2011). To ease the presentation, we focus on the special case with one worker group and refer to Appendix J.5 for formal derivations.

Figure 2 depicts two equilibrium loci for locations $i = \{1, 2\}$. The solid lines refer to location 1 while the dashed schedules indicate location 2. The *housing equilibrium locus* (HH_i) is a log-linearised version of Eq. (8) collecting all combinations of real living costs and employment that satisfy all housing-market related conditions that must hold in the TSE (and the SSE). Under plausible parameterisations, the expenditure on housing increases faster in city size (due to inelastically supplied land) than the wage (due to agglomeration economies). Therefore, the housing equilibrium locus is positively sloped. Greater housing productivity η_i shifts the housing equilibrium locus downwards.

Likewise, the *migration equilibrium locus* (LL_i) collects all combinations of real living costs and employment that satisfy all migration-related conditions that must hold in the SSE. It is derived from Eq. (12). Intuitively, the migration equilibrium locus is downward sloping since the preference of the marginal resident joining the city decreases as city size increases due to taste heterogeneity (Arnott and Stiglitz, 1979; Moretti, 2011). The slope of the migration equilibrium locus is inversely related to the migration elasticity γ^{θ}. Higher QoL A_i^{θ} shifts the migration equilibrium locus upwards. The intersection of both equilibrium loci is the only combination of real wages and employment that satisfies all SSE equilibrium conditions and, hence, we can use it to quantify the model and derive QoL premiums.

The two vertical dashed lines mark two cities of different size with $L_2 > L_1$. Housing productivity η_i is higher in the larger city, which gives the city an edge in the competition for workers since the housing sector provides more housing at the same equilibrium price (HH_2 is below HH_1). Yet, despite the housing productivity advantage, the city size differential can only be rationalised by a greater labour supply in the larger city and an upward-shifted migration equilibrium locus (LL_2 vs. LL_1). Intuitively, the lower idiosyncratic amenity of the marginal resident must be compensated for by a higher average group-specific QoL A_i^{θ} in the larger city. Hence, there is a positive urban QoL premium.

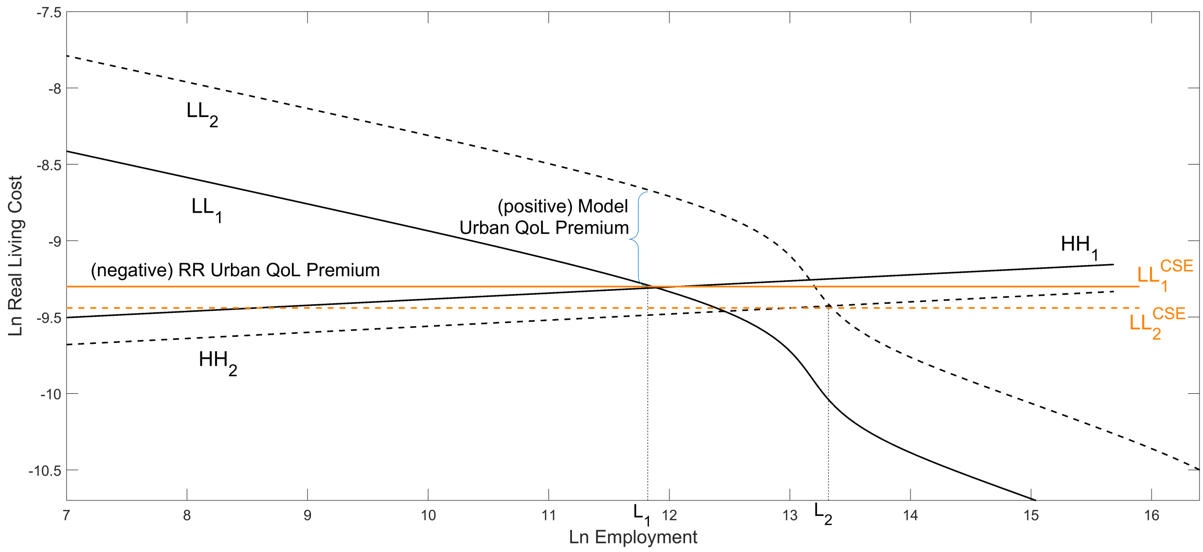
With decreasing taste heterogeneity, the migration elasticity γ^{θ} increases, the slope of the migration equilibrium flattens, and the urban QoL premium shrinks. For the limit case $\gamma^{\theta} \to \infty$, our model nests the canonical CSE framework in which the migration equilibrium locus is simply a horizontal line shifted by $\mathcal{A}_i^{\theta}$ (see Eq. (15)). The corresponding migration equilibrium schedules are described by LL_i^{CSE}. In the given example, because the larger city has a fundamental housing productivity advantage, we qualitatively misrepresent the urban QoL premium if we abstract from taste heterogeneity.

The important takeaway is that the urban QoL premium in the DSM with taste heterogeneity is necessarily more positive than in the canonical spatial equilibrium framework unless the migration elasticity γ^{θ} is large. More generally, we necessarily recover larger QoL differentials from a model with taste heterogeneity. Since, consistent with the literature (Caliendo et al., 2019b), we estimate relatively low values of γ^{θ} for all groups, we expect our quantitative framework to deliver larger valuations of local non-marketed goods than the canonical Rosen-Roback framework.

D Quantification

The quantification of the model consists of two steps. First, we obtain values of the structural parameters $\{\alpha,\, \beta,\, \rho,\, \iota,\, \gamma^{\theta},\, \kappa^{\theta},\, B_{ij,t}^{\theta},\, \tau_{ij}^{\theta}\}$. We borrow $\{\alpha, \iota, \rho\}$ from the literature and estimate the remaining parameters using variables observed in data $\{L_{i,t}^{\theta}, \bar{T}_i, w_{i,t}^{\theta}, \chi_{ij|i,t}^{\theta}, p_{i,t}\}$ and the structure of the model. Second, we use data, the estimated parameter values, and the structure of the model to invert the structural fundamentals $\{\psi_{i,t}^{\theta},\, \eta_{i,t},\, A_{i,t}^{\theta}\}$ and to solve for the region-group-time-specific employment vector $\mathbf{L}_{\mathbf{i},\mathbf{t}}^{\theta}$ that references the dynamic equilibrium.

Figure 2: Urban quality of life premium



Notes: A formal derivation of demand and supply shifters and elasticities is in Appendix J.5. We use parameter values $\gamma = 0.5$ and $\beta = 0.2$ which are within the range of estimates in the literature and our own estimates in Section D. We use the structural fundamentals quantified in Section D. To ease the presentation, we derive all curves for one worker group (middle-aged, skilled male workers) exclusively.

D.1 Data

As an empirical correspondent to locations indexed by i in the model, we choose 141 German labour market regions defined by Kosfeld and Werner (2012) based on commuting data. The centre of a labour market region is the municipality with the largest number of workers. We treat periods t in our model as years in the data. We briefly discuss the sources and processing of our data below and refer to Appendix K.1 for details.

Employment. Our measure of employment $L_{i,t}^{\theta}$ is constructed from the Employment History (BeH) covering the years 1993-2017. This dataset is provided by the Institute of Employment Research (IAB) and contains information on the universe of employees in Germany (with the exception of civil servants and the self-employed) on a daily basis. We only select those workers who are employed subject to social security contributions (including apprentices) and compute region-year-specific employment levels for different groups which are defined according to the interactions between sex, three skill categories (no apprenticeship, completed apprenticeship and tertiary education) and three age categories (16-30 years, 31-50 years and 51-65 years).

Migration. We assign workers to labour market regions using their place of employment as reported in the BeH. Bilateral group-specific migration flows $M_{ij,t}^{\theta}$ are then constructed by computing the number of workers belonging to group θ who were employed in region i in year t and in region j in year $t+1$. The conditional migration probabilities are then observed as $\chi_{ij|i,t}^{\theta} = M_{ij,t}^{\theta}/L_{i,t}^{\theta}$.

Wages. We follow the standard approach in labour and urban economics and identify the region-group-year wage $w^{\theta}_{i,t}$ from movers by regressing individual wages against region-group-year fixed effects, controlling for individual fixed effects (Abowd et al., 1999; Combes et al., 2008). We use matched employer-employee data including nominal wages from the IAB covering the universe of German workers and establishments from 1993 to 2017.

Rents. We follow Combes et al. (2019) and compute a house price index for a representative property at the centre of a labour market area. Assuming a monocentric region, this is the only location where we can abstract from commuting costs when inferring QoL (Albouy and Lue, 2015). The price index maps into rent p_i via a constant cap rate of 0.035 (Koster and Pinchbeck, 2021). The property micro data we use is from Immoscout24 covering more than 16.5 million sales proposals for apartments and houses between 2007-2017. The data were accessed via the FDZ-Ruhr (Boelmann and Schaffner, 2019).

Geographic variables. We use a geographic information system (GIS) to compute the land area $\bar{T}_i$ of all regions and the great circle distance between all pairs of regions. For a cultural distance measure, we use the inverse of the county-based dialect similarity index by Falck et al. (2012), which we aggregate to labour markets.

Big data. We use social media data from Facebook, Flickr, and Picasa to approximate regional amenity value and social connectedness. We use those data to over-identify estimated structural parameters and inverted structural fundamentals.

Location characteristic. For our policy application, we collect the concentration of particular matter (PM^{10}), the spatial distribution of coal deposits, the locations of coal power plants, and the distribution of winds by direction for all regions. We also collect a comprehensive data set on fundamental first-nature characteristics that potentially affect productivity (e.g. access to navigable rivers), amenity (e.g. opera houses, World War II destruction), and housing TFP (e.g. physical constraints to development).

D.2 Structural parameters

We set the housing expenditure share to $1 - \alpha = 0.33$, which is in line with a literature summarised in Ahlfeldt and Pietrostefani (2019) and official data from Germany (Statistisches Bundesamt, 2020). We use a tax rate of $\iota = 0.49$ which incorporates social insurance contributions that are proportionate to income in Germany (OECD, 2017). Likewise, we set the intertemporal discount rate to $\rho = 0.11$ following the economics literature on time-preferences (Moore and Viscusi, 1988; Frederick et al., 2002). Lastly, we impose that stayers face no migration cost ($\tau^{\theta}_{ij=i} = 0$).

We estimate all other parameters using estimation equations that we derive from the structure of the model. For identification, we generally follow the current best-practice examples in the respective fields. Our main empirical contribution is to exploit our rich data to account for greater inter-group heterogeneity than in previous work. We briefly discuss the parameter values along with references to the identification strategies and the relevant literature below. For a formal derivation of all estimation equations and full estimation results we refer to Appendix K.2.

Density elasticity of productivity (κ^θ). The estimating equation for κ^θ is a log-linearised version of Eq. (5). Identification comes from between-labour-market-area movers and is conditional on individual effects (Combes et al., 2008). We use a 100-year lag of population density following a literature that argues that production fundamentals that determined productivity in history are no longer relevant today (Ciccone and Hall, 1996). With this approach, we estimate the agglomeration elasticity for $\Theta = 18$ groups and find that returns to agglomeration (κ^θ) are not only biased with respect to skills (Baum-Snow and Pavan, 2013), but also gender, with women benefiting more from agglomeration. The weighted average elasticity estimate of 0.024 is close to the typical result in the literature (Combes and Gobillon, 2015).

Land share (β). The estimating equation for β is a log-linearised version of Eq. (8). The estimation equation is similar to the one in Combes et al. (2019), although, following from our general equilibrium setting, the main independent variable is GDP density rather than population. Following the literature we, again, use the 100-year lag of population density as an instrument. Our estimate of $\beta = 0.18$ implies a population density elasticity of house prices of 0.2, which is within the typical range in the literature (Ahlfeldt and Pietrostefani, 2019). The implied intensive-margin housing supply elasticity $(1 - \beta)/\beta = 4.2$ is close to existing structural estimates (Epple et al., 2010).

Migration elasticity (γ^θ). The estimating equation for γ^θ is a log-linearised and spatially differenced version of Eq. (10) in which leading migration probabilities control for future utility flows according to the Bellmann's principle (Artuç et al., 2010). We follow the literature and estimate γ^θ using GMM. In our preferred approach, we restrict the identifying variation to lagged group-specific average wage differences between eastern and western states that likely capture a legacy of the cold-war era. The estimated average elasticity of 0.3 is somewhat larger than when we use the standard IVs (lagged wage and migration probabilities), but somewhat smaller than previous estimates for the U.S. (Caliendo et al., 2019a). Novel to the literature using this estimation approach, we find that middle-aged and middle-skilled male workers are those that are most responsive to economic migration incentives.

Migration costs (τ^θ_{ij}). The estimating equation for τ^θ_{ij} is a log-linearised version of Eq. (10) using a PPML estimator. Destination-group-year and origin-group-year effects control for arbitrary pull factors and multilateral resistance (Head and Mayer, 2014). Exploiting the panel-dimension, origin-destination-time effects non-parametrically identify origin-destination-group-specific migration resistance $\tau^\theta_{ij} \times \gamma^\theta$ up to a constant. Exploiting the no-internal-migration-cost constraint $\tau^\theta_{i,j=i} = 0$, we derive theory-consistent estimates of τ^θ_{ij} for given values of γ^θ.

Female, old, and middle-skilled workers have the largest resistance to migrate. Yet, middle-skilled workers experience low migration costs. Because their tastes are relatively homogeneous (large γ^θ), small differences in migration costs rationalise large differences in migration flows. In monetary terms, the weighted average migration cost corresponds to about €170K which is more than revealed in survey-based research for the average U.S. citizen, though much less than for those who report themselves as "rooted" (Koşar et al., 2021).

Bilateral amenity. The estimating equation for B^θ_{ij} is the same gravity migration equation from which we infer migration resistance $\tau^\theta_{ij} \times \gamma^\theta$. For given values of γ^θ, we infer B^θ_{ij} from the structural residual. Consistent with theory, we rationalise migration flows of zero by setting $B^\theta_{ij} = 0$.

D.3 Structural fundamentals

Labour and housing productivity. Given our estimates of the agglomeration elasticity κ^θ and observed wages $w^\theta_{i,t}$, regional employment $\sum_\theta L^\theta_{i,t}$, and land area $\bar{T}_i$, we invert fundamental labour productivity $\psi_{i,t}$ using the first-order condition of labour demand, Eq. (6). Likewise, we use our estimate of the land share β and observed rents $p_{i,t}$, output $\sum_\theta w^\theta_{i,t} L^\theta_{i,t}$ and land area $\bar{T}_i$ to invert fundamental housing productivity $\eta_{i,t}$ using housing market clearing, Eq. (8).

Quality of life. Owing to the dynamic structure of our model, the inversion of QoL $A^\theta_{i,t}$ is less straightforward. Given observed data on conditional migration probabilities $\chi^\theta_{ij|i,t}$ and estimates of bilateral amenities $B^\theta_{ij,t}$, migration costs τ^θ_{ij} and the migration elasticity γ^θ, we can invert the within-group QoL $\bar{A}^\theta_{i,t}$ up to the group-year constant ζ^θ_t for a given dynamic employment vector $\mathbf{L^\theta_{i,t}}$ that determines future wages $w^\theta_{i,t+s}$ and rents $p_{i,t+s}$ (see Section C.4) using the migration gravity Eq. (10). However, to forecast $\mathbf{L^\theta_{i,t}}$ using the dynamic structure of the model, we require values of $\bar{A}^\theta_{i,t}$ that feed into labour supply, Eq. (12), via the migration gravity Eq. (10).

Therefore, we solve for the endogenous employment vector $\mathbf{L^\theta_{i,t}}$ that references the dynamic spatial equilibrium and the exogenous structural fundamental $\bar{A}^\theta_{i,t}$ simultaneously. To this end, we use a nested dynamic programming algorithm to which we refer as *dynamic solver* for convenience. The dynamic solver consists of three components. First, a fixed-point programming algorithm (FP) that delivers a numerical solution for $\bar{A}^\theta_{i,t}$ (output) for given guessed values of $\mathbf{L^\theta_{i,t}}$ (input) using Eq. (10). Second, a dynamic programming algorithm (DP) which forecasts $\mathbf{L^\theta_{i,t}}$ (output) for guessed values of $\bar{A}^\theta_{i,t}$ (input). The DP iterates over Eqs. (10), (12), (6) and (8) to forecast $\chi^\theta_{ij|i,t+s}$, $L^\theta_{i,t+s+1}$, $w^\theta_{i,t+s+1}$, and $p_{i,t+s+1}$ which feed into $\chi^\theta_{ij|i,t+s+1}$ and $L^\theta_{i,t+s+2}$ until $L^\theta_{i,t+S}$ is stationary. Third, an outer loop (OL) that nests the two other algorithms and forwards the outputs from the FP as inputs to the DP and vice versa.

In taking the dynamic solver to the data, we set a time horizon of $H = 1000$ years which exceeds the transition period to the SSE in all our applications. As initial guesses for the employment vector $\mathbf{L^\theta_{i,t}}^{\mathbf{0}}$ we use the values we observe in year t for which the model is being quantified:

$$\mathbf{L}_{\mathbf{i,t}}^{\theta}{}^{\mathbf{0}} = \underbrace{L_{i,t}^{\theta}, L_{i,t}^{\theta}, ..., L_{i,t}^{\theta}}_{H \text{ elements}}.$$

Given $\mathbf{L}_{\mathbf{i,t}}^{\theta}{}^{\mathbf{0}}$, we use the FP to generate starting values $\bar{A}_{i,t}^{\theta}{}^{0}$. We then start the OL where in each iteration l the DP delivers an output $\mathbf{L}_{\mathbf{i,t}}^{\theta}{}^{l}$ that is an input into the FP which in turn delivers $\bar{A}_{i,t}^{\theta}{}^{l+1}$ as an output that serves as an input into the PF in the next iteration. In this OL, we treat $\mathbf{L}_{\mathbf{i,t}}^{\theta}$ and $\bar{A}_{i,t}^{\theta}$ as fixed points that are found in iteration $\mathcal{L}$ when the input into the FP corresponds to the output from the DP and vice versa. Once the OL converges, we crop $\mathbf{L}_{\mathbf{i,t}}^{\theta}{}^{\mathcal{L}}$ to the $S^{\mathcal{L}}$ elements forecasted by the DF in the last iteration of the OL. Saving $A_{i,t}^{\theta}{}^{\mathcal{L}}$ concludes the quantification of the model. For further details on the dynamic solver, we refer to Appendix K.3.

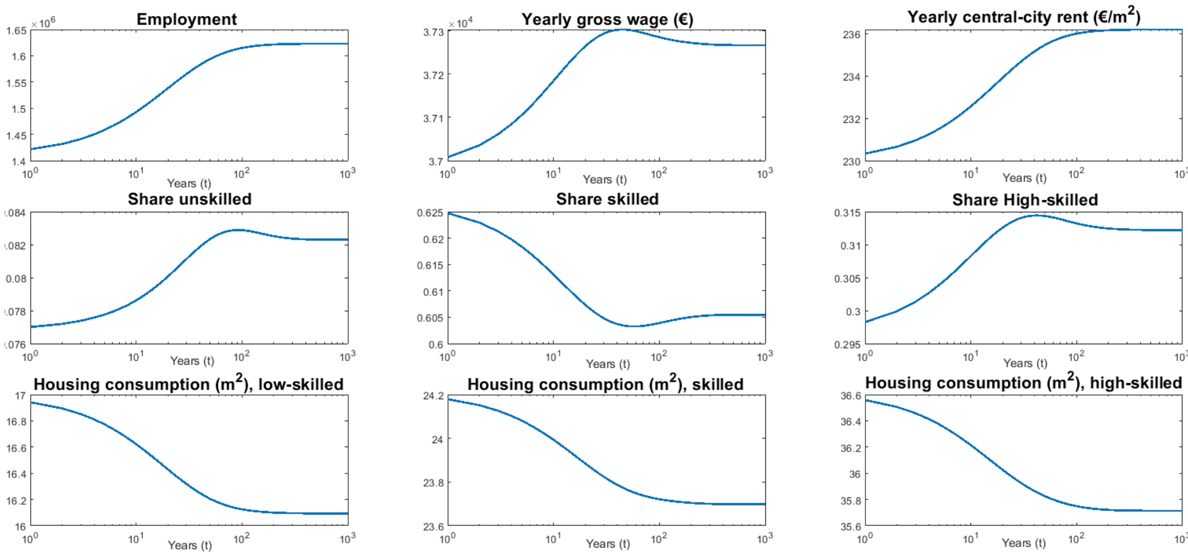
D.4 Transition into the stationary spatial equilibrium

Figure 3 exemplarily illustrates the transition path from the TSE observed in our data into the SSE found by the dynamic solver introduced in Section D.3. For Berlin, Germany's largest local labour market, the model forecasts that employment would grow by 14% as the economy transitions into the SSE if there were no further shocks to fundamentals. The average wage would increase by about 0.8%. This would be more than the agglomeration-induced productivity effect and driven by a 5.4%-increase in the high-skilled share. The increase in housing demand would map to a higher house price and lower housing consumption for all skill groups.

While it takes more than 700 years for group-region employment to become stationary, almost all of the adjustment in Berlin takes place within the first 100 years. Zooming out, we find that the sum of the absolute difference between TSE and SSE values across all groups and regions shrinks by about 35% within the first 10 years, and by about 65% within the first 30 years, with some variation depending on the outcome (see Figure A8 in Appendix K.4).

The main takeaway from the aggregate outcomes in Table 1 is that during the transition into the SSE workers of all skill groups relocate to local labour markets with higher QoL, but lower density, on average. This tendency is strongest for the unskilled. The effect of relocating to lower-density labour markets dominates the QoL-effect on housing cost, resulting in a slight increase in housing consumption. In contrast, the high-skilled tend to remain in denser labour markets, so that the effect of sorting into higher QoL labour markets dominates and housing consumption decreases. The reduction in the weighted average density by 4% leads to a mild reduction in aggregate output owing to lower agglomeration economies. A comparison of the TSE to the SSE at the regional level reveals an increase in employment in the eastern states by nearly one million workers (at the expense of the western states), partially offsetting domestic migration during the first 25 years after the end of the Cold War era. This increase in employment in the eastern states drives rents, but does not map to higher average wages due to a moderate decrease in the high skilled share (see Appendix Section K.4).

Figure 3: Transition from TSE into SSE in Berlin



Notes: Model-based forecasts using the dynamic solver introduced in Section D.3. 2017 starting values. Yearly gross wage, skill shares and housing consumption are weighted by group shares.

D.5 Overidentification

To subject the model-derived structural parameters and fundamentals to a reality check, we correlate fundamental labour productivity $\psi^{\theta}_{i,t}$, fundamental housing productivity $\eta_{i,t}$ and migration costs τ^{θ}_{ij} with observable characteristics not used in the quantification of the model. The results are generally plausible. As an example, fundamental labour productivity is lower in the eastern states, likely a legacy of the Cold War era, and where tradable services are over-represented. Housing productivity is low in the mountainous region near the Alps where the geography is less favorable for development. Migration costs increase in geographic and social distance, consistent with greater costs of rebuilding social capital. Since the structural fundamental $A^{\theta}_{i,t}$ is the focus of our analysis, we explore the correlation with observable characteristics more extensively in the next section.

Inverting the model from the TSE observed in $t = 2007$, we find that the model-based forecasts of employment $L^{\theta}_{i,t+s}$ over the 2007-2017 period are positively correlated with observed employment data. Conditional on region and year effects, a log-point increase in the out-of-sample forecast of regional employment is associated with a 0.75-log-point increase in observed employment, with a standard error of just 0.03. Hence, the model successfully captures a mean reversion tendency that is a feature of the data. We refer to Appendix Section K.5 for estimation results and a detailed discussion.

E Quality of life

In this section, we illustrate the spatial variation in the within-group measure of QoL, $\bar{A}^{\theta}_{i,t}$, inverted from the DSM (DSM-QoL) and how it correlates with a range of amenity measures

Table 1: TSE vs. SSE

Outcome	TSE	SSE	Ratio
Output in bn.	1.058	1.056	0.998
QoL index	1.616	1.639	1.014
Weighted average density (emp./km^2)	150.710	144.462	0.959
QoL index, unskilled	2.169	2.272	1.048
QoL index, skilled	1.412	1.414	1.002
QoL index, high-skilled	2.149	2.218	1.032
Weighted density, unskilled	167.221	161.940	0.968
Weighted density, skilled	143.493	136.149	0.949
Weighted density, high-skilled	170.666	168.040	0.985
Yearly wage (€), unskilled	23222	23239	1.001
Yearly wage (€), skilled	33804	33722	0.998
Yearly wage (€), high-skilled	50784	50773	1.000
Yearly housing cost (€/m^2), unskilled	132.364	133.601	1.009
Yearly housing cost (€/m^2), skilled	121.136	120.420	0.994
Yearly housing cost (€/m^2), high-skilled	150.365	152.752	1.016
Housing consumption m^2, unskilled	43.353	43.656	1.007
Housing consumption m^2, skilled	70.165	70.489	1.005
Housing consumption m^2, high-skilled	86.407	85.577	0.990

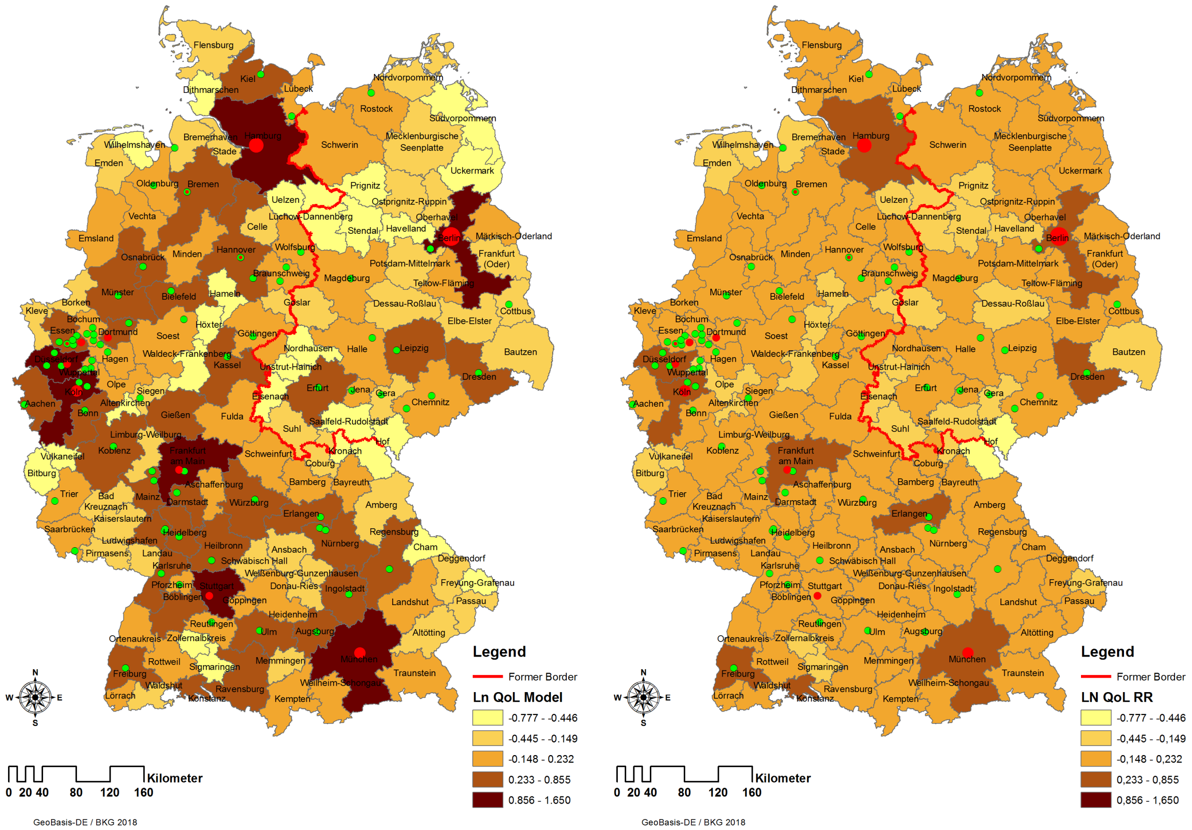
Notes: TSE values observed in the data except for QoL which is inverted using the dynamic solver introduced in Section D.3. All SSE values are model-based forecasts of the dynamic solver. QoL index is normalised within-group measure $\bar{A}^{\theta}_{i,t}$, weighted by group-region employment $L^{\theta}_{i,t}$

typically employed in the literature as well as a composite amenity index derived from 'big data'. We provide a comparison to a Rosen-Roback type QoL measure $\mathcal{A}^{\theta}_{i,t}$ (RR-QoL) and evaluate how the migration elasticity γ^{θ} moderates the relationship between the two QoL measures.

E.1 Spatial variation in quality of life

Two important stylised facts arise from a comparison of the two QoL measures in Figure 4. First, the spatial distribution of QoL is similar, which is arguably reassuring. In particular, there is a positive urban QoL premium. Large labour markets in Germany are not only good places to work, but also good places to live. Second, there is significantly more variation in DSM-QoL than in the canonical RR-QoL. This is consistent with our theoretical analysis in Section C.7 and substantiated by Figure 5 which correlates the two QoL measures across regions allowing for inter-group heterogeneity. For all 18 groups, RR-QoL increases less than proportionately in DSM-QoL, confirming our theoretical prior that the canonical framework understates QoL differentials if the migration elasticity γ^{θ} is low. The bias is quantitatively large as group-specific regressions of $\ln \mathcal{A}_{i,t=2017}$ against $\ln \bar{A}_{i,t=2017}$ yield point estimates in the range of 0.16-0.45, with an unweighted mean of 0.27 (see Table A14 in Appendix L).

Figure 4: Spatial variation in quality of life



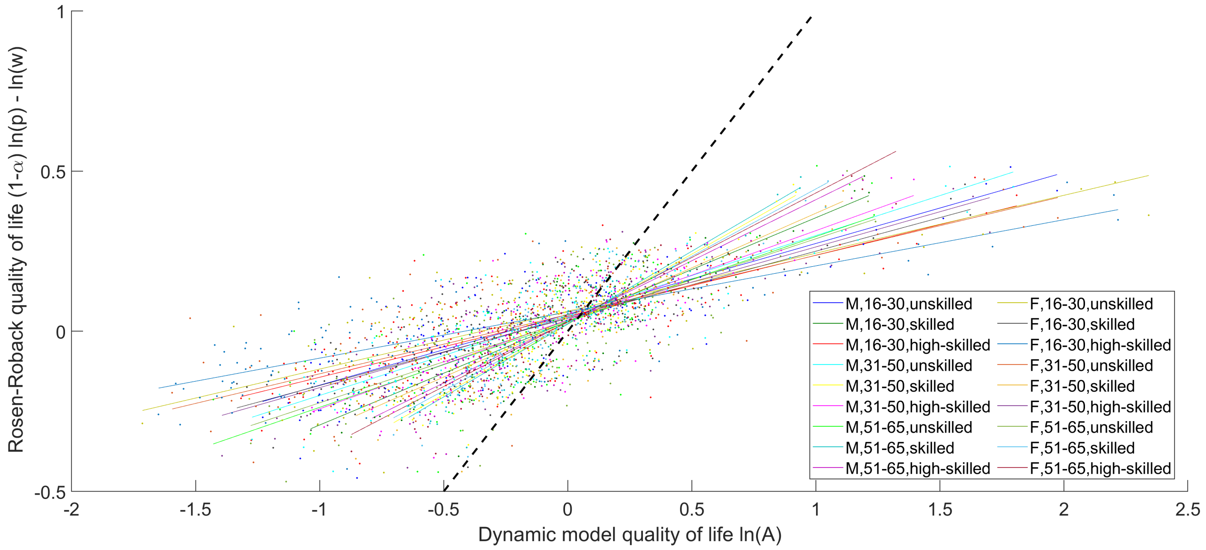
(a) Dynamic model ($\bar{A}_i$) (b) Rosen-Roback ($\mathcal{A}_i$)

Note: Unit of observation is 141 labour market areas as defined by Kosfeld and Werner (2012). Group adjustment in auxiliary regressions of ln(QoL) against group and region fixed effects, the latter being shown on the maps.

E.2 Determinants of quality of life

Since Roback (1982), it is conventional to regress inverted QoL measures against regional amenity variables to infer the value of amenities. In Table 2, we illustrate how the larger variation in the DSM-QoL leads to larger utility effects of regional amenities. We begin by considering the number of geo-tagged photos shared in social media as a "big data" composite amenity index that was originally proposed by Ahlfeldt (2013) and has gained popularity recently (Gaigné et al., 2017; Saiz et al., 2018; Carlino and Saiz, 2019). This measure assumes that social media users share visually appealing content (e.g. distinctive architecture or scenic views) or interesting activities (e.g. hiking tours or restaurant visits) that are related to a location's endowment with amenities (see Appendix K.1.7 for details). For the purpose of overidentification of our DSM-QoL, the appealing feature of the big data amenity index is that it does not rely on an arbitrary selection of observable characteristics that are more or less readily available. A simple bi-variate log-linear pooled cross-sectional regression (excluding group, region, or year effects) of the DSM-QoL on the amenity index explains almost 60% of

Figure 5: Rosen-Roback vs. DSM QoL estimates



Notes: 2017 values. Unit of observation is region-group. Thick dashed line is the 45-degree line. Model-based amenity inverts QoL from a TSE assuming that agents have perfect foresight. Rosen-Roback assumes that the economy is in a SSE without spatial frictions. We tabulate the slope parameters of the log-linear fits in Table A14 in the appendix.

the variation (Column 1). This high correlation simultaneously lends support to the DSM-QoL and suggests that big data can be a similarly powerful predictor of QoL as lights at night are for GDP (Henderson et al., 2012). The point estimate in Column (1) has a structural interpretation in that it is the inverse of the QoL elasticity in the photo production function ($1/0.356 = 2.81$), but it seems fair to assume that this large estimate is to some extent driven by high QoL regions being more populated (see Appendix L.2 for further discussion).

In the next two columns, we use DSM-QoL in 2007 (Column 2) and 2017 (Column 3) as dependent variables and add traditional amenity measures as explanatory variables, taking inspiration from a literature that has been concerned with the role of city size (Albouy, 2011), climate (Roback, 1982), crime (Linden and Rockoff, 2008), air pollution (Chay and Greenstone, 2005), as well as natural and consumption amenities (Glaeser et al., 2001). We use three supra-regional dummy variables to capture the effects of fresh and rainy summers (near coast), cold winters (near Alps), and the legacy of the Cold War era (East), none of which exhibits precisely estimated effects. There is no persistent QoL effect of World War II bombings, consistent with rapid mean reversion in city size documented by Brakman et al. (2004). We also do not find significant effects for crime or bodies of water, likely because of limited variation across German regions.

In contrast, the positive urban QoL premium suggested in Figure 4 is substantiated by a precisely estimated employment elasticity of QoL of about 0.45 (Columns 2 and 3). The employment effect on the DSM-QoL is about four times as large as on the RR-QoL (Columns 5 and 6), the latter being larger than found by Albouy (2011) for the U.S., but close to the residential spillover effect found by Ahlfeldt et al. (2015) for Berlin. This comparison

Table 2: Quality-of-life determinants

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln($\bar{A}_i^\theta$)	Ln($\bar{A}_i^\theta$)	Ln($\bar{A}_i^\theta$)	Ln($\mathcal{A}_i^\theta$)	Ln($\mathcal{A}_i^\theta$)	Ln($\mathcal{A}_i^\theta$)
	All	2007	2017	All	2007	2017
Ln big data amenity	0.356***	0.114***	0.129***	0.054**	0.064***	0.058**
(residualised)	(0.02)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)
Ln employment		0.409***	0.455***		0.096***	0.123***
		(0.04)	(0.05)		(0.02)	(0.02)
Near Alps (dummy)		-0.068	-0.016		-0.009	0.054
		(0.06)	(0.08)		(0.05)	(0.06)
Near coast (dummy)		-0.090^{+}	-0.050		-0.007	0.011
		(0.06)	(0.06)		(0.04)	(0.04)
East (dummy)		-0.025	-0.024		0.037	0.008
		(0.06)	(0.06)		(0.03)	(0.04)
Ln crime per capita		0.027	-0.032		-0.032	-0.063
		(0.06)	(0.07)		(0.04)	(0.04)
Ln pollution		-0.302*	-0.402**		-0.148	-0.223^{+}
concentration (pm10)		(0.16)	(0.19)		(0.10)	(0.13)
Housing stock destroyed		-0.001	-0.001		-0.000	-0.001
in WWII (%)		(0.00)	(0.00)		(0.00)	(0.00)
# Opera houses		0.059**	0.051*		0.009	0.010
		(0.02)	(0.03)		(0.01)	(0.02)
Ln water area		0.063*	0.064^{+}		0.024	0.030
		(0.03)	(0.04)		(0.02)	(0.02)
Ln area		-0.072^{+}	-0.085^{+}		-0.005	-0.035
		(0.05)	(0.05)		(0.03)	(0.04)
Group effects	-	Yes	Yes	-	Yes	Yes
Observations	27918	2538	2538	27918	2538	2538
R^2	.593	.737	.721	.0379	.458	.459

Notes: Unit of observation is group-region. OLS estimation. Ln($\bar{A}_i^\theta$) is the region-group amenity shifter in the DSM developed in this paper. Ln($\mathcal{A}_i^\theta$) is the region-group amenity shifter implied by the Rosen-Roback framework (see section C.4). Standard errors clustered on regions in (1) and (4) and on regions and groups in all other columns. Big data amenity is the log of the number of geotagged photos shared on social media (flick and picasa) residualised in regressions against all other covariates reported in a column. $^{+}$ $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

highlights how in a quantitative model with preference heterogeneity, a high fundamental QoL is required to rationalise why, for example, Berlin has almost 10 times the employment of the average labour market. For Germany, the urban QoL premium is much larger than even the unadjusted urban wage premium (see Figure 1, panel a), let alone the skill-adjusted urban wage premium (see Section D.2).

The pollution effect illustrates how the same logic extends to non-marketed goods of immediate policy interest. Descriptively, the DSM-QoL decreases in the concentration of particulate matter at an elasticity of -0.4 (we turn to causal effects in Section F). For the RR-QoL the estimated elasticity is not even half as large. Hence, the case for preserving clean air is significantly stronger if we account for frictional migration. This finding is consistent with previous evidence by Luechinger (2009), who finds larger pollution effects on life satisfaction than on house prices, and Bayer et al. (2009), who show that the willingness to pay for clean air is larger in a discrete choice model allowing for mobility cost than in a conventional hedonic model. The same conclusion extends to cultural amenities as opera houses are more strongly positively associated with the DSM-QoL measure than with the RR-QoL measure.

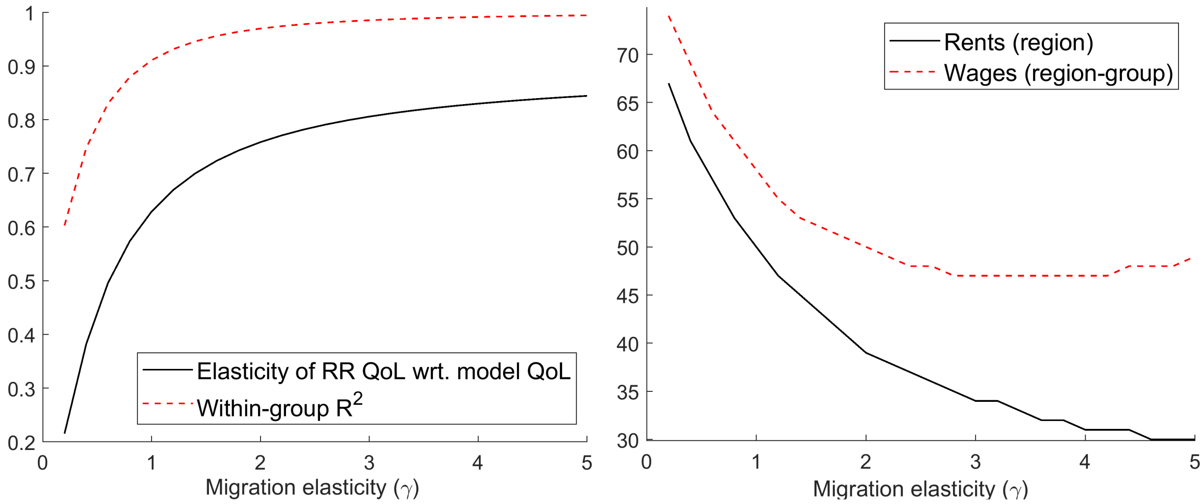
To ensure that the big data amenity captures the effects of unobserved QoL determinants, exclusively, we residualise the measure in auxiliary regressions against all covariates in Columns (2-3) and (5-6). While the point estimate drops, in particular in the DSM-QoL models, it remains statistically and economically significant, highlighting the role big data can play in controlling for QoL determinants that are difficult to observe.

E.3 The role of the migration elasticity

To evaluate how sensitive the relationship between DSM-QoL and RR-QoL is to the choice of the migration elasticity, we quantify the model under varying group-independent values for γ. The left panel of Figure 6 confirms the theoretical expectation that DSM-QoL approaches RR-QoL for large values of γ (see Section C.7). If one is willing to believe that $\gamma \geq 3$, the elasticity of RR-QoL with respect to DSM-QoL exceeds 0.8. The R^2 of a log-linear regression then exceeds 0.9. For smaller values suggested by the empirical literature, however, a small change in the set or estimated value of the migration elasticity can have large effects on inverted QoL.

Since for observed migration probabilities, migration resistance $\gamma^\theta \times \tau^\theta_{ij}$ is exactly identified by the migration gravity Eq. (10) (see Appendix K.2.4 for the empirical counterpart), an increase in γ^θ implies a proportionate decrease in migration costs τ^θ_{ij}. Lower $\tau^\theta_{ij,j\neq i}$ relative to $\tau^\theta_{ij,j=i} = 0$ imply larger between-city migration flows, with implications for the speed of spatial adjustments. The right panel of Figure 6 illustrates the negative relationship between γ and the years it takes until 80% of the transition to the SSE are completed. In terms of rents, which are log-proportionate to city employment, the adjustment period shrinks from more than 65 years to 30 years. In terms of group-region wages, we observe a decrease from close to 75 years to less than 55 years.

Figure 6: The role of the migration elasticity (γ)



(a) Fit of Rosen-Roback with model QoL

(b) Years until 80% convergence to SSE

Notes: Elasticity estimates and within-R^2 are from regressions of ln RR-QoL ($\mathcal{A}$) against ln DSM-QoL (A), controlling for group fixed effects. An increase in γ implies a decrease in migration cost $\tau^{\theta}_{i,j\neq i}$ since the migration resistance $\gamma \times \tau^{\theta}_{i,j}$ is exactly identified by the gravity migration equation. Convergence to the SSE is measured in terms of a reduction in the sum of the absolute difference between TSE and SSE values. In all iterations, the model is quantified using 2017 values observed in the data.

F Policy evaluation

In this section, we outline how to use the quantified model for the evaluation of policies that seek to improve regional QoL. The first step is to establish a causal relationship between the structural fundamental $A^{\theta}_{i,t}$ and some QoL determinant that is amenable to policy-induced change. This challenge is shared with a reduced-form literature exploring capitalisation effects of QoL determinants in house prices or inverse real wages. The second step is novel to the QoL literature. Starting from the SSE, we use the causal estimate from the first step to update $A^{\theta}_{i,t}$, and then re-solve for a counterfactual SSE. A comparison between the initial and the counterfactual SSE delivers general equilibrium comparative statics that account for aggregate effects as well as inter-group and inter-region distributional effects. Unlike in the canonical CSE that is anchored by a spatially invariant reservation utility, spatial policies can have positive or negative effects on expected regional utility.

Our case in point is an improvement in air quality. Air pollution causes 400 thousand premature deaths per year in the EU and is by far the number one environmental factor driving disease (European Environment Agency, 2020). Negative effects of dirty air on health (Deryugina et al., 2019), property prices (Chay and Greenstone, 2005; Bayer et al., 2009) and self-reported life satisfaction (Luechinger, 2009; Levinson, 2012) are well established. Our policy counterfactual is a reduction in PM^{10} concentration in the most polluted regions to the 75$^{\text{th}}$ percentile in the distribution across all regions. Since this application is intended to serve as an illustrative example, we keep the estimation strategy and the policy experiment simple and transparent. For future applications, researchers are, of course, invited to expand on our application, e.g. by exploiting natural experiments or randomised policies for identification, or considering more sophisticated policy interventions in the counterfactuals. Naturally, the procedure outlined below can be applied to any other QoL determinant or, more generally, any determinant that affects any of the structural fundamentals $\{\psi^{\theta}_{i,t}, \eta_{i,t}, A^{\theta}_{i,t}\}$ in the model.

F.1 Procedure

Transition to counterfactual SSE. Adopting the conventional *exact hat algebra* notation where hats represent ratios of counterfactual values over initial values (Dekle et al., 2007),[6] we model a policy as an exogenously induced relative change in QoL $\widehat{A^{\theta}_{i,t}} = \widehat{A^{\theta}_{i,t}}(b^{\theta}\widehat{\mathcal{X}_{i,t}})$ that results in a counterfactual QoL $A^{\theta C}_{i,t} = \widehat{A^{\theta}_{i,t}} A^{\theta}_{i,t}$. $\widehat{\mathcal{X}_{i,t}}$ is a relative change in an exogenous QoL determinant and b^{θ} is a group-specific parameter that describes a causal relationship between $A^{\theta}_{i,t}$ and $\mathcal{X}_{i,t}$.

Starting from the initial SSE, we use a simplified version of the dynamic solver introduced in Section D.3 that takes $A^{\theta}_{i,t}{}^{C}$ as given to solve for a counterfactual SSE. As with the initial SSE, the counterfactual SSE is referenced by stationary employment $L^{\theta}_{i,t}{}^{C}$ that maps to the other endogenous variables as discussed in Section C.4. The transition into the counterfactual SSE is moderated by a sequence of migration flows that restore the SSE through the model-endogenous agglomeration and dispersion forces. The comparison of the initial and the counterfactual SSE delivers a policy effect that is causal in the sense that it is not confounded by the mean-reversion tendency of a spatial economy in the TSE. Hence, our approach yields results that are comparable to the comparative statics employed for economic policy evaluation in static models.

Welfare. Consider a social planner that extrapolates the expected indirect utility of stayers in the SSE into the infinite future to create a group-region welfare measure:

$$\mathcal{R}^{\theta}_{i,t} = \frac{V^{\theta}_{i|i,t}}{\rho} = \frac{(1-\iota)}{\rho} \frac{w^{\theta}_{i,t}}{p^{1-\alpha}_{i,t}} A^{\theta}_{i,t} \exp\left[\ln B^{\theta}_{ii,t} - \tau^{\theta}_{ii}\right] = \frac{(1-\iota)}{\rho} \frac{w^{\theta}_{i,t}}{p^{1-\alpha}_{i,t}} A^{\theta}_{i,t} \tag{16}$$

Since unlike in the canonical CSE framework utility is not equalised across regions in our DSM, it is particularly important to specify the social welfare function when aggregating group-region-specific welfare. We define a social welfare function in the tradition of Atkinson (1970) as

$$\mathcal{W}_t(\varepsilon) = \frac{1}{1-\varepsilon} \sum_i \sum_{\theta} \left(\mathcal{R}^{\theta}_{i,t}\right)^{1-\varepsilon} \frac{L^{\theta}_{i,t}}{\bar{L}_t} = \mathcal{R}_t\left(1 - \mathcal{I}_t(\varepsilon)\right), \tag{17}$$

where $\mathcal{R}_t$ is the weighted average of group-region utility and $\mathcal{I}_t \in [0, 1]$ represents the Atkinson measure of inequality (see Appendix M.1 for derivation details). This formulation separates social welfare into a scale-dependent part (average utility) that enters positively into social

[6] See Heblich et al. (2020b) for a recent application.

welfare and a scale-independent inequality measure that imposes a penalty on inequality. The strength of the penalty is governed by the inequality aversion parameter $0 \leq \varepsilon \neq 1$. If $\varepsilon = 0$, $1 - \mathcal{I} = 1$, such that social welfare is solely determined by the aggregate (utilitarian case). The inequality penalty increases in ε, with $\varepsilon \to \infty$ representing the limiting Rawlsian case in which the penalty is entirely determined by the weakest region-group.

Based on $\mathcal{W}$ for the baseline ($*$) and the counterfactual (c) SSE, we obtain the change in social welfare from the initial to the counterfactual SSE for a given level of inequality aversion as

$$\widehat{\mathcal{W}}_t(\varepsilon) = \frac{\mathcal{R}_t^c}{\mathcal{R}_t^*} \frac{1 - \mathcal{I}_t(\varepsilon)^c}{1 - \mathcal{I}_t(\varepsilon)^*}. \tag{18}$$

With this formulation, we acknowledge the efficiency-equity trade-off that is inherent to many spatial shocks and policies. If there is a positive effect on aggregate welfare accompanied by an increase in inequality, the effect on social welfare qualitatively and quantitatively depends on inequality aversion.

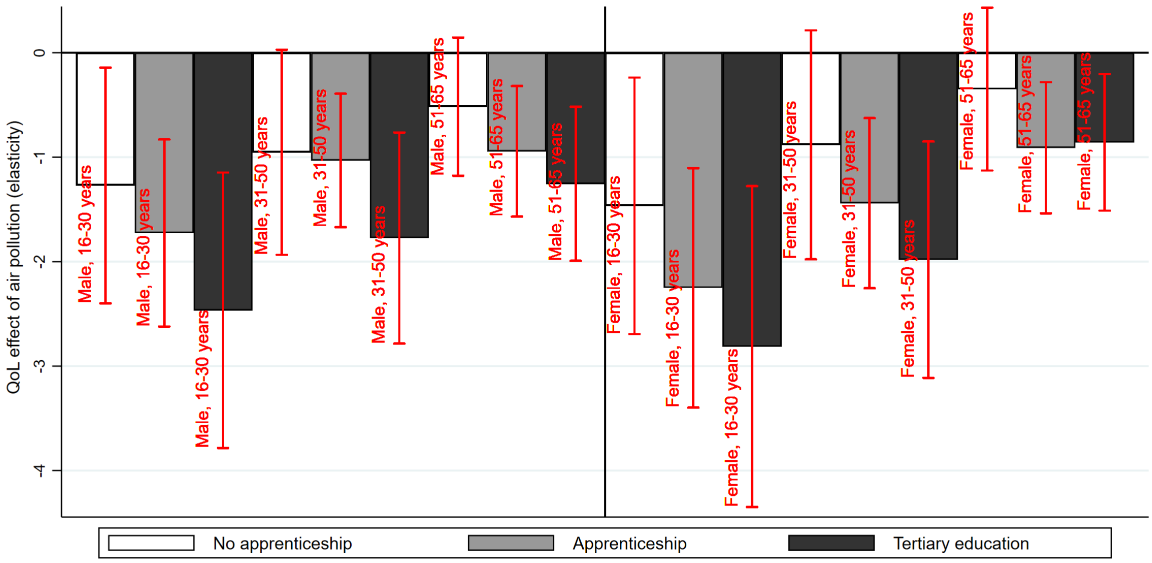
F.2 Application

Estimation. The descriptive results reported in Table 2 point to a negative effect of particulate matter air pollution (PM^{10}) on QoL. To obtain a causal estimate of the effect of air pollution on group-specific QoL, we require an identification strategy that addresses the obvious concern that air pollution may be correlated with unobserved QoL determinants. As an example, a more extensive road network may induce traffic and increase air pollution while having a positive QoL effect due to reduced travel times. The potential for a downward bias in the air pollution effect on QoL is significant as transport accounts for 20% of particulate matter emissions in Germany, on average, with a greater share in urbanised regions (Umweltbundesamt, 2020).

Therefore, we use an instrumental variable approach which exploits that the spatial diffusion of air pollution is shaped by winds (Deryugina et al., 2019; Heblich et al., 2020a). To this end, we compute black coal and brown coal exposure measures that aggregate over black or brown coal deposits in surrounding regions, weighted by wind-adjusted distance. Intuitively, we scale down the crow-flight distance from region j to i if winds typically blow from j to i and scale the distance up if the opposite is true. We normalise these exposure measures by the naive spatial aggregate of coal deposits and exclude any coal deposits in the own region. Hence, when we use the resulting (log) coal exposure measures as instrumental variables for air pollution, identification stems from exogenous variation introduced by wind directions, exclusively. The rationale for using coal fields in the exposure measures is that, historically, energy-intensive industries and coal power plants co-located with coal fields as shipping costs were high until the mid 20^{th} century (Fernihough and O'Rourke, 2021; Mohammed and Williamson, 2004). Unlike for industries and power plants, we can rule out reverse causality from QoL to the locations of coal fields. Since we exclude the own region ($j = i$) in the exposure measures, the instrumental variables exclude localised disamenities, for example due to unpleasant views. For a more detailed discussion of the construction, the relevance and the validity of the instruments as well as the underlying mechanisms, we refer to Appendix M.2.

In Figure 7, we display the estimated pollution effects from group-specific instrumental variable regressions in which we also control for all covariates used in Table 2. In keeping with intuition, the point estimates are negative for all 18 groups. On average, the effect is larger than in Table 2, suggesting a role for unobserved confounders that are positively correlated with QoL and negatively correlated with pollution, such as transport. There is a notable age gap, with the QoL of younger workers being more sensitive to air pollution.

Figure 7: Quality-of-life effect of air pollution by group

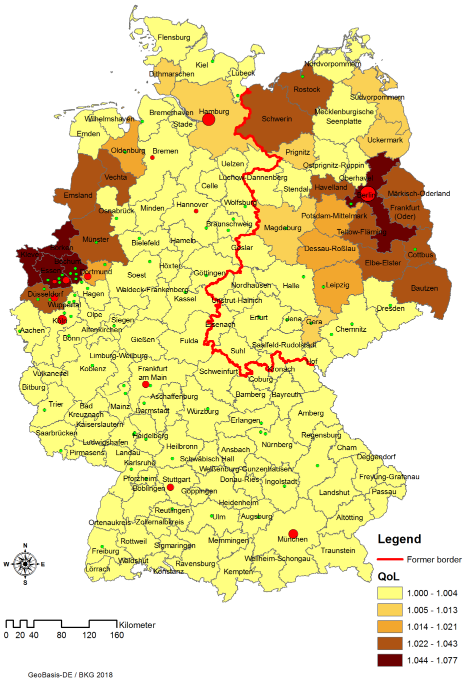


Note: Elasticity estimates are from group-specific regressions of the log of DSM-QoL ($A^{\theta}_{i,t}$) inverted as discussed in D.3 against the log of particular matter (PM^{10}), controlling for the remaining covariates listed in Table 2. We use the log of the wind-adjusted-distance-weighted aggregates of black and brown coal deposits in surrounding regions (excluding the self-potential) as instrumental variables for pollution. These coal exposure measures are normalised by the non-wind-adjusted spatial lags of black and and brown coal deposits, so that identification is driven by wind direction exclusively.

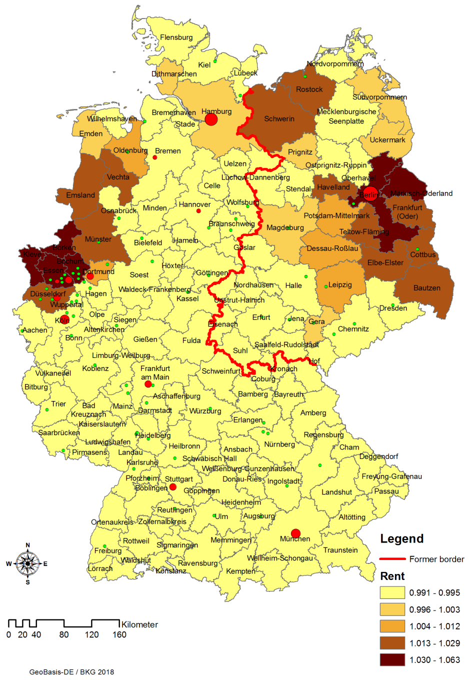
Regional effects. To generate an exogenous change in QoL $\widehat{A^{\theta}_{i,t}}$, we combine the group-specific estimates of the air pollution effect on QoL from Figure 7 with a hypothetical region-specific policy. Specifically, we reduce the regional PM^{10} concentration to the 75$^{\text{th}}$ percentile in the distribution of pollution levels across regions where the levels exceed that threshold. While we choose the threshold arbitrarily with no particular policy in mind, the general design vaguely resembles the US Clean Air Act. Panel (a) in Figure 8 illustrates the simulated policy effect on the weighted (by group employment) average regional QoL. Three broader regions stand out as being treated owing to relatively high air pollution levels: The west, home to black coal fields; the north, home to various seaports; the east, home to brown coal fields. The QoL effects are sizable, with the largest increase in average QoL of 7.7% in Bochum (in the west).

The policy-induced positive change in regional QoL naturally creates incentives for workers to relocate. As workers move to the treated regions, they congest the housing market, leading

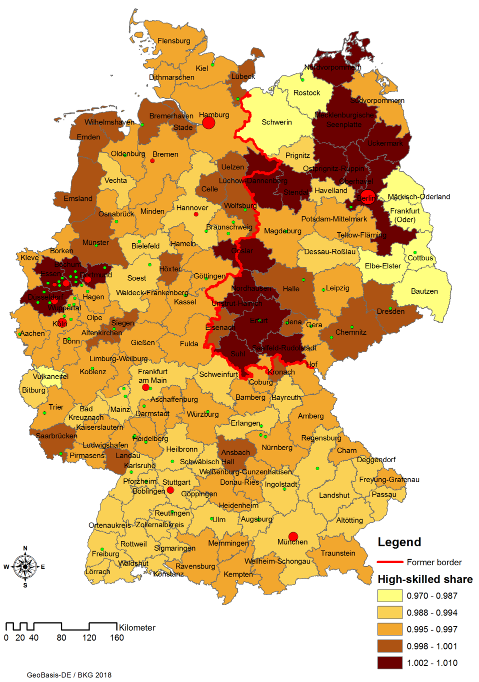
Figure 8: Counterfactual analysis: Regional effects



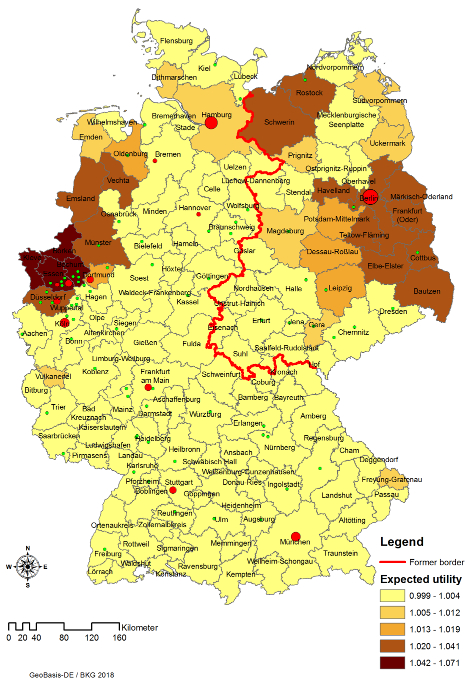
(a) QoL



(b) Rent



(c) High-skilled share



(d) Expected utility (stayers)

Note: We aggregate the model solutions for the initial and the counterfactual SSE from the region-group level to the region level using the respective SSE employment shares as weights. We then display the ratio of the counterfactual regional aggregates over the initial regional aggregates.

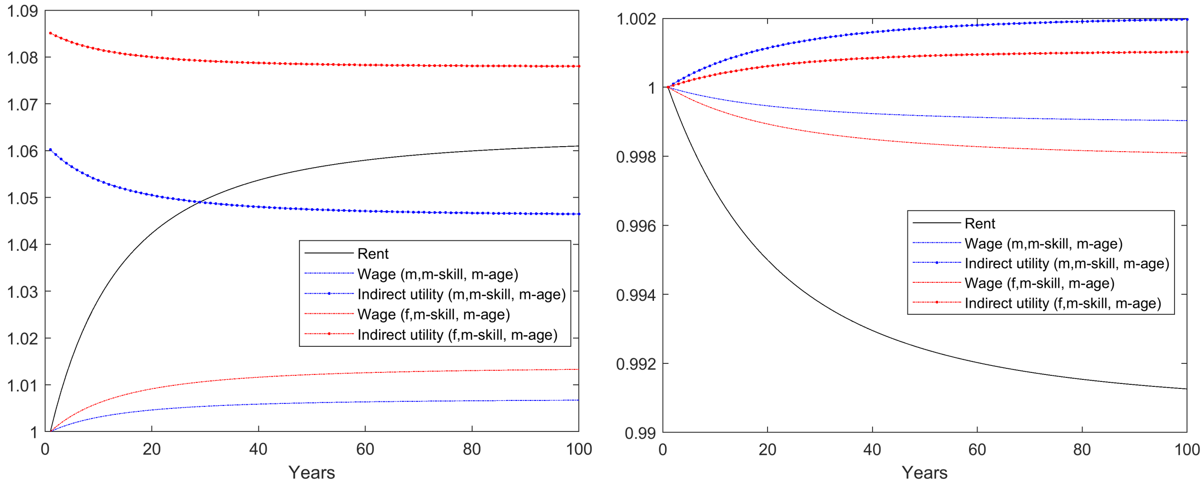
to higher rents as illustrated in panel (b). Unsurprisingly, we find the largest increase in rent of 6.3% in Bochum where QoL increased the most. There are small decreases in rent in the range of -0.9% to -0.7% throughout the non-treated regions as these lose workers to the treated regions. Accounting for relocation effects is a natural strength of quantitative models compared to reduced-form settings, where indirect treatments represent a challenge for the identification and interpretation of treatment effects.

Since we quantify the model for 18 age-gender-skill groups, our model-based counterfactuals deliver rich sorting effects. Panel (c) shows how the policy leads to an increase in the high-skilled share in the urbanised treated regions in particular. This increase is driven by a combination of the high-skilled having a relatively large valuation of air quality (b^{θ}) and a relatively large migration elasticity (γ^{θ}) while facing relatively low migration costs ($\tau^{\theta}_{ij \neq i}$) and net-costs of agglomeration ($\beta \times (1-\alpha) - \kappa^{\theta}$).

A distinctive feature of our DSM is that there is no exogenous reservation utility that anchors the spatial economy. Because of migration costs, spatial differences in expected group-specific utility are not arbitraged away, not even in the SSE. Hence, while migration leads to capitalisation of a change in QoL into rents (see panel b), capitalisation remains imperfect so that we see persistent effects on regional utility in panel (d). Since $1 - \alpha = 33\%$ of the income is spent on housing, the 6.3% increase in rent in Bochum, for instance, implies a $6.3\% \times 0.33\% = 2.1\%$ decline in utility, ceteris paribus, compared to the 7.7% QoL-induced utility gain. In other words, only about one fourth of policy-induced QoL increase capitalises into rents. The remaining fraction boosts utility persistently. Note that the net-effect on expected indirect utility of 7.1% in Bochum does not amount exactly to the difference between the equivalent utility effects of the QoL and rent increases due to agglomeration-induced wage effects and sorting (see panel c).

Temporal effects. While there are sizable utility gains in the positively treated regions in the SSE, utility increases even more during the transition period. In Figure 9, we plot the evolution of wages, rents, and indirect utility during the first 100 years of the SSE-to-SSE transition. We show group-specific wage and indirect utility effects for middle-aged, middle-skilled men and women separately. In Bochum, the selected group of female workers receives an 8.5% increase in QoL which maps one-to-one to an indirect utility effect in the initial period. The respective male group receives a smaller gain due to a lower valuation of clean air. Over the subsequent years, the heightened QoL attracts workers from other regions, increasing rent (due to inelastically supplied land) and wage (due to agglomeration economies) levels. Since the effect of the former dominates that of the latter, indirect utility decreases over time, and so does the incentive for workers from other regions to relocate to Bochum. Since women enjoy greater returns to agglomeration, spatial arbitrage neutralises a smaller fraction of their utility gain, which adds to the long-run benefits they experience relative to men. Munich is not directly treated by our simulated policy. The city is indirectly affected by the policy, however, as it loses workers to the positively treated regions. Rents and wages decrease and since the effect of the former dominates the latter, indirect utility increases. Hence, there is a positive policy spillover effect that operates through migration and a de-congested housing market. Net benefits to women are lower since they take a greater wage cut due to larger returns to agglomeration.

Figure 9: Counterfactual analysis: Temporal effects



(a) Treated: *Bochum* (positive QoL effect)

(b) Not treated: *Munich* (displacement effect)

Note: Model-based numerical simulation of the SSE-to-SSE transition. Pre-policy values in all variables normalized to one. Policy is a region-group-specific increase in QoL due to a hypothetical reduction in air pollution in the most polluted regions. f: female, m: male, m-skill: middle-skilled (apprenticeship), m-age: middle-aged (31-50).

Aggregate effects. We aggregate the SSE-to-SSE region-group effects delivered by the model simulations to relative changes in aggregate outcomes in Table 3. In doing so, we distinguish between treated regions where the policy bites and the remaining non-treated regions which are only indirectly affected through displacement. Although our estimated migration elasticity parameters (γ^θ) are relatively small, we observe a sizable worker flow, increasing employment in the treated regions by almost 10% in total. GDP increases more than proportionately compared to employment in the treated area since agglomeration economies and sorting raise wages. Rents naturally increase in the treated area due to more congested housing markets. Since the non-treated area accounts for about twice as many workers (20M) as the treated area (10M) in the initial SSE, the relative decline in employment in the non-treated area is about half as large (-4.5%). The displacement effect naturally leads to adjustments in wages, rents and group composition in the opposite direction of those in the treated area.

The employment-weighted group-region utility increases by 2.2% across all regions. This increase is driven primarily by the treated area where the group-weighted average utility increases by slightly more than 3.5%. There is a small positive effect within the non-treated area owing to lower real living cost. The spatially differentiated utility effect once more highlights that, unlike in the canonical CSE framework, spatial policies can help targeted regions if there are mobility frictions. The effect on social welfare $\mathcal{W}$ is about 13% lower if we aggregate group-region utility $\mathcal{R}_i^\theta$ using an inequality parameter $\varepsilon = 0.5$, which is towards the lower end of the

Table 3: Counterfactual analysis: Aggregate effects

Outcome	All regions	Treated area	Non-treated area
Population	1.0000	1.0949	0.9536
GDP	0.9991	1.0996	0.9515
Average wage	0.9991	1.0043	0.9978
Average rent	1.0021	1.0175	0.9911
High-skilled share	1.0000	1.0109	0.9946
Skilled share	1.0000	1.0118	0.9976
Average utility	1.0219	1.0350	1.0003
Social welfare (inequality adjusted)	1.0191	.	.
Monetised average utility (bn. €)	23.1	.	.
Monetised social welfare (bn. €)	20.2	.	.

Notes: Results from model-based numerical simulations. Treated regions are those where a hypothetical policy improves QoL via lower air pollution. Non-treated regions are affected indirectly through displacement. All outcomes except for the last two are given in ratios of counterfactual (SSE) values over initial (SSE) initial values. Social welfare deflates average utility in group-region inequality using the Atkinson (1970) measure ($\epsilon = 0.5$). Monetised average utility and social welfare are yearly flow measures obtained by multiplying the utility and welfare ratios by initial GDP.

range considered by Atkinson (1970). If we use $\varepsilon = 2$ (towards the higher end of the considered range), the discount increases to 35%. Since we obtain virtually the same inequality-adjusted social welfare effect if we discount on inter-regional inequality, exclusively, we can conclude that the cost of the policy comes in the form of increased spatial inequality.

A simple way to monetise the welfare effect is to multiply the relative change in welfare by the total wage bill in the initial SSE. If we abstract from inequality aversion, a proportionate increase in yearly region-group wages that totals €23.1 bn would achieve the same utility effect as the policy. If we adjust for the policy effect on inequality using $\varepsilon = 0.5$, a fully equitable increase in the total wage bill of €20.2 bn would suffice. With $\varepsilon = 2$, the monetised welfare effect drops to €15.0 bn. In this application of the model, we abstract from the cost of the measures used to achieve the pollution reduction. Yet, it is clear from the example that once we move beyond the canonical CSE framework, the result of a cost-benefit test of a spatial policy will critically depend on the social welfare function.

F.3 Other applications

The Covid-19 pandemic has spurred a debate about the future of big cities (Nathan and Overman, 2020). A typical argument brought forth is that the pandemic erodes the main comparative advantage of big cities: economic and social benefits of proximity. We apply the procedure developed in this section to quantitatively evaluate three apocalyptic scenarios: a) a reduction in productivity due to an elimination of all agglomeration benefits arising from density; b) a reduction in QoL due to a loss of amenities that relate to social interaction (captured by our big data amenity); c) the combination of a) and b). The headline findings for the scenarios a)/b)/c) are as follows: Large labour markets (>250k employed workers) lose

8.2%/36.7%/37.9% percent of their workforce to small labour markets; aggregate GDP decreases by 10.5%/1%/10.9%; rents fall by 3.1%/9.3%/11.1% in large labour markets whereas they increase (decrease) by (0.4)%/6.7%/5.1% in small labour markets; despite a larger reduction in the urban wage premium, the high-skilled are more likely to remain in large cities due to their amenity preference. While these simulated effects on big cities are large, they are not nearly as devastating as predicted by a frictionless CSE model. We refer the interested reader to Appendix M.3 for details.

G Conclusion

We develop the first DSM with heterogeneous forward-looking agents in which all unobserved structural fundamentals can be quantified without assuming that the economy is in a stationary spatial equilibrium. We exploit this novel feature of our model to make several contributions that are of academic and policy interest alike.

A key insight from our analysis is that differentials in QoL across regions are much larger once we quantitatively account for idiosyncratic taste heterogeneity. While the existence of an urban wage premium that reflects productivity advantages of cities is by now uncontroversial, the evidence for an urban QoL premium is weak at best. Our results show that accounting for idiosyncratic tastes that reduce mobility, the consumption value of cities is key to rationalising why more than 50% of the world's population lives in cities. CSE models have been the workhorse tool for the evaluation of non-marketed goods such as clean air, education, safety, or transport, just to name a few. Our results show that consensus estimates of the value of such local public goods are likely lower bounds, implying a stronger case for policies that seek to improve QoL.

The existence of localised place-based policies such as Enterprise Zones or broader regional redistribution schemes such as the EU Cohesion Fund suggests that spatial equity matters to policy makers and voters. There is an obvious tension between such spatial policies and the workhorse spatial equilibrium models which rule out spatial effects of spatial policies by assumption. We provide a quantitative framework for the evaluation of the aggregate and distributional welfare consequences of place-based policies that allows for spatial incidence and relocation effects. This framework closes the gap between QSMs that assume perfect spatial arbitrage and the reality of spatial policy-making where trading efficiency against equity is the order of the day. We show that even a moderate spatial inequality aversion can have a sizable impact on the social welfare effect of a spatially targeted policy. As the literature on spatial policy evaluation moves beyond the canonical framework in the tradition of Rosen-Roback, the spatial aggregation of welfare effects will require an explicitly defined social welfare function.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Ahlfeldt, Gabriel M., "Urbanity," *CESIfo working paper*, 2013, *4533.*

— **and Elisabetta Pietrostefani**, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

— **, Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf**, "The Economics of Density : Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," *NBER Working Paper*, 2011, *14472.*

— **and Bert Lue**, "Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life," *Journal of Urban Economics*, 2015, *89*, 74–92.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

Almagro, Milena and Tomas Domínguez-Iino, "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Working paper*, 2020.

Althoff, Lukas, Fabian Eckert, Sharat Ganapati, and Conor Walsh, "The City Paradox: Skilled Services and Remote Work," *Working paper*, 2020.

Arnott, Richard J. and Joseph E. Stiglitz, "Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size," *The Quarterly Journal of Economics*, 1979, *93* (4), 471–500.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 2010, *100* (3), 1008–1045.

Atkinson, Anthony B., "On the measurement of inequality," *Journal of Economic Theory*, 1970, *2* (3), 244–263.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 2018, *32* (3), 259–280.

Balboni, Clare, "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities," 2019, *Working paper.*

Baum-Snow, Nathaniel and Ronni Pavan, "Inequality and City Size," *The Review of Economics and Statistics*, 2013, *95* (5), 1535–1548.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins, "Migration and hedonic valuation: The case of air quality," *Journal of Environmental Economics and Management*, 2009, *58* (1), 1–14.

Blomquist, Glenn C., Mark C. Berger, and John P. Hoehn, "New Estimates of Quality of Life in Urban Areas," *The American Economic Review*, 1988, *78* (1), 89–107.

Blouri, Yashar and Maximilian V. Ehrlich, "On the optimal design of place-based policies: A structural evaluation of EU regional transfers," *Journal of International Economics*, 2020, *125*, 103319.

Boelmann, Barbara and Sandra Schaffner, "Real-Estate Data for Germany (RWI-GEO-RED v1) - Advertisements on the Internet Platform ImmobilienScout24 2007-03/2019," Technical Report, RWI Leibniz-Institut für Wirtschaftsforschung 2019.

Brakman, Steven, Harry Garretsen, and Marc Schramm, "The strategic bombing of German cities during World War II and its impact on city growth," *Journal of Economic Geography*, 2004, *4* (2), 201–218.

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

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement," *NBER Working paper*, 2019, *23695.*

_ , Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Carlino, Gerald A. and Albert Saiz, "Beautiful city: Leisure amenities and urban growth," *Journal of Regional Science*, 2019, *59* (3), 369–408.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein, "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *The Quarterly Journal of Economics*, 2010, *125* (1), 215–261.

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Ciccone, Antonio and Robert E. Hall, "Productivity and the Density of Economic Activity," *The American Economic Review*, 1996, *86* (1), 54–70.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, chapter 5, pp. 247–348.

_ , Gilles Duranton, and Laurent Gobillon, "Spatial wage disparities: Sorting matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

_ , _ , and _ , "The Costs of Agglomeration: House and Land Prices in French Cities," *The Review of Economic Studies*, 2019, *86* (4), 1556–1589.

Conte, Bruno, Klaus Desmet, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "Local Sectoral Specialization in a Warming World," *NBER Working Paper*, 2020, *28163.*

Criscuolo, Chiara, Ralf Martin, Henry G. Overman, and John Van Reenen, "Some Causal Effects of an Industrial Policy," *American Economic Review*, 2019, *109* (1), 48–85.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Unbalanced Trade," *American Economic Review*, 2007, *97* (2), 351–355.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif, "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, 2019, *109* (12), 4178–4219.

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *Journal of Political Economy*, 2018, *126* (3), 903–983.

Diamond, Rebecca, "The determinants and welfare implications of US Workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

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

Epple, Dennis, Brett Gordon, and Holger Sieg, "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, 2010, *100* (3), 905–924.

European Environment Agency, "Healthy environment, healthy lives: how the environment influences health and well-being in Europe," Technical Report, Luxembourg 2020.

Fajgelbaum, Pablo D. and Cecile Gaubert, "Optimal Spatial Policies, Geography, and Sorting," *The Quarterly Journal of Economics*, 2020, *135* (2), 959–1036.

Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Südekum, "Dialects, cultural identity, and economic exchange," *Journal of Urban Economics*, 2012, *72* (2-3), 225–239.

Fernihough, Alan and Kevin H. O'Rourke, "Coal and the European Industrial Revolution," *The Economic Journal*, 11 2021, *131* (635), 1135–1149.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 2002, *40* (2), 351–401.

Gaigné, Carl, Hans R.A. Koster, Fabien Moizeau, and Jacques-François Thisse, "Who Lives Where in the City? Amenities, Commuting and Income Sorting," *CEPR Discussion Paper*, 2017, *11958.*

Glaeser, Edward L., *Cities, Agglomeration, and Spatial Equilibrium*, Oxford: Oxford University Press, 2008.

_ , David Laibson, and Bruce Sacerdote, "An Economic Approach to Social Capital," *The Economic Journal*, 2002, *112* (483), F437–F458.

_ , Jed Kolko, and Albert Saiz, "Consumer city," *Journal of Economic Geography*, 2001, *1* (1), 27–50.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse,Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, pp. 131–195.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg, "East Side Story: Historical Pollution and Persistent Neighborhood Sorting," *Journal of Political Economy*, 2020, *forthcoming.*

_ , Stephen J Redding, and Daniel M Sturm, "The Making of the Modern Metropolis: Evidence from London," *The Quarterly Journal of Economics*, 11 2020, *135* (4), 2059–2133.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028.

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kline, Patrick and Enrico Moretti, "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *The Quarterly Journal of Economics*, 2013, *129* (1), 275–331.

_ and _ , "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," *Annual Review of Economics*, 2014, *6* (1), 629–662.

Koşar, Gizem, Tyler Ransom, and Wilbert van der Klaauw, "Understanding Migration Aversion Using Elicited Counterfactual Choice Probabilities," *Journal of Econometrics*, 2021, *forthcoming.*

Kosfeld, Reinhold and Alexander Werner, "Deutsche Arbeitsmarktregionen-Neuabgrenzung nach den Kreisgebietsreformen 2007-2011," *Raumforschung und Raumordnung*, 2012, *70* (1), 49–64.

Koster, Hans R. A. and Edward W. Pinchbeck, "How do Households Value the Future? Evidence from Property Taxes," *American Economic Journal: Economic Policy*, 2021, *forthcoming.*

Levinson, Arik, "Valuing public goods using happiness data: The case of air quality," *Journal of Public Economics*, 2012, *96* (9-10), 869–880.

Lin, Jeffrey, Sanghoon Lee, and Seung Hoon Lee, "The Well-Being of Nations: Estimating Welfare from International Migration," *Working paper*, 2019.

Linden, Leigh and Jonah E. Rockoff, "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 2008, *98* (3), 1103–1127.

Luechinger, Simon, "Valuing air quality using the life satisfaction approach," *The Economic Journal*, 2009, *119* (536), 482–515.

McFadden, Daniel, "The measurement of urban travel demand," *Journal of Public Economics*, 1974, *3* (4), 303–328.

— **and Kenneth Train**, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 2000, *15* (5), 447–470.

Mohammed, Saif I.Shah and Jeffrey G. Williamson, "Freight rates and productivity gains in British tramp shipping 1869 - 1950," *Explorations in Economic History*, 2004, *41* (2), 172–203.

Monras, Joan, "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis," *Journal of Political Economy*, 2020, *128* (8), 3017–3089.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Moore, Michael J. and W. Kip Viscusi, "The quantity-adjusted value of life," *Economic Inquiry*, 1988, *26* (3), 369–388.

Moretti, Enrico, "Local Labor Markets," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, 2011, chapter 14, pp. 1237–1313.

Nathan, Max and Henry G. Overman, "Will coronavirus cause a big city exodus?," *Environment and Planning B*, 2020, *47* (9), 1537–1542.

Neumark, David and Helen Simpson, "Place-Based Policies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, chapter 18, pp. 1197–1287.

OECD, "Taxing Wages 2017," 2017.

Redding, Stephen J. and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9* (1), 21–58.

Ricardo, David, *On the Principles of Political Economy and Taxation*, London: John Murrey, 1817.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Sherwin, "Wages-based Indexes of Urban Quality of Life," in Peter Mieszkowski and Mahlon Straszheim, eds., *Current Issues in Urban Economics*, Baltimore, MA: Johns Hopkins University Press, 1979, pp. 74–104.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

— **, Arianna Salazar, and James Bernard**, "Crowdsourcing architectural beauty: Online photo frequency predicts building aesthetic ratings," *PLoS ONE*, 2018, *13* (7).

Statistisches Bundesamt, "Einkommens- und Verbrauchsstichprobe Konsumausgaben privater Haushalte," *Fachserie*, 2020, *15* (5).

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872.

Umweltbundesamt, "Nationale Trendtabellen für die deutsche Berichterstattung atmosphärischer Emissionen seit 1990, Emissionsentwicklung 1990 bis 2018," Technical Report, Umweltbundesamt 2020.

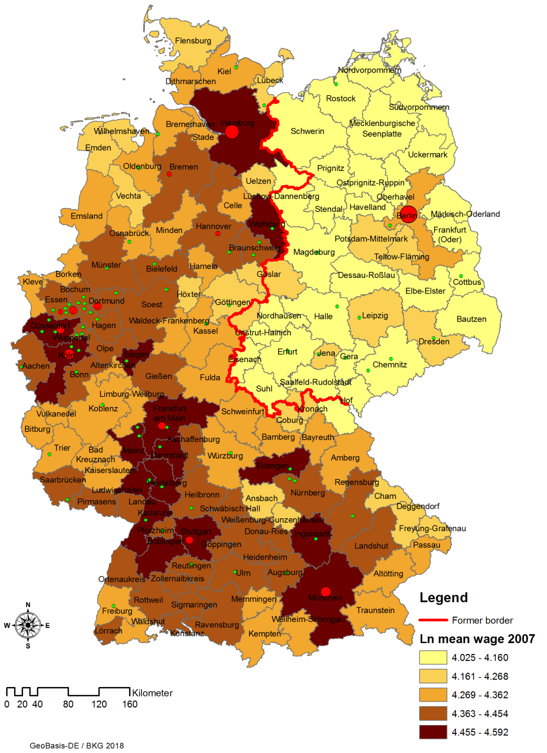
APPENDIX FOR ONLINE PUBLICATION

This section presents an online appendix containing complementary material not intended for publication. It does not replace the reading of the main paper.

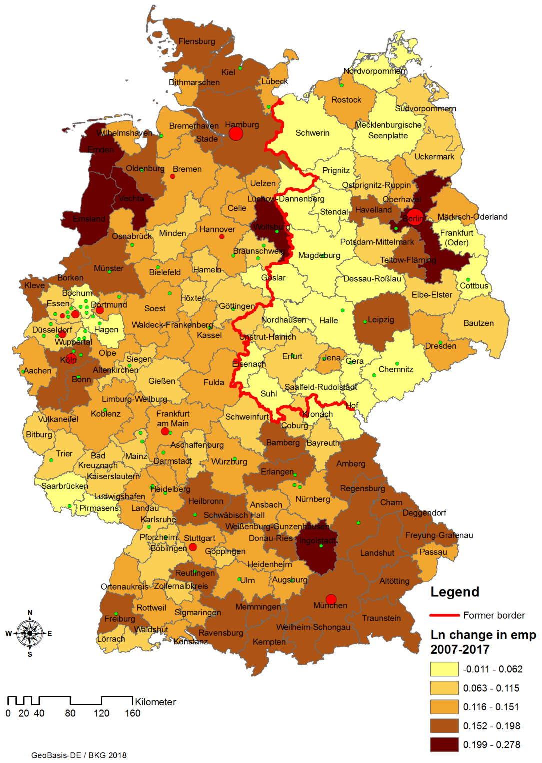
H Stylised facts

Complementing Figure 1 in the main paper, Figure A1 visualises causes and consequences of migration in three illustrative maps. Panel (a) plots the spatial distribution of nominal wages. In keeping with intuition, wages tend to be higher in agglomerated areas such as Rhine-Ruhr, Rhine-Main or the metropolitan areas of Hamburg, Munich or Stuttgart. Panel (b) shows the spatial distribution of net-migration over the 2007 to 2017 period. High-wage areas tend to experience positive net-migration, suggesting that workers respond to economic incentives when making migration decisions. Panel (c) shows a strong correlation between net-migration and changes in rents, in line with housing markets representing a congestion force.

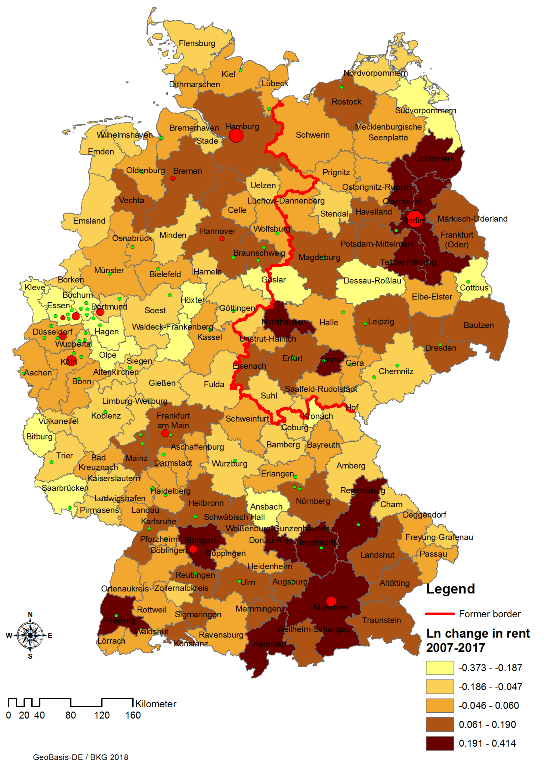
Figure A1: Wages, migration and rents



(a) Wages 2007



(b) Change in employment 2007-2017



(c) Change in rent 2007-2017

Note: Data from the IAB and Immobilienscout24 accessed via FDZ Ruhr.

I Literature appendix

Table A1 summarizes the recent literature on DSM and QSM that explicitly model migration. The distinctive feature of our model is the invertability of all structural fundamentals under perfect foresight from a TSE.

Table A1: Dynamic and quantitative spatial models

Authors	Model type	Expectations	Inversion		Counterfactual
Ahlfeldt et al. (2020) WP	DSM, GE, MC	PF	P,H,A,bMC	TSE	SSE to SSE
Balboni (2019), R&R AER	DSM, GE, MC	PF	P,MA	TSE	TSE to ED
Bryan and Morten (2019), JPE	QSM, MC	Static	-	-	-
Caliendo et al. (2019b), Ecta	DSM, GE, MC	PF	-	-	TSE to ED
Caliendo et al. (2019a), R&R JPE	DSM, GE, MC	PF	-	-	TSE to ED
Conte et al. (2020), WP	DSM, GE, MC	Static	P,A,uMC	TSE	TSE to SSE
Desmet et al. (2018), JPE	DSM, GE, MC	Static	P,A,uMC	TSE	TSE to SSE
Fan (2019), AEJ: Macro	QSM, GE, MC	Static	bMC, TC	SSE	SSE to SSE
Monras (2020), JPE	DSM, GE	PF	P,H,A,MR	SSE	TSE to SSE
Tombe and Zhu (2019), AER	QSM, MC	Static	-	-	-

Abbreviations:
Model type: QSM = Quantitative spatial model; DSM = Dynamic spatial model; GE = General equilibrium; MC = Migration cost
Expectations: PF = Perfect foresight
Inversion: P = Exogenous productivity; H = Exogenous housing supply; A = Exogenous amenity; uMC = Unilateral migration costs; bMC = Bilateral migration costs; MR: Migration rate; MA = Market access; TC: Trade costs;
SSE = Stationary spatial equilibrium; TSE = Transitory spatial equilibrium;
Counterfactual: ED = Given end date

J Theory appendix

This section complements Section C in the main paper which develops our model.

J.1 Housing market

In this appendix, we derive the housing market equilibrium condition Eq. (8). Developers produce housing according to the Cobb-Douglas housing production function in Eq. (7) and seek to maximise profits:

$$\pi_{i,t}^{h} = p_{i,t}\eta_{i,t}\left(\frac{\bar{T}_i}{\beta}\right)^{\beta}\left(\frac{K_{i,t}}{1-\beta}\right)^{1-\beta} - r_{i,t}^{T}\bar{T}_i - K_{i,t}, \tag{19}$$

where we have normalised the internationally competitive interest rate for capital to unity and $r_{i,t}^{T}$ is the local rental rate for developable land. From the first-order conditions, we obtain:

$$r_{i,t}^{T} = \frac{\beta}{1-\beta}\frac{K_{i,t}}{\bar{T}_i}. \tag{20}$$

Using Eq. (20) in Eq. (19) and assuming zero-profit delivers

$$p_{i,t} = \frac{(r_{i,t}^{T})^{\beta}}{\eta_{i,t}}. \tag{21}$$

Using Eqs. (21) and (20) in Eq. (19), we can express housing supply as

$$H_{i,t}^{S} = \eta_{i,t}^{\frac{1}{\beta}} \left(\frac{\bar{T}_i}{\beta} \right) p_{i,t}^{\frac{1-\beta}{\beta}}, \tag{22}$$

where $\frac{1-\beta}{\beta}$ is the housing supply elasticity. From Eq. (2), housing demand in region i is given by

$$H_{i,t}^{D} = (1-\alpha)(1-\iota) \sum_{\theta} L_{i,t}^{\theta} \varphi_{i,t}^{\theta} = (1-\alpha)(1-\iota) X_{i,t}. \tag{23}$$

Housing market clearing implies that $H_{i,t}^{D} = H_{i,t}^{S}$, which leads to Eq. (8). Alternatively, we can express the regional housing rent as a function of structural parameters, structural fundamentals, and employment density:

$$p_{i,t} = \left(\frac{(1-\alpha)(1-\iota)\beta}{\eta_{i,t}^{\frac{1}{\beta}}} \right)^{\beta} \left[\sum_{\theta} \frac{L_{i,t}^{\theta}}{\bar{T}_i} \left(\frac{L_{i,t}}{\bar{T}_i} \right)^{\kappa^{\theta}} \psi_{i,t}^{\theta} \right]^{\beta}. \tag{24}$$

The first term in the sum captures the direct effect of employment density on the supply of housing: inelastically supplied land generates a congestion force in the form of higher rents when immigration into i raises employment. The second term in the sum captures the indirect effect of employment density: Density increases productivity and in turn wages via agglomeration economies, which further increases housing demand. While our model can provide the microfoundations for a regression of the log of housing rents against the log of employment density as in Combes et al. (2019), the estimated elasticity of that regression would not directly correspond to β in our model. Using employment (or population) density instead of output density as a regressor (see Eq. (31) below), we would underestimate the land share, the housing supply elasticity, and the congestion force generated by the housing market.

J.2 Net present value of utility

This section complements Section C.3 in which we introduce the migration net present value.

Strictly monotonic transformations of utility functions still represent the same underlying preferences. We follow the conventions in the DSM literature (see e.g. Caliendo et al., 2019b) and employ a logarithmic formulation of the net present value of utility, which allows to derive simple closed-form solutions for expected utility in Appendix J.3.

This net present value of a worker of type θ currently living in region i, and who lived in region k at time period $t-1$ depends on current period utility and the maximal discounted future utility, which in turn is a function of bilateral utility between all other (potentially) different regions $j, m, ..., n, h$ in all future periods $t+1, t+2, ..., (t+T)+(t+1+T)$. Workers expect to stay at their final destination $h \in J$ from time period $(t+1)+T$ onward, such that the net present value of utility is given as

$$
\begin{aligned}
\ln NPV^{\theta}_{i|k,t}(\omega) = &\ln\left[\frac{(1-\iota)w^{\theta}_{i,t}}{p^{1-\alpha}_{i,t}}A^{\theta}_{i,t}\exp\left(a^{\theta}_{ki,t}(\omega)-\tau^{\theta}_{ki}\right)\right] \\
&+ \max_{\{j,m,\dots,n,h\}^{J}_{j,m,\dots,n,h=1}}\Bigg\{\frac{1}{1+\rho}\ln\left(\frac{(1-\iota)w^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,t+1}}A^{\theta}_{j,t+1}\exp\left(a^{\theta}_{ij,t+1}(\omega)-\tau^{\theta}_{ij}\right)\right) \\
&+ \left(\frac{1}{1+\rho}\right)^{2}E\left[\ln\left(\frac{(1-\iota)w^{\theta}_{m,t+2}}{p^{1-\alpha}_{m,t+2}}A^{\theta}_{m,t+2}\exp\left(a^{\theta}_{jm,t+2}(\omega)-\tau^{\theta}_{jm}\right)\right)\right] + \ \dots \\
&+ \left(\frac{1}{1+\rho}\right)^{T+1}\Bigg[E\left[\ln\left(\frac{(1-\iota)w^{\theta}_{h,(t+1)+T}}{p^{1-\alpha}_{h,(t+1)+T}}A^{\theta}_{h,(t+1)+T}\exp\left(a^{\theta}_{nh,(t+1)+T}(\omega)-\tau^{\theta}_{nh}\right)\right)\right] \\
&+ \sum_{s=(t+2)+T}^{\infty}\left(\frac{1}{1+\rho}\right)^{s-(t+1+T)}E\left[\ln\left(\frac{(1-\iota)w^{\theta}_{h,s}}{p^{1-\alpha}_{h,s}}A^{\theta}_{h,s}\exp\left(a^{\theta}_{hh,s}(\omega)\right)\right)\right]\Bigg]\Bigg\},
\end{aligned}
$$

where $\frac{1}{1+\rho} \in (0,1)$ is the time discount factor. Combining with the definitions of utility (1) and demand functions (2) these results lead to the net present value of utility in Eq. (9).

J.3 Expected utilities and migration probabilities

This section complements Section C.3 in which we introduce the migration gravity Eq. (10).

J.3.1 Expected utility

We are interested in the expected net present value of workers of type θ when migrating from region i to j at the end of time period t. Taking the expectation over idiosyncratic Gumbel-distributed amenity shocks involves solving both the unconditional expectation over current shock realisations as well as the expectation of future shocks, conditional on some region $n \in J$ offering the highest expected utility in future time periods to these workers.

J.3.2 Unconditional expectation of current period utility

Random amenity shocks are distributed according to a Gumbel distribution with the following cumulative distribution and density function:

$$
F^{\theta}_{ij,t}(a) = \exp\left(-\tilde{B}^{\theta}_{ij,t}\exp\{-\left[\gamma^{\theta}a+\Gamma\right]\}\right)
$$

$$
f^{\theta}_{ij,t}(a) = \gamma^{\theta}\tilde{B}^{\theta}_{ij,t}\exp\left(-\gamma^{\theta}a-\Gamma-\tilde{B}^{\theta}_{ij,t}\exp\{-\left[\gamma^{\theta}a+\Gamma\right]\}\right)
$$

We first solve for the unconditional expectation over current the component of log-transformed net present value of utility :

$$
\begin{aligned}
E\left[v^{\theta}_{i|k,t}(\omega)\right] &\equiv E\left[\ln\left(\frac{(1-\iota)w^{\theta}_{i,t}A^{\theta}_{i,t}\exp\left(a^{\theta}_{ki,t}(\omega)-\tau^{\theta}_{ki}\right)}{p^{1-\alpha}_{i,t}}\right)\right] \\
&= E\left[\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)+a^{\theta}_{ki,t}(\omega)-\tau^{\theta}_{ki}\right] \\
&= \int_{-\infty}^{\infty}\left[\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)+a^{\theta}_{ki,t}(\omega)-\tau^{\theta}_{ki}\right] \\
&* f\left(a^{\theta}_{ki,t}(\omega)\right)da^{\theta}_{ki,t}(\omega) \\
&= \int_{-\infty}^{\infty}\left[\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)+a^{\theta}_{ki,t}(\omega)-\tau^{\theta}_{ki}\right] \\
&* \gamma^{\theta}\tilde{B}^{\theta}_{ki,t}\exp\left(-\gamma^{\theta}a^{\theta}_{ki,t}(\omega)-\Gamma-\tilde{B}^{\theta}_{ki,t}\exp\{-\left[\gamma^{\theta}a^{\theta}_{ki,t}(\omega)+\Gamma\right]\}\right)da^{\theta}_{ki,t}(\omega),
\end{aligned}
$$

where we substituted the density function for bilateral amenity shocks from above. We then re-define the following variables:

$$
\begin{aligned}
x_t &\equiv \gamma^{\theta}a^{\theta}_{ki,t}(\omega)+\Gamma \\
\lambda_t &\equiv \ln\tilde{B}^{\theta}_{ki,t} \\
y_t &= x_t-\lambda_t
\end{aligned}
$$

Substituting into the integral above yields:

$$
\begin{aligned}
E\left[v^{\theta}_{i|k,t}(\omega)\right] = \int_{-\infty}^{\infty}\gamma^{\theta}\tilde{B}^{\theta}_{ki,t}&\left[\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)-\tau^{\theta}_{ki}+\frac{1}{\gamma^{\theta}}\left(x_t-\Gamma\right)\right] \\
&* \exp\left(-x_t\right)\exp\left(-\exp\left(\lambda_t\right)\exp\left(-x_t\right)\right)\frac{1}{\gamma^{\theta}}dx_t
\end{aligned}
$$

$$
\begin{aligned}
E\left[v^{\theta}_{i|k,t}(\omega)\right] = \int_{-\infty}^{\infty}\tilde{B}^{\theta}_{ki,t}&\left[\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)-\tau^{\theta}_{ki}+\frac{1}{\gamma^{\theta}}\left(x_t-\Gamma\right)\right] \\
&* \exp\left(-x_t-\exp\left(-\left[x_t-\lambda_t\right]\right)\right)dx_t
\end{aligned}
$$

Then note that the derivative of $\exp\left(-\exp\left(-y_t\right)\right)$ is $\exp\left(-y_t-\exp\left(-y_t\right)\right)$ and $\int y_t\exp\left(-y_t-\exp\left(-y_t\right)\right)=\Gamma$. This allows to evaluate the integral at its boundaries:

$$
\begin{aligned}
E\left[v^{\theta}_{i|k,t}(\omega)\right] &= \int_{-\infty}^{\infty}\left[\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)-\tau^{\theta}_{ki}+\frac{1}{\gamma^{\theta}}\left(y_t+\lambda_t-\Gamma\right)\right] \\
&* \exp\left(-y_t-\exp\left(-y_t\right)\right)dy_t \\
&= \left(\ln\left((1-\iota)w^{\theta}_{i,t}\right)+\ln A^{\theta}_{i,t}-\ln\left(p^{1-\alpha}_{i,t}\right)-\tau^{\theta}_{ki}+\frac{1}{\gamma^{\theta}}\left(\lambda_t-\Gamma\right)\right) \\
&* \int_{-\infty}^{\infty}\exp\left(-y_t-\exp\left(-y_t\right)\right)dy_t+\frac{1}{\gamma^{\theta}}\int_{-\infty}^{\infty}y_t\exp\left(-y_t-\exp\left(-y_t\right)\right)dy_t
\end{aligned}
$$

Furthermore note that $[\exp(-\exp(-y_t))]_{-\infty}^{\infty} = 1$. This yields

$$E\left[v^{\theta}_{i|k,t}(\omega)\right] = \left(\ln\left((1-\iota)w^{\theta}_{i,t}\right) + \ln A^{\theta}_{i,t} - \ln\left(p^{1-\alpha}_{i,t}\right) - \tau^{\theta}_{ki} + \frac{1}{\gamma^{\theta}}\lambda_t\right)$$
$$= \ln\left(\frac{\exp\left(-\tau^{\theta}_{ki}\right)(1-\iota)\,w^{\theta}_{i,t}A^{\theta}_{i,t}B^{\theta}_{ki,t}}{p^{1-\alpha}_{i,t}}\right).$$

In line with the definition of per-period utility in Eq. (1) we subsequently define the average per-period welfare for workers of type θ as

$$V^{\theta}_{i|k,t} = \exp E\left[v^{\theta}_{i|k,t}(\omega)\right] = \frac{\exp\left(-\tau^{\theta}_{ki}\right)(1-\iota)\,w^{\theta}_{i,t}A^{\theta}_{i,t}B^{\theta}_{ki,t}}{p^{1-\alpha}_{i,t}}.$$

J.3.3 Migration option values

Assume that all workers expect to stay infinite time periods at their final destination $h \in J$ after $(S+1)$ moves between regions, and incorporate expectations about future amenity shocks up to time period $(t+1)+T$ for all bilateral region pairs already when deciding on destinations at the end of time period $t+1$.

We solve this dynamic problem by backwards induction, that is we first solve for the conditional expectation over idiosyncratic bilateral amenity shocks, given that region $h \in J$ offers the highest life-time utility compared to all other regions $l \neq h$ from time period time period $(t+1)+T$ onward. Let
$\ln \mathcal{V}^{\theta}_{h,(t+1)+T} \equiv \left[\frac{\ln A^{\theta}_{h,(t+1)+T}}{\rho} + \frac{E(a^{\theta}_{hh,(t+2)+T}(\omega))}{\rho(1+\rho)} + \sum_{s=(t+1)+T}^{\infty}\left(\frac{1}{1+\rho}\right)^{s-((t+1)+T)}\ln\left(\frac{(1-\iota)w^{\theta}_{h,s}}{p^{1-\alpha}_{h,s}}\right)\right]$
be the discounted infinite sum over future utilities at time period $(t+1)+T$, then it holds that

$$E\left[v^{\theta}_{n,(t+1)+T}(\omega)\right](1+\rho)^{T} \equiv E\left[\max_{h\in J}\left(\frac{1}{1+\rho}\right)\left[a^{\theta}_{nh,(t+1)+T}(\omega) - \tau^{\theta}_{nh} + \ln \mathcal{V}^{\theta}_{h,(t+1)+T}\right]\right]$$
$$= \sum_{h\in J}\int_{-\infty}^{\infty}\left(\left(\frac{1}{1+\rho}\right)\left[a^{\theta}_{nh,(t+1)+T}(\omega) - \tau^{\theta}_{nh} + \ln \mathcal{V}^{\theta}_{h,(t+1)+T}\right]\right)$$
$$* f\left(a^{\theta}_{nh,(t+1)+T}(\omega)\right)$$
$$* \prod_{h\neq l} F\left[\tau^{\theta}_{nl} - \tau^{\theta}_{nh} + \ln\frac{\mathcal{V}^{\theta}_{h,(t+1)+T}}{\mathcal{V}^{\theta}_{l,(t+1)+T}} + a^{\theta}_{nh,(t+1)+T}(\omega)\right] da^{\theta}_{nh,(t+1)+T}(\omega)$$
$$= \sum_{h\in J}\int_{-\infty}^{\infty}\left(\left(\frac{1}{1+\rho}\right)\left[a^{\theta}_{nh,(t+1)+T}(\omega) - \tau^{\theta}_{nh} + \ln \mathcal{V}^{\theta}_{h,(t+1)+T}\right]\right)$$
$$* f\left(a^{\theta}_{nh,(t+1)+T}(\omega)\right)\prod_{l\neq h} F\left[\Omega^{\theta}_{nhl} + a^{\theta}_{nh,(t+1)+T}(\omega)\right] da^{\theta}_{nh,(t+1)+T}(\omega),$$

where we define the compound parameter $\Omega^{\theta}_{nhl} \equiv \tau^{\theta}_{nl} - \tau^{\theta}_{nh} + \ln\frac{\mathcal{V}^{\theta}_{h,(t+1)+T}}{\mathcal{V}^{\theta}_{l,(t+1)+T}}$. In a next step, we substitute the cumulative distribution and density function for idiosyncratic amenity shocks from above:

$$
\begin{aligned}
E\left[v^{\theta}_{n,(t+1)+T}(\omega)\right](1+\rho)^{T} = & \sum_{h\in J}\int_{-\infty}^{\infty}\left(\frac{1}{1+\rho}\left[a^{\theta}_{nh,(t+1)+T}(\omega)-\tau^{\theta}_{nh}+\ln\mathcal{V}^{\theta}_{h,(t+1)+T}\right]\right)\gamma^{\theta}\tilde{B}^{\theta}_{nh,(t+1)+T} \\
& * \exp\left(-\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left\{-\gamma^{\theta}\Omega^{\theta}_{nhl}-\gamma^{\theta}a^{\theta}_{nh,(t+1)+T}-\Gamma\right\}\right) \\
& * \exp\left(-\gamma^{\theta}a^{\theta}_{nh,(t+1)+T}(\omega)-\Gamma\right)da^{\theta}_{nh,(t+1)+T}(\omega)
\end{aligned}
$$

Similar to the proofs above we re-define variables:

$$
\begin{aligned}
x_{(t+1)+T} &\equiv \gamma^{\theta}a^{\theta}_{nh,(t+1)+T}(\omega)+\Gamma \\
\lambda_{(t+1)+T} &\equiv \ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(-\gamma^{\theta}\Omega^{\theta}_{nhl}\right) \\
y_{(t+1)+T} &= x_{(t+1)+T}-\lambda_{(t+1)+T}
\end{aligned}
$$

If we substitute for the re-defined variables we get:

$$
\begin{aligned}
E\left[v^{\theta}_{n,(t+1)+T}(\omega)\right](1+\rho)^{T} = & \sum_{h\in J}\int_{-\infty}^{\infty}\left(-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{1}{(1+\rho)\gamma^{\theta}}\left(x_{(t+1)+T}-\Gamma\right)\right) \\
& * \tilde{B}^{\theta}_{nh,(t+1)+T}\exp\left(-x_{(t+1)+T}\right) \\
& * \exp\left(-\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(-x_{(t+1)+T}\right)\exp\left(-\gamma^{\theta}\Omega^{\theta}_{nhl}\right)\right)dx_{(t+1)+T}
\end{aligned}
$$

$$
\begin{aligned}
E\left[v^{\theta}_{n,(t+1)+T}(\omega)\right](1+\rho)^{T} = & \sum_{h\in J}\int_{-\infty}^{\infty}\tilde{B}^{\theta}_{nh,(t+1)+T}\left(-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{1}{(1+\rho)\gamma^{\theta}}\left(x_{(t+1)+T}-\Gamma\right)\right) \\
& * \exp\left(-x_{(t+1)+T}-\exp\left(-x_{(t+1)+T}+\lambda_{(t+1)+T}\right)\right)dx_{(t+1)+T} \\
& = \sum_{h\in J}\tilde{B}^{\theta}_{nh,(t+1)+T}\int_{-\infty}^{\infty}\left(-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{y_{(t+1)+T}+\lambda_{(t+1)+T}-\Gamma}{(1+\rho)\gamma^{\theta}}\right) \\
& * \exp\left(-\lambda^{\theta}_{(t+1)+T}\right)\exp\left(-y_{(t+1)+T}-\exp\left(-y_{(t+1)+T}\right)\right)dy_{(t+1)+T} \\
& = \sum_{h\in J}\tilde{B}^{\theta}_{nh,(t+1)+T}\exp\left(-\lambda^{\theta}_{(t+1)+T}\right) \\
& * \Bigg[\left(-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{\left(\lambda_{(t+1)+T}-\Gamma\right)}{(1+\rho)\gamma^{\theta}}\right) \\
& * \int_{-\infty}^{\infty}\exp\left(-y_{(t+1)+T}-\exp\left(-y_{(t+1)+T}\right)\right)dy_{(t+1)+T} \\
& + \frac{1}{(1+\rho)\gamma^{\theta}}\int_{-\infty}^{\infty}y_{(t+1)+T}\exp\left(-y_{(t+1)+T}-\exp\left(-y_{(t+1)+T}\right)\right)dy_{(t+1)+T}\Bigg]
\end{aligned}
$$

Note that the derivative of $\exp\left(-\exp\left(-y_{(t+1)+T}\right)\right)$ is $\exp\left(-y_{(t+1)+T}-\exp\left(-y_{(t+1)+T}\right)\right)$ and furthermore $\int y_{(t+1)+T}\exp\left(-y_{(t+1)+T}-\exp\left(-y_{(t+1)+T}\right)\right)=\Gamma$, such that, similarly to the proofs above, we can evaluate the integrals at their boundaries:

$$
\begin{aligned}
E\left[v^{\theta}_{n,(t+1)+T}(\omega)\right](1+\rho)^{T} &= \sum_{h\in J}\tilde{B}^{\theta}_{nh,(t+1)+T}\exp\left(-\lambda^{\theta}_{(t+1)+T}\right)\left(-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{\lambda_{(t+1)+T}}{(1+\rho)\,\gamma^{\theta}}\right)\\
&= \sum_{h\in J}\tilde{B}^{\theta}_{nh,(t+1)+T}\exp\left\{-\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(-\gamma^{\theta}\Omega^{\theta}_{nhl}\right)\right\}\\
&* \left(-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{1}{(1+\rho)\,\gamma^{\theta}}\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(-\gamma^{\theta}\Omega^{\theta}_{nhl}\right)\right)\\
&= \sum_{h\in J}\exp\left\{-\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(-\gamma^{\theta}\left(\tau^{\theta}_{nl}-\tau^{\theta}_{nh}+\ln\frac{\mathcal{V}^{\theta}_{h,(t+1)+T}}{\mathcal{V}^{\theta}_{l,(t+1)+T}}\right)\right)\right\}\\
&* \tilde{B}^{\theta}_{nh,(t+1)+T}\Bigg[-\frac{\tau^{\theta}_{nh}}{1+\rho}+\frac{\ln\mathcal{V}^{\theta}_{h,(t+1)+T}}{1+\rho}+\frac{1}{(1+\rho)\,\gamma^{\theta}}\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\\
&* \exp\left(-\gamma^{\theta}\left(\tau^{\theta}_{nl}-\tau^{\theta}_{nh}+\ln\frac{\mathcal{V}^{\theta}_{h,(t+1)+T}}{\mathcal{V}^{\theta}_{l,(t+1)+T}}\right)\right)\Bigg]
\end{aligned}
$$

Re-arranging terms and simplifying we thus get:

$$
\begin{aligned}
E\left[v^{\theta}_{n,(t+1)+T}(\omega)\right](1+\rho)^{T} &= \sum_{h\in J}\tilde{B}^{\theta}_{nh,(t+1)+T}\exp\Big\{\gamma^{\theta}\left[\ln\mathcal{V}^{\theta}_{h,(t+1)+T}-\tau^{\theta}_{nh}\right]\\
&- \ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(\gamma^{\theta}\left[\ln\mathcal{V}^{\theta}_{h,(t+1)+T}-\tau^{\theta}_{nl}\right]\right)\Big\}\\
&* \frac{1}{(1+\rho)\,\gamma^{\theta}}\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(\gamma^{\theta}\left[\ln\mathcal{V}^{\theta}_{l,(t+1)+T}-\tau^{\theta}_{nl}\right]\right)\\
&= \frac{\sum_{h\in J}B^{\theta}_{nh,(t+1)+T}\exp\left\{\gamma^{\theta}\left[\ln\mathcal{V}^{\theta}_{h,(t+1)+T}-\tau^{\theta}_{nh}\right]\right\}}{\sum_{l\in J}B^{\theta}_{nl,(t+1)+T}\exp\left\{\gamma^{\theta}\left[\ln\mathcal{V}^{\theta}_{l,(t+1)+T}-\tau^{\theta}_{nl}\right]\right\}}\\
&* \frac{1}{(1+\rho)\,\gamma^{\theta}}\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(\gamma^{\theta}\left[\ln\mathcal{V}^{\theta}_{l,(t+1)+T}-\tau^{\theta}_{nl}\right]\right)\\
&= \frac{1}{(1+\rho)\,\gamma^{\theta}}\ln\sum_{l\in J}\tilde{B}^{\theta}_{nl,(t+1)+T}\exp\left(\ln\mathcal{V}^{\theta}_{l,(t+1)+T}-\tau^{\theta}_{nl}\right)^{\gamma^{\theta}}\\
&= \frac{1}{1+\rho}\ln\left[\sum_{l\in J}\left\{\exp\left(-\tau^{\theta}_{nl}\right)B^{\theta}_{nl,(t+1)+T}\mathcal{V}^{\theta}_{l,(t+1)+T}\right\}^{\gamma^{\theta}}\right]^{\frac{1}{\gamma^{\theta}}}
\end{aligned}
$$

where we denote as $\mathcal{O}^{\theta}_{n,(t+1)+T}=\frac{1}{1+\rho}\ln\left[\sum_{l\in J}\left\{\exp\left(-\tau^{\theta}_{nl}\right)B^{\theta}_{nl,(t+1)+T}\mathcal{V}^{\theta}_{l,(t+1)+T}\right\}^{\gamma^{\theta}}\right]^{\frac{1}{\gamma^{\theta}}}$ the migration option for region $n\in J$ and with $\ln\mathcal{V}^{\theta}_{l,(t+1)+T}$ as defined above.

When deciding on a migration destination at time period $t+T$ all workers incorporate expectations about future amenity shocks and time paths of wages, prices and amenities in their decisions as well as their (discounted) migration option values. The expected utility of workers of type θ and who were living in region o one period before is then given as

$$E\left[v^{\theta}_{o,t+T}(\omega)\right](1+\rho)^{T-1} \equiv E\left[\max_{n\in J}\left(\frac{1}{1+\rho}\right)\left[a^{\theta}_{on,t+T}(\omega)-\tau^{\theta}_{on}+\ln\mathcal{V}^{\theta}_{n,t+T}\right]\right],$$

with $\ln\mathcal{V}^{\theta}_{n,t+T} \equiv \ln\left[\frac{(1-\iota)w^{\theta}_{n,t+T}A^{\theta}_{n,t+T}}{p^{1-\alpha}_{n,t+T}}\right]+\mathcal{O}^{\theta}_{n,(t+1)+T}$.

With a proof similar to the one above (and a slight change in notation) it is straightforward to show that

$$E\left[v^{\theta}_{o,t+T}(\omega)\right](1+\rho)^{T-1} = \frac{1}{1+\rho}\ln\Big[\sum_{m\in J}\Big\{\exp\left(-\tau^{\theta}_{om}\right)B^{\theta}_{om,t+T}\mathcal{V}^{\theta}_{m,t+T}\Big\}^{\gamma^{\theta}}\Big]^{\frac{1}{\gamma^{\theta}}}.$$

By backwards induction we thus obtain expected worker utility at time period t, as a function of wages, prices and amenities at the destination as well as migration option values

$$\ln\mathcal{U}^{\theta}_{i|k,t} = E\left[v^{\theta}_{i|k,t}(\omega)\right]+\frac{1}{1+\rho}\left[E\left(v^{\theta}_{j|i,t+1}(\omega)\right)+\mathcal{O}^{\theta}_{j,t+2}\right] \tag{25}$$

and $\mathcal{O}^{\theta}_{j,t+2} = \frac{1}{1+\rho}\ln\Big[\sum_{m\in J}\Big(\exp\Big\{E\left[v^{\theta}_{m|j,t+2}(\omega)\right]+\mathcal{O}^{\theta}_{m,t+3}\left[\mathcal{O}^{\theta}_{l,t+4}...\mathcal{O}^{\theta}_{n,(t+1)+T}\right]\Big\}\Big)^{\gamma^{\theta}}\Big]^{\frac{1}{\gamma^{\theta}}}$.

J.3.4 Conditional migration probability

Let $\ln\mathcal{V}^{\theta}_{j,t+1} \equiv \ln\left[\frac{(1-\iota)w^{\theta}_{j,t+1}A^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,t+1}}\right]+\mathcal{O}^{\theta}_{j,t+2}$ be the discounted life-time utility at destination j starting from time period $t+1$ onward. This allows to derive the share of workers $\chi^{\theta}_{ij|i,t}$ who are located in region i and for whom region j offers the highest life-time utility among alternatives $n\in J$ when incorporating forward-looking expectations:

$$\begin{aligned}\chi^{\theta}_{ij|i,t} &= Pr\left\{-\tau^{\theta}_{ij}+\tau^{\theta}_{in}+\ln\frac{\mathcal{V}^{\theta}_{j,t+1}}{\mathcal{V}^{\theta}_{n,t+1}}+a^{\theta}_{ij,t+1}(\omega)\geq a^{\theta}_{in,t+1}(\omega)\ \forall\ \ n\in J\right\}\\ &= \int_{-\infty}^{\infty}f\left(a^{\theta}_{ij,t+1}(\omega)\right)\prod_{n\neq j}F\left[\Omega^{\theta}_{ijn,t+1}+a^{\theta}_{ij,t+1}(\omega)\right]da^{\theta}_{ij,t+1}(\omega),\end{aligned}$$

where we define the compound parameter $\Omega^{\theta}_{ijn,t+1} \equiv -\tau^{\theta}_{ij}+\tau^{\theta}_{in}+\ln\frac{\mathcal{V}^{\theta}_{j,t+1}}{\mathcal{V}^{\theta}_{n,t+1}}$. Substituting the cumulative distribution and density function we get:

$$\begin{aligned}\chi^{\theta}_{ij|i,t} = \int_{-\infty}^{\infty}&\gamma^{\theta}\tilde{B}^{\theta}_{ij,t+1}\exp\left(-\gamma^{\theta}a^{\theta}_{ij,t+1}(\omega)-\Gamma-\tilde{B}^{\theta}_{ij,t+1}\exp\{-\left[\gamma^{\theta}a^{\theta}_{ij,t+1}(\omega)+\Gamma\right]\}\right)*\\ &\prod_{n\neq j}\exp\left(-\tilde{B}^{\theta}_{in,t+1}\exp\{-\gamma^{\theta}\Omega^{\theta}_{ijn,t+1}-\gamma^{\theta}a^{\theta}_{ij,t+1}(\omega)-\Gamma\}\right)da^{\theta}_{ij,t+1}(\omega)\end{aligned}$$

$$\chi^{\theta}_{ij|i,t} = \int_{-\infty}^{\infty} \gamma^{\theta} \tilde{B}^{\theta}_{ij,t+1} \exp\left(-\gamma^{\theta} a^{\theta}_{ij,t+1}(\omega) - \Gamma\right) *$$

$$\exp\left(-\sum_{n\in J} \tilde{B}^{\theta}_{in,t+1} \exp\left[-\gamma^{\theta}\Omega^{\theta}_{ijn,t+1} - \gamma^{\theta} a^{\theta}_{ij,t+1}(\omega) - \Gamma\right]\right) da^{\theta}_{ij,t+1}(\omega)$$

To solve this integral we re-define variables. In particular, we define the following variables:

$$x_{t+1} \equiv \gamma^{\theta} a^{\theta}_{ij,t+1}(\omega) + \Gamma$$

$$\lambda_{t+1} \equiv \ln \sum_{n\in J} \tilde{B}^{\theta}_{in,t+1} \exp\left(-\gamma^{\theta}\Omega^{\theta}_{ijn,t+1}\right)$$

$$y_{t+1} = x_{t+1} - \lambda_{t+1}$$

Substituting in the re-defined variables delivers

$$\begin{aligned}
\chi^{\theta}_{ij|i,t} &= \int_{-\infty}^{\infty} \gamma^{\theta} \tilde{B}^{\theta}_{ij,t+1} \exp\left(-x_{t+1}\right) \exp\left\{-\exp\left(\lambda_{t+1}\right)\exp\left(-x_{t+1}\right)\right\} \frac{1}{\gamma^{\theta}} dx_{t+1} \\
&= \int_{-\infty}^{\infty} \tilde{B}^{\theta}_{ij,t+1} \exp\left(-y_{t+1} - \lambda_{t+1}\right) \exp\left\{-\exp\left(\lambda_{t+1}\right)\exp\left(-y_{t+1} - \lambda_{t+1}\right)\right\} dy_{t+1} \\
&= \tilde{B}^{\theta}_{ij,t+1} \exp\left(-\lambda_{t+1}\right) \int_{-\infty}^{\infty} \exp\left(-y_{t+1} - \exp\left(-y_{t+1}\right)\right) dy_{t+1}
\end{aligned}$$

Then note that the derivative of $\exp\left(-\exp\left(-y_{t+1}\right)\right)$ is $\exp\left(-y_{t+1} - \exp\left(-y_{t+1}\right)\right)$, such that we can evaluate the integral at its boundaries:

$$\chi^{\theta}_{ij|i,t} = \tilde{B}^{\theta}_{ij,t+1} \exp\left(-\lambda_{t+1}\right) * \left[\exp\left(-\exp\left(-y_{t+1}\right)\right)\right]_{-\infty}^{\infty} = \tilde{B}^{\theta}_{ij,t+1} \exp\left(-\lambda_{t+1}\right)$$

Re-substituting for λ_{t+1} and $\Omega^{\theta}_{ijn,t+1}$, we derive the probability of workers of type θ to migrate from region i to region j between time periods t and $t+1$ as

$$\begin{aligned}
\chi^{\theta}_{ij|i,t} &= \frac{\tilde{B}^{\theta}_{ij,t+1}}{\sum_{n\in J} \tilde{B}^{\theta}_{in,t+1} \exp\left(-\gamma^{\theta}\Omega^{\theta}_{ijn,t+1}\right)} \\
&= \frac{\tilde{B}^{\theta}_{ij,t+1}}{\sum_{n\in J} \tilde{B}^{\theta}_{in,t+1} \left[\exp\left(-\tau^{\theta}_{ij} + \tau^{\theta}_{in} + \left[\ln \frac{\mathcal{V}^{\theta}_{j,t+1}}{\mathcal{V}^{\theta}_{n,t+1}}\right]\right)\right]^{-\gamma^{\theta}}}.
\end{aligned}$$

The share of workers of type θ who migrate from region i to region j is increasing in utility at j, the migration option value and bilateral amenities, but decreasing in bilateral migration costs:

$$\chi^{\theta}_{ij|i,t} = \frac{\left(m^{\theta}_{ij} B^{\theta}_{ij,t+1} \mathcal{V}^{\theta}_{j,t+1}\right)^{\gamma^{\theta}}}{\sum_{n\in J}\left(m^{\theta}_{in} B^{\theta}_{in,t+1} \mathcal{V}^{\theta}_{n,t+1}\right)^{\gamma^{\theta}}}, \tag{26}$$

where $\mathcal{V}^{\theta}_{j,t+1} = \exp\left[\ln\left(\frac{(1-\iota)w^{\theta}_{j,t+1}A^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,t+1}}\right) + \mathcal{O}^{\theta}_{j,t+2}\right]$ and $m^{\theta}_{ij} = \exp\left[-\tau^{\theta}_{ij}\right]$.

J.3.5 Expected utilities and sequential moves

In what follows, we derive expected worker utility and migration probabilities under differing worker expectations. We refer to the correct anticipation of future prices as *perfect foresight* and to anticipated future moves subsequent to an initial migration decision as *sequential moves*. We treat regimes with with restrictive assumptions as special cases of a general case in which workers are fully informed. Concretely, we consider the following four cases: The general case with perfect foresight and an arbitrary number of sequential moves ($S \geq 2$); the special case with perfect foresight and exactly one sequential move ($S \geq 1$); perfect foresight and no sequential moves ($S = 0$); static expectations with no sequential moves. For a detailed discussion of the assumptions underlying each of the four cases and a case-by-case comparison of results we refer the interested reader to Section N.

General case with perfect foresight and several sequential moves (S ≥ 2). In the general case, workers anticipate to migrate at least $(S+1) \geq 3$ times over their employment history. When comparing life-time utility at destinations $j \in J$, workers correctly anticipate the whole time path of wages and prices in all regions and build sophisticated expectations about future bilateral amenity shocks. Note that the migration option value in Eq. (26) entails two components: the value of being able to move to any high-utility region $m \in J$ following stochastic amenity shocks one period forward as well as their migration option values, that is the migration option value $\mathcal{O}^{\theta}_{j,t+2}$ will itself be a function of the option values in time period $t+3$. Furthermore under the assumption of perfect foresight the option value $\mathcal{O}^{\theta}_{m,t+3}\left[\mathcal{O}^{\theta}_{l,t+4} \dots \mathcal{O}^{\theta}_{n,(t+1)+T}\right]$ in turn necessarily depends on the whole forward time path of migration option values and is given as

$$\mathcal{O}^{\theta}_{j,t+2} = \frac{1}{1+\rho}\ln\left[\sum_{m\in J}\left(\exp\left\{E\left[v^{\theta}_{m|j,t+2}(\omega)\right] + \mathcal{O}^{\theta}_{m,(t+3)}\left[\mathcal{O}^{\theta}_{l,(t+4)}\dots\mathcal{O}^{\theta}_{n,(t+1)+T}\right]\right\}\right)^{\gamma^{\theta}}\right]^{\frac{1}{\gamma^{\theta}}}.$$

Special case with perfect foresight and one sequential move (S=1). In this special case we assume that workers expect to relocate from their destination $j \in J$ only once. Under this assumption the migration option value in $t+2$ will only depend on the ease at which the discounted time path of wages, prices and amenities in all regions can be accessed from $j \in J$, but not on the forward time path of migration option values. The migration option value in expected worker utility Eq. (25) and migration probabilities in Eq. (26) then simplifies to

$$\mathcal{O}^{\theta}_{j,t+2} = \frac{1}{1+\rho}\ln\left[\sum_{m\in J}\left(\exp\left(-\tau^{\theta}_{jm}\right)B^{\theta}_{jm,t+2}\mathcal{V}^{\theta}_{m,t+2}\right)^{\gamma^{\theta}}\right]^{\frac{1}{\gamma^{\theta}}}$$

$$\text{with}\quad \ln\mathcal{V}^{\theta}_{m,t+2} \equiv \frac{\ln A^{\theta}_{m,t+2}}{\rho} + \frac{E(a^{\theta}_{mm,t+3}(\omega))}{\rho(1+\rho)} + \sum_{s=t+2}^{\infty}\left(\frac{1}{1+\rho}\right)^{s-((t+2)}\ln\left(\frac{(1-\iota)w^{\theta}_{m,s}}{p^{1-\alpha}_{m,s}}\right).$$

As discussed in detail in Section N.2, quantifying the model under the assumption of an arbitrary number of sequential moves creates a dimensionality problem. Guided by the empirical observation that almost 90% of workers switch local labour markets at most twice over their employment history, we solve and quantify the model for this special case of one sequential move.

Special case with perfect foresight and no sequential move (S=0). In this special case we assume that workers do not expect to relocate from their destination $j \in J$. We can therefore set $\mathcal{O}^{\theta}_{j,t+2} = 0$ in Eqs. (25) and (26), such that workers only incorporate the infinite time paths of wages, prices and amenities in j when making their migration decisions. The formulation of expected worker utility and migration probabilities then simplifies further to

$$\ln \mathcal{U}^{\theta}_{i|k,t} = E\left[v^{\theta}_{i|k,t}(\omega)\right] + \frac{1}{1+\rho}\ln\left[\exp\left(-\tau^{\theta}_{ij}\right) B^{\theta}_{ij,t+1}\mathcal{V}^{\theta}_{j,t+1}\right],$$

$$\text{and} \quad \ln \mathcal{V}^{\theta}_{j,t+1} \equiv \frac{\ln A^{\theta}_{j,t+1}}{\rho} + \frac{E(a^{\theta}_{jj,t+2}(\omega))}{\rho(1+\rho)} + \sum_{s=t+1}^{\infty}\left(\frac{1}{1+\rho}\right)^{s-((t+1)}\ln\left(\frac{(1-\iota)w^{\theta}_{j,s}}{p^{1-\alpha}_{j,s}}\right).$$

Special case with static expectations and no sequential move (S=0) In this last special case we assume that workers have *static expectations*, that is they extrapolate the current realization of wages and prices into the future. Under the assumption of an infinite time horizon and constant real wages as well as amenities, this special scenario yields the simplest formulation of expected worker utility and migration probabilities:

$$\ln \mathcal{U}^{\theta}_{i|k,t} = E\left[v^{\theta}_{i|k,t}(\omega)\right] + \frac{1}{1+\rho}\ln\left[\exp\left(-\tau^{\theta}_{ij}\right) B^{\theta}_{ij,t+1}\mathcal{V}^{\theta}_{j,t+1}\right]$$

$$\text{and} \quad \ln \mathcal{V}^{\theta}_{j,t+1} \equiv \frac{1}{\rho}\left(\ln\frac{(1-\iota)w^{\theta}_{j,s}A^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,s}} + \frac{E(a^{\theta}_{jj,t+2}(\omega))}{1+\rho}\right).$$

J.4 Uniqueness

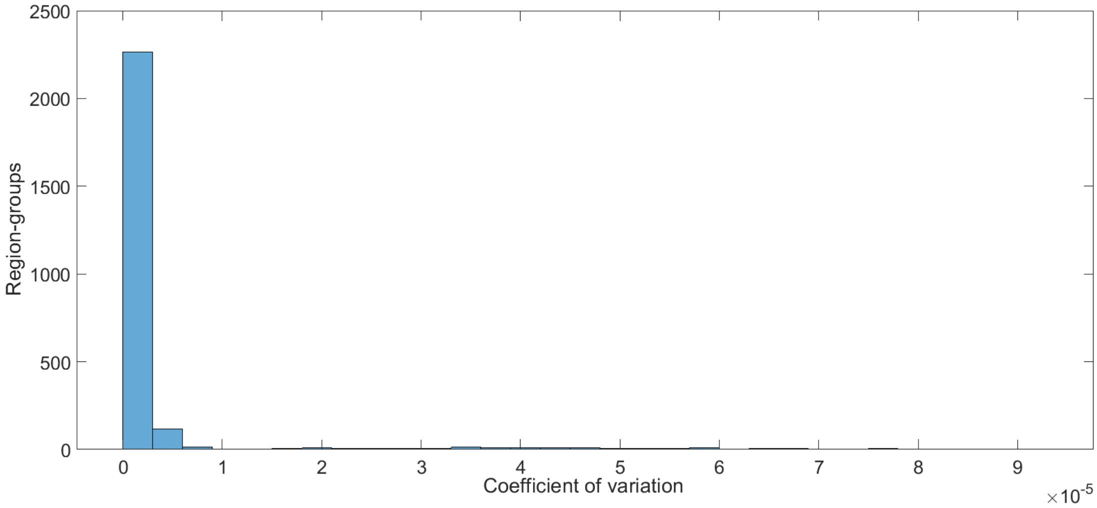
This appendix section complements Section C.4 in the main paper and provides a discussion of equilibrium properties.

Our model features a direct mapping from group-region employment to local wages and rents conditional on structural parameters according to Eqs. (6) and (8). Further, applying the condition that adding one worker to a location raises expenditure more than income ensures mean reversion of the model and all locations will be populated. With respect to income, an immigrating worker exerts a positive production externality on a θ-type worker in the destination region measured by the elasticity κ^{θ} according to Eq. (6). Individual expenditure changes due to responses in housing rents. Combining demand and supply effects and building on Appendix J.1, we obtain

$$(1-\alpha)\frac{\partial p_{i,t}}{\partial L_{i,t}^{\theta}} = (1-\alpha)\beta p_{i,t}\left[\sum_{\theta}\frac{L_{i,t}^{\theta}}{\bar{T}_i}\psi_{i,t}^{\theta}\left(\frac{L_{i,t}}{\bar{T}_i}\right)^{\kappa^{\theta}}\right]^{-1}\frac{L_{i,t}^{\theta}}{\bar{T}_i}\psi_{i,t}^{\theta}\left(\frac{L_{i,t}}{\bar{T}_i}\right)^{\kappa^{\theta}}\left[\frac{1}{L_{i,t}^{\theta}}+\frac{\kappa^{\theta}}{L_{i,t}}\right].$$

For the SSE to hold, Eq. (14) has to be satisfied for all region-group pairs. Conditional on primitives and given the mean-reversion tendency of the model, we find in Monte Carlo simulations that there is a unique employment vector to which the economy converges in the long run. Figure A2 illustrates this insight.

Figure A2: Monte-Carlo simulation - SSE employment



Note: The figure summarises the outcome from 250 Monte Carlo experiments. In each experiment, we hold all primitives constant and use random values of $L_{i,t}^{\theta}$ drawn from a uniform distribution under the constraint $\sum_i L_{i,t}^{\theta} = \bar{L}_t^{\theta}$ to generate a TSE from which we solve for the SSE using the dynamic solver discussed in Sections D.3 and K.3. The histogram illustrates the variation in SSE employment across Monte Carlo experiments within $J \times \Theta = 2,538$ region-groups. The variation is essentially zero, implying that the solver has converged to the same employment values that reference a SSE in all experiments.

J.5 Quality-of-life premiums

In this section, we derive the migration and housing equilibrium loci displayed in Figure 2 in Section C.7 from the structure of our model. Intuitively, the housing equilibrium locus in the real living cost-employment space is a collection of points that satisfy all housing-market-related conditions that must hold in the TSE (and the SSE). Likewise, the migration equilibrium locus satisfies all migration-related conditions that must hold in the SSE. The intersection of both loci is the only point where all equilibrium conditions of the SSE are satisfied and, hence, we can use it to quantify the model and derive QoL premiums.

Housing equilibrium. Inelastically supplied land implies that the cost of supplying housing increases in the regional housing provision. Profit maximisation by developers, perfect competition, and housing market clearing give Eq. (8), which we can rearrange to represent how real living costs are related to housing demand and exogenous housing productivity in equilibrium (housing markets clear in the TSE and the SSE):

$$\ln\left(\frac{p_{i,t}^{1-\alpha}}{w_{i,t}^{\theta}}\right) = (1-\alpha)\,\beta\left(\ln\left[\beta\,(1-\alpha)\,(1-\iota)\,X_{i,t}\right] - \frac{1}{\beta}\ln\left(\eta_{i,t}\right) - \ln\left(\bar{T}_i\right)\right) - \ln\left(w_{i,t}^{\theta}\right).$$

Regional output is the sum over the wage bill of all groups $X_{i,t} = \sum_\theta w_{i,t}^\theta L_{i,t}^\theta$. Wages $w_{i,t}^\theta$ are a function of employment $L_{i,t}^\theta$, exogenous labour productivity $\psi_{i,t}^\theta$ and exogenous land $\bar{T}_i$ as defined in Eq. (6). Therefore, there is a one-to-one mapping from employment to real living cost under the parametrisation discussed in Section D. For the illustration in Figure 2, we use the structural fundamentals inverted for the city of Essen and the parameters estimated for the group of middle-aged, middle-skilled, male workers to derive the housing equilibrium locus HH_1. To obtain the housing equilibrium locus HH_2, we increase housing productivity $\eta_{i,t}$ by 70%.

Since we are already in the real living cost-employment space, it is straightforward to derive the total differential with respect to (log) employment.

$$d\ln\left(p_{i,t}^{1-\alpha}/w_{i,t}^{\theta}\right) = \sum_\theta\left[\frac{(1-\alpha)\,\beta w_{i,t}^\theta L_{i,t}^\theta}{\sum_\theta w_{i,t}^\theta L_{i,t}^\theta}\left(1+\frac{\kappa^\theta L_{i,t}^\theta}{\sum_\theta L_{i,t}^\theta}\right) - \frac{\kappa^\theta L_{i,t}^\theta}{\sum_\theta L_{i,t}^\theta}\right] d\ln L_{i,t}^\theta,$$

where $d\ln L_{i,t}^\theta$ denotes the change in group-specific (log) employment $\ln L_{i,t}^\theta$. For the special case of $\Theta = 1$ (one worker group), the elasticity of real living costs with respect to employment (the slope of the housing equilibrium locus) simplifies to

$$\frac{d\ln\left(p_{i,t}^{1-\alpha}/w_{i,t}\right)}{d\ln L_{i,t}} = (1-\alpha)\,\beta(1+\kappa) - \kappa.$$

In keeping with intuition, real living costs increase faster in city size the larger the land share β (and hence, the smaller the housing supply elasticity) and the smaller the agglomeration elasticity κ.

Migration equilibrium. The supply of labour $L_{i,t}^\theta$ of group θ in city i in period t is the sum over the products of the inbound migration probabilities $\chi_{ji|j,t-1}^\theta$ and employment $L_{j,t-1}^\theta$ across all migration origins j ($\sum_j \chi_{ji|j,t-1}^\theta L_{j,t-1}^\theta$) according to Eq. (12).

Intuitively, higher real living costs make a location less attractive as a migration destination, ceteris paribus. In the SSE, migration markets clear in the sense that the region-group employment is stationary. As a result, the prices of labour and housing are also stationary.

To derive the migration equilibrium locus LL_1 in Figure 2, we again use the structural fundamentals inverted for the city of Essen and the parameters estimated for the group of middle-aged, middle-skilled, male workers. We then take a numerical approach and compute LL_1 under varying living costs. To obtain LL_2, we repeat the exercises, increasing the QoL shifter $A_{i,t}^\theta$ by 60%.

Since the SSE assumption simplifies the expected wage and rent vectors to an infinite projection of the stationary realisations in t, we can derive an analytical solution for the slope for the migration equilibrium locus when furthermore abstracting from the migration option values. Starting from labour supply defined by Eq. (12), we take logs, and then differentiate with respect to the log of real living costs $\frac{d\ln L_{i,t}^{\theta}}{d\ln(p_i^{(1-\alpha)}/w_{i,t}^{\theta})}$. The inverse of this derivative gives the elasticity of real living cost to employment:

$$\frac{d\ln\left(p_{i,t}^{1-\alpha}/w_{i,t}^{\theta}\right)}{d\ln L_{i,t}^{\theta}} = \frac{\rho}{\gamma^{\theta}}\frac{\ln\left(p_{i,t}^{1-\alpha}/w_{i,t}^{\theta}\right)}{\left(1-\chi_{ii|i,t-1}^{\theta}\right)}\frac{\sum_{j\in J}\chi_{ji|j,t-1}^{\theta}L_{j,t-1}^{\theta}}{\chi_{ii|i,t-1}^{\theta}L_{i,t-1}^{\theta}} < 0.$$

Hence, the migration equilibrium locus establishes a negative relationship between real living cost and city employment, which is intuitive given that the inbound migration probabilities $\chi_{ji|,t-1}^{\theta}$ are positively related to the real wage at i via the migration gravity Eq. (10).

The elasticity of real living cost to employment is governed by the variance of idiosyncratic amenities that captures worker heterogeneity. Intuitively, greater worker heterogeneity implies a lower aggregate migration response to real living cost differentials as economic migration incentives will dominate idiosyncratic factors for fewer workers. In the limit $\gamma^{\theta}\to 0$, labour supply becomes perfectly inelastic (a vertical migration equilibrium locus). If workers are homogeneous $\left(\gamma^{\theta}\to\infty\right)$, marginal differences in real living costs trigger large frictionless migration adjustments, resulting in a horizontal migration equilibrium locus.

K Quantification appendix

K.1 Data

This section complements Section D.1 in the main paper. To estimate the crucial structural parameters and invert the structural fundamentals, we require four sets of data compiled for consistent spatial units: Employment, wages, floor space prices, and bilateral migration. In addition, we collect data on determinants of migration costs as well as various location characteristics for overidentification tests and policy simulations. A detailed description of our data is below.

K.1.1 Spatial unit

As an empirical correspondent to locations indexed by i in the model we choose the 141 German labour market regions defined by Kosfeld and Werner (2012). The delineation of these areas is based on combining one or more administrative regions at the county level with the aim of creating self-contained labour markets. The boundary of local labour markets are defined such that commuting within labour market regions is relatively large compared to commuting between regions (subject to an upper limit on commuting time of 45-60 minutes).

K.1.2 Employment

Our measure of employment $L^{\theta}_{i,t}$ is constructed from the Employment History (BeH) covering the years 1993-2018.[7] This dataset is provided by the Institute of Employment Research (IAB) and contains information on the universe of employees in Germany (with the exception of civil servants and the self-employed) on a daily basis. We only select those workers who are employed subject to social security contributions (including apprentices) and who are aged between 16 and 65 years.[8]

Based on this selection we compute the number of employees in each year and labour market region. In addition, we compute region-year-specific employment levels for different groups which are defined according to the interactions between sex, three skill categories (no apprenticeship, completed apprenticeship and tertiary education) and three age categories (16-30 years, 31-50 years and 51-65 years).[9] Employment size varies considerably between labour market regions. While the average number of employees stands at 201,000 in the year 2017, values range from 17,000 in the labour market region *Vulkaneifel* to 1.4 million in *Berlin*.

[7] We use version 10.04.00-190819.

[8] We extract all relevant information from the employment record that contains 30 June of a given year. If a person has multiple employment records, we select according to 1) the average daily wage, 2) the duration of the employment record, 3) at random.

[9] Individuals are assigned the highest qualification level that they achieve over the course of their working life. Consequently, while a person's age changes over time, sex and skill are time-invariant. The educational qualification variable has been processed based on Imputation Procedure 1 described in Fitzenberger et al. (2006).

K.1.3 Migration

We assign workers to labour market regions using their place of employment as reported in the BeH. Bilateral group-specific migration flows are then constructed by computing the number of workers belonging to group θ who used to be employed in region i in year t but who are working in region j in year $t+1$ for every pair of origin region i and destination region j. Based on these bilateral flows we construct group-specific migration probabilities $\chi^{\theta}_{ij,t}$ that are defined as the ratio of the flows from i to j over the level of employment in origin region i in year t. Since labour market regions are designed with the aim of reflecting commuting patterns in a region, we propose that a change in the place of employment across labour market regions is likely to go along with a change of residence.[10]

There are gaps in a worker's employment record in our data, for example if a person was in a different form of employment that is not subject to social security contributions, unemployed or had withdrawn from the labour market. We close such gaps by creating artificial records that duplicate the last available employment record and, in particular, the place of employment. In doing so, we implicitly assume that a person remains in the same labour market region until they find a new regular job in another region, which will be recorded in our data.[11]

K.1.4 Productivity

We use information from the BeH on the universe of workers who are observed as employed subject to social security (including apprentices) on June 30 during the 1993-2018 time period to estimate the group-region-year-specific productivity which maps into the wage. In line with the standard approach in the agglomeration literature (Combes and Gobillon, 2015), we assume in Eq. (5) that worker productivity $\varphi^{\theta}_{i,t}(\omega)$ is a multiplicative function of a group-region-year component $\varphi^{\theta}_{i,t}$ and an individual component $\delta^{\theta}_{i,t}(\omega)$. Following the conventions in labour economics (Abowd et al., 1999), we define $\delta^{\theta}_{i,t}(\omega) = \exp(\bar{\delta}_{\omega} S^{L}_{i,t} z^{L} f^{L,\theta}_{i,\omega,t})$ as a function of unobserved time-invariant individual productivity $\bar{\delta}_{\omega}$ (we use ω as a subscript to index workers), observable worker characteristics $S^{L}_{\omega,t}$ (dummies for whether a worker is in an apprenticeship or works part-time, with z^{L} being the marginal effects) and a stochastic residual term $f^{L,\theta}_{i,\omega,t}$. Log-linearisation and setting individual productivity equal to the nominal wage $\varphi^{\theta}_{i,t}(\omega) = w^{\theta}_{i,\omega,t}$ as predicted under perfect competition (see Section C.2) then gives the estimation equation:

$$\ln w^{\theta}_{i,\omega,t} = \bar{\delta}_{\omega} + S^{L}_{i,t} z^{L} + \tilde{\varphi}^{\theta}_{i,t} + f^{L,\theta}_{i,\omega,t}. \tag{27}$$

In estimating Eq. (27), we remove all observations of individuals who never change their place of employment and estimate the model separately by gender-skill groups for computa-

[10]This assumption is backed up by a considerable degree of overlap between the place of employment and the place of residence. For the year 2017, we find that approximately 75% of employees who work in a specific labour market region also live there. Moreover, use of the place of residence would reduce the available data as this information is only available from 1999 onward.

[11]Notice that this procedure is only used for the computation of migration flows. Estimation of individual-level productivity is therefore unaffected. Approximately 19% of the employment records in the data set are constructed in this way.

tional efficiency. Table A2 shows the results of estimating Eq. (27) for each of the six sex-skill groups. As expected, part-time workers and apprentices have lower expected daily wages. In both cases the wage discount is larger for males than for females and it increases in magnitude with the skill level. Moreover, wages are lower on average in periods when the worker has not yet reached the highest skill level. Given the skill group, male regular full-time workers who have reached their highest skill level have higher wages than females. Likewise, within sex groups the expected wage of regular full-time workers at their highest skill level increases with skill.

Table A2: Estimation of group-region-year productivity

	(1) Female No apprenticeship	(2) Female Apprenticeship	(3) Female Tertiary	(4) Male No apprenticeship	(5) Male Apprenticeship	(6) Male Tertiary
Part-time	-0.331***	-0.351***	-0.441***	-0.437***	-0.455***	-0.559***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Apprentice	-0.798***	-0.854***	-0.953***	-0.933***	-1.017***	-1.039***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Below highest skill	-	-0.162***	-0.167***	-	-0.111***	-0.149***
	(.)	(0.00)	(0.00)	(.)	(0.00)	(0.00)
Constant	3.886***	4.143***	4.415***	4.158***	4.367***	4.689***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Worker effects	Yes	Yes	Yes	Yes	Yes	Yes
Group-region-year effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,690,790	71,274,252	21,087,352	5,427,142	107,566,946	36,3674,105
R^2	.777	.763	.752	.805	.831	.830

Notes: Units of observation are individual-level employment records. The dependent variable is the log average daily wage. $^{+}$ $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We recover $\hat{\hat{\varphi}}^{\theta}_{i,t}$ as a log index of group-region-year-specific productivity which we re-scale such that the group-averages match the group-specific log annual earnings in the raw wage data. We remove a common national trend by running an auxiliary regression of $\hat{\hat{\varphi}}^{\theta}_{i,t}$ against region and year effects and subtracting the latter (using 2017 as the reference category). Exponentiating the regression-adjusted $\hat{\hat{\varphi}}^{\theta}_{i,t}$, we obtain our final region-group-year-specific productivity index $\varphi^{\theta}_{i,t}$.

In Table A3, we test for systematic differences in $\ln \varphi^{\theta}_{i,t}$ across age, gender, and skill groups. Results are shown separately for the period 2007-17, which is used in the empirical analysis (as information on housing prices is only available for those years) as well as for the full period, 1993-2018. Ceteris paribus, female worker productivity is 27% (=(exp(-0.315)-1)*100%) lower than male productivity, with no discernible difference between the two time periods. Workers with an apprenticeship have a predicted productivity that is approximately 45% (=(exp(0.371)-1)*100%) higher than among workers without an apprenticeship, while it is almost twice as high for workers with tertiary education. Whereas the difference in productivities between workers with and without an apprenticeship are almost identical in both time periods, it has increased for university-educated workers. Expected productivity increases with age. It is is 46% (=(exp(0.380)-1)*100%) higher among the age group 31-50 and 64% (=(exp(0.495)-

1)*100%) among the age group 51-65 compared to the youngest age group. Compared to the full time period, it appears that the age gradient has become smaller.

Table A3: Productivity differences

	(1)	(2)
	2007-17	1993-2018
Female	-0.315***	-0.315***
	(0.00)	(0.00)
31-50 years	0.380***	0.472***
	(0.00)	(0.00)
51-65 years	0.495***	0.624***
	(0.00)	(0.00)
Apprenticeship	0.371***	0.373***
	(0.00)	(0.00)
Tertiary education	0.707***	0.667***
	(0.00)	(0.00)
Constant	9.878***	9.817***
	(0.00)	(0.00)
Region effects	Yes	Yes
Year effects	Yes	Yes
Observations	27,918	65,988
R^2	.916	.898

Notes: Units of observation are group-region-year cells. The dependent variable is a group-region-year-specific log productivity measure that is derived as a fixed effect from an individual-level regression of log daily wages that also controls for individual fixed effects. $^{+}$ $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

K.1.5 Housing costs

To compute mix-adjusted indices of purchase prices for a panel of labour market area-year observations, we use the "Real-Estate Data for Germany (RWI-GEO-RED)" micro data discussed in detail by Boelmann and Schaffner (2019). The data originally come from the internet platform ImmobilienScout24 and have been processed and made available for scientific research by the FDZ (Forschungsdatenzentrum) Ruhr. It covers apartments and houses for sale from 2007 to 2017. ImmobilienScout24 is the leading online platform for real estate listings, with a self-reported market share of about 50% (Georgi and Barkow, 2010).

In line with standard practice in urban economics, we model the cost of housing as a rental price whereas in our data we observe purchase prices. Following conventions, we assume that property markets are competitive and investors and owner-occupiers apply a 0.035 discount rate to future streams of actual or imputed rents over an infinite horizon (Koster and Pinchbeck, 2021). Our empirical measure of rent then is $p_{i,t} = 0.035 P_{i,t}$, where $P_{i,t}$ is a location-time-specific house price index following Combes et al. (2019), who in turn build on a long tradition of urban gradient regressions going back to Clark (1951):

$$\ln P_{s,i,t} = \ln D^P_{s,i} u_i + \tilde{S}^P_{s,i,t} z^P_i + \tilde{P}_{i,t} + f^P_{s,i,t}, \tag{28}$$

where $\ln P_{s,i,t}$ is the log of price per square meter floor area of property s, $\ln D^P_{s,i}$ is the distance from the geographic centroid of the municipality with the largest employment in a labour market area, u_i are the destination-specific gradients, $\tilde{S}_{s,t} = S_{s,t} - \bar{S}$ is a vector of property characteristics $S_{s,t}$ net of the national average $\bar{S}$, z^P_i is a vector of destination-specific implicit prices, $\tilde{P}_{i,t}$ is a location-year fixed effect and $f^P_{s,i,t}$ is an unobserved residual. To remove a common national trend, we run an auxiliary regression of $\tilde{P}_{i,t}$ against region effects and year effects and subtract the latter. From the adjusted location-year fixed effect we infer a property price index $P_{i,t} = e^{\tilde{P}_{i,t}}$, which is mix-adjusted for property characteristics and location and representative for a property with the national average characteristics at the centre of a labour market area. In following Combes et al. (2019), we assume that workers are fully mobile and indifferent between locations within monocentric regions indexed by i. Decreasing prices at greater distances from the regional centre offset one for one increasing within-region transport costs. At any other location within a region, quantifying QoL requires accounting for commuting costs (Albouy and Lue, 2015).

The processed data contain a detailed geo-reference, accurate to the level of 1x1 square kilometer grid cells in the European standard ETRS89-LAEA projection. This makes it straightforward to calculate the straight-line distance from a property to the centre of a labour market area, defined as the geographic centroid of the municipality with the largest employment number. Moreover, the data set contains a wide range of property characteristics. However, the degree of coverage varies significantly, with missing values accounting for the majority of observations for selected variables. We focus on control variables with reasonably wide coverage, which include attributes that are typical in hedonic analyses such as the floor area, the number of rooms, the type of property (house vs. flat), the type of heating system and whether features such a balcony, a garden, or a basement belong to the property. There are a limited number of missing values within these variables. For each variable, we set the missing values to zero and generate an auxiliary indicator variable that identifies all observations with a missing value in the selected variable. The mix-adjusted hedonic index we generate then gives the price of a property with the national average in observable characteristics and the average unobserved characteristics of properties with non-missing values in observables, which is located right at the centre of the labour market area. We report summary statistics of observable characteristics in Table A4. The average property has a floor area of about 140 square meters, approximately five rooms, and a 40-percent chance of being an apartment.

K.1.6 Migration distance

We first compute the distance between every pair of municipalities in Germany using the delineation that is valid on 31 December 2018. To derive the distance between any two labour market regions, we form the population-weighted geometric mean of the corresponding munic-

Table A4: House price index: Descriptive statistics

	N	Mean	Stand. dev.	10^{th} pct.	90^{th} pct.
Price per square meter	16,591,919	2,317	225,608	714	3,258
Distance to CBD (in km)	16,591,919	17.45	13.4	2.89	35.98
Living space (in square meter)	16,591,919	141.81	130.13	59	232
Rooms	16,591,919	4.75	2.77	2	8
Type of housing	16,591,919	0.4	0.49	0	1
Balcony	16,591,919	0.28	0.45	0	1
Garden	16,591,919	0.08	0.27	0	0
Basement	16,591,919	0.35	0.48	0	1
Type of heating	16,591,919	7.1	6.14	0	13

Notes: *Type of heating* is a categorical variable between 1 and 13. *Type of housing* is a binary variable with value one for apartments and zero for houses. *Balcony*, *Garden* and *Basement* are also binary variables. Micro data from RWI-Leibniz Institute for Economic Research (Boelmann and Schaffner, 2019).

ipal distances. For a cultural distance measure, we use the inverse of the county-based dialect similarity index by Falck et al. (2012), which we aggregate to labour markets regions.

K.1.7 Big data

Big data amenity. To generate a big-data amenity index, we use geotagged photos shared in social media. They originally stem from Eric Fisher's Geotaggers' World Atlas, whose observations are taken from Flickr and Picasa search APIs.[12] We use about 1.5 million photos taken within the boundaries of German labour market regions, most of which are from the early 2010s, roughly in the middle of our core study period (2007-2017). The idea to use geotagged photos to capture the amenity value of locations was originally proposed by Ahlfeldt (2013), with recent applications including Gaigné et al. (2017), Saiz et al. (2018), and Carlino and Saiz (2019).

We follow Ahlfeldt (2013) and assume that there is a photo production function that links the amenity value A_i^θ to the number of photos shared on social media:

$$\mathcal{P}_i^\theta = c^{\theta P} A_i^{\theta^{\zeta^\theta}} \prod_n (\mathcal{X}_{i,n}^{b_n^{\theta P}}) \epsilon_i^{\theta P}, \tag{29}$$

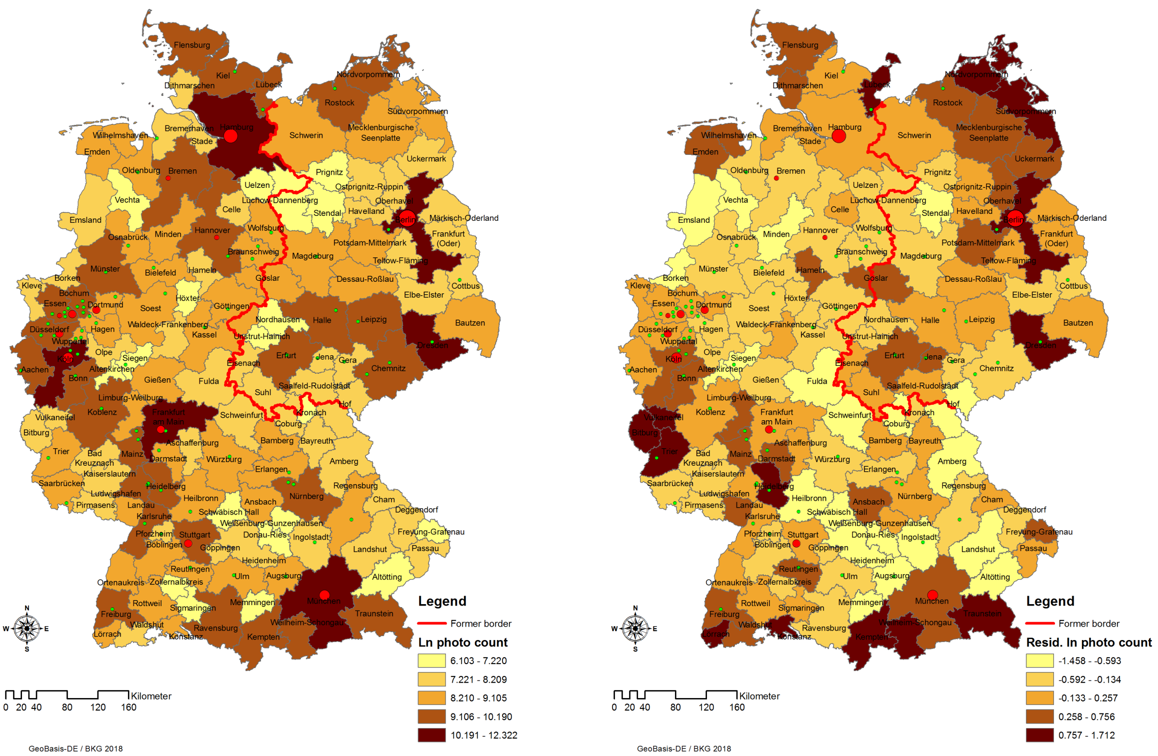
where $\mathcal{X}_{i,n}$ is a set of production factors indexed by n to be specified and $\mathcal{P}_i^\theta = \bar{\mathcal{P}}_i \ \forall \ \theta \in \Theta$ with $\bar{\mathcal{P}}_i$ being the total number of photos. As an example, regional employment L_i may be included since more residents may generate more photos at a constant photo propensity. ζ^θ is the amenity elasticity of photo production, which will be positive if social media users share visually appealing content (e.g. distinctive architecture or scenic views) or interesting activities (e.g. hiking tours or restaurant visits) that are related to a location's endowment with amenities. $\epsilon^{\theta P}$ is a residual term and $\{c^{\theta P}, b_n^{\theta P}\}$ are parameters. Under the assumptions made, we retrieve a big data amenity as $\mathcal{D}_i^\theta = \zeta^\theta \ln A_i{}^\theta + \ln \epsilon_i^{\theta P}$ from the following regression:

[12] See for details http://www.flickr.com/photos/walkingsf/sets/72157623971287575/).

$$\ln \mathcal{P}_i^{\theta} = \tilde{c}^{\theta P} + \sum_n (b_n^{\theta\,P} \ln \mathcal{X}_{i,n}) + \mathcal{D}_i^{\theta}.$$

The interpretation of the big data amenity naturally depends on the covariates in $\mathcal{X}_{i,n}$. We plot an unconditional version excluding any controls in the left panel of Figure A3. Evidently, large urban labour market regions generate more photos. However, this is not necessarily an amenity effect since we expect more populated areas to generate more photos simply because there are more users. In the right panel, we plot a version where we condition on employment and geographic land area. Now, some regions close to the Baltic Sea in the North and the Alps in the South that are popular holiday destinations are also identified as high-amenity areas. From the large labour markets, only Berlin remains in the top category of amenity value. However, controlling for population not only removes the effect on photo production, but also a potential urban quality-of-life premium. Thus, this conditional version of the big data amenity is best interpreted as capturing amenities such as a favourable geography offering scenic views, or historic buildings, but not a vibrant cultural landscape due to restaurant variety, which are typical for large cities.

Figure A3: Photo count and big data amenity



(a) Unconditional big data amenity

(b) Conditional big data amenity

Note: Unit of observation is 141 labour market areas as defined by Kosfeld and Werner (2012). Conditional big data amenity is the log photo count stripped off the effect of log employment and log geographic area in an auxiliary regression.

Social connectedness. We use the Social Connectedness Index (SCI) to measure the strength of social ties that exist between two regions. The SCI is defined as the ratio between the number of friendship connections that exist between Facebook users of any two regions i and j over the product of Facebook users in each of the two regions:

$$SCI_{ij} = \frac{Connections_{ij}}{Users_i \times Users_j}.$$

The variable is then re-scaled so that it ranges from 1 to $1,000,000,000$. A more detailed discussion of the SCI can be found in (Bailey et al., 2018).

Facebook provides the SCI data at the regional level for a broad range of countries (see Bailey et al. (2020) for an application). Within Europe data are available at the third level of the Nomenclature of Territorial Units for Statistics (NUTS). In Germany, NUTS3 regions correspond to counties, so that the data can be aggregated to the level of the labour market region. We first select all region pairs for which both counties are in Germany. We then proceed to compute a weighted average of the SCI over all county pairs within a pair of labour market regions using the sum of the populations in each county pair as a weight.

K.1.8 Location characteristics

Air pollution. We use the concentration of particulate matter to measure air pollution. According to the German Environment Agency (Umweltbundesamt), particulates with a diameter of less than 10 micrometer (PM^{10}) exhibit a particular health risk. We access raw data at the municipality level from the German Environment Agency for 2019. Since there is a direct mapping from municipalities to the local labour markets defined by Kosfeld and Werner (2012), aggregation of the data is straightforward.

Coal deposits and power plants. To compute the coal exposure measure used in the policy application in Section F, we collect data on the spatial distribution of energy resources, especially brown and black coal, from the Federal Institute for Geosciences and Natural Resources in Germany (www.bgr.bund.de). To explore the mechanisms underlying our identification strategy for the estimation of pollution effects, we collect the locations of active coal power plants from the Bundesnetzagentur (www.bundesnetzagentur.de, list of power plants from 1 April 2020).

Wind directions. We obtain wind frequencies by 36 directions for all local labour markets from Kasperski (2002), which we use to generate a wind-adjusted coal exposure measure that serves as an instrument for air pollution.

Fundamentals. We compute a comprehensive data set on fundamental first-nature characteristics that potentially affect productivity (e.g. access to navigable rivers), amenity (e.g. climate), and housing TFP (e.g. physical constraints to development).

World War II destruction. We compile a new dataset based on Hohn (1991) documenting the share of living space destroyed during World War II. The data are available for all German cities with more than 2,000 inhabitants in 1939. Combining this information with average destruction rates per state and population weights for each location, we construct the weighted average share of destroyed living space per labour market region.

K.1.9 Summary statistics

Table A5 provides descriptive statistics for the central variables from the year 2017 that are used for the quantification of our model.

K.2 Structural parameters

This section complements Section D.2 in the main paper by formally deriving estimation equations and providing full estimation results. Before we introduce the technical details and full estimation results in the following subsections, we provide an accessible summary of the key parameters of the model in Table A6 and summarise the variation in group-specific parameters by means of regressions against group-dummies in Table A7 for convenience.

K.2.1 Density elasticity of productivity (κ^θ)

Our empirical approach to the identification of exogenous and endogenous productivity effects is inspired by Combes et al. (2008). We use a conventional AKM-regression described in the data section K.1.4 to separate the group-region-year specific component of productivity $\varphi^\theta_{i,t}$ defined in Eq. (5) from the worker-specific component. Next, we define the exogenous group-region-year productivity as $\psi^\theta_{i,t} = \exp(a^{L,\theta}_g + e^{L,\theta}_{i,t})$, where $a^{L,\theta}_g$ is a group-zone specific effect and $e^{L,\theta}_{i,t}$ is a structural residual. Zone effects capture differences in exogenous productivity between former East Germany and West Germany, indexed by g, due to persistent effects of the division period. Log-linearisation yields the following group-specific regression model, which exactly identifies the group-specific density elasticity of productivity κ^θ and the exogenous group-region-year productivity $\psi^\theta_{i,t}$:

$$\ln \varphi^\theta_{i,g,t} = a^{L,\theta}_g + \kappa^\theta \ln\left(\frac{L_{i,t}}{\bar{T}_i}\right) + e^{L,\theta}_{i,g,t}. \tag{30}$$

Unobserved fundamentals correlated with density pose a threat to identification of κ^θ. Following Ciccone and Hall (1996), we use the deep lag of log population density (1907) as an instrument for the log of contemporary density, arguing that fundamentals that gave rise to density a century ago are of limited relevance for productivity today. Since the instrumental variable is time-invariant, we cluster standard errors on regions.

The resulting estimates of the density elasticity of productivity are presented in Figure A4. The employment-weighted average estimate for κ is 0.025, close to the consensus of about 0.02 in the literature (Combes and Gobillon, 2015). There is significant heterogeneity across

Table A5: Summary statistics

	N	Mean	Std dev
Bilateral flows and distances			
Migration flow	357,858	81.12	1,929.79
Ln distance	19,740	5.62	0.60
Ln cultural distance	19,881	0.03	0.01
Ln social connectedness	19,881	-8.99	0.99
Employment variables			
Ln employment	2,538	8.25	1.58
Employment share: Female (%)	141	46.38	2.95
Employment share: Apprenticeship (%)	141	75.79	5.68
Employment share: Tertiary education (%)	141	15.76	4.99
Employment share: 31-50 years (%)	141	46.15	1.64
Employment share: 51-65 years (%)	141	32.48	2.93
Employment share: Agriculture (%)	141	1.09	1.14
Employment share: Construction (%)	141	6.69	1.90
Employment share: Tradable services (%)	141	9.43	4.20
Employment share: Manufacturing (%)	141	25.60	9.06
Employment share: Energy-intensive heavy industry (%)	141	5.37	3.34
Ln employment density	141	4.07	0.79
Wages and rents			
Ln wage	2,538	10.37	0.36
Ln price	141	4.22	0.57
Structural fundamentals			
Ln quality of life (DSM)	2,538	0.53	0.55
Ln quality of life (Rosen-Roback)	2,538	-8.94	0.39
Regional characteristics			
Ln area	141	7.69	0.56
East Germany (dummy)	141	0.23	0.42
Near Alps (dummy)	141	0.02	0.14
Near coast (dummy)	141	0.11	0.31
Ln historic population density	141	4.65	0.68
Ln crime per capita	141	-6.30	0.31
Housing stock destroyed in WWII (%)	141	9.60	9.72
Number of opera houses	141	0.80	1.08
Ln water area	141	17.39	1.00
Big data amenity index (residualised)	141	0.00	0.43
Pollution variables			
Ln pollution (PM^{10})	141	2.62	0.12
Number of active coal plants	141	0.52	1.37
Ln meteorological black coal exposure (net of geographical exposure)	141	-0.11	0.46
Ln meteorological brown coal exposure (net of geographical exposure)	141	-0.24	0.26

Notes: Number of observations differ: 141 regions; 141 regions x 18 groups = 1,551 region-groups; 141 regions x 141 regions = 19,881 region pairs; 141 regions x 141 regions x 18 groups = 357,858 region-pair-groups. Distance is not defined when origin and destination regions are identical.

Table A6: Parameter values

	Parameter	Value	Approach	Source	Appendix
1-α	Share of expenditure on housing	0.33	Set	Statistisches Bundesamt (2020)	-
ρ	Discount rate	0.11	Set	Moore and Viscusi (1988)	-
ι	Tax rate	0.49	Set	OECD (2017)	-
β	Share of land in housing	0.19	Estimated	Combes et al. (2019)	K.2.2
κ^θ	Density elasticity of productivity	0.001-0.042	Estimated	Combes et al. (2008)	K.2.1
γ^θ	Migration elasticity	0.043-0.60	Estimated	Artuç et al. (2010)	K.2.3
$\tau^\theta_{i,j=i}$	Migration cost (iceberg)	0	Set	Assumption	-
$\tau^\theta_{i,j\neq i}$	Migration cost (iceberg)	6.4-68.3	Estimated	OD-FE in migration gravity | γ^θ	K.2.4
$B^\theta_{ij,t}$	Bilateral amenity	0.03-17.84	Estimated	Residual in migration gravity | γ^θ	K.2.5

Notes: If the approach is "set", we borrow a parameter value from the paper given under "source". If the approach is "estimated", we estimate the parameter following the estimation strategy in the paper given under "source". For details, we refer to the section given under "appendix".

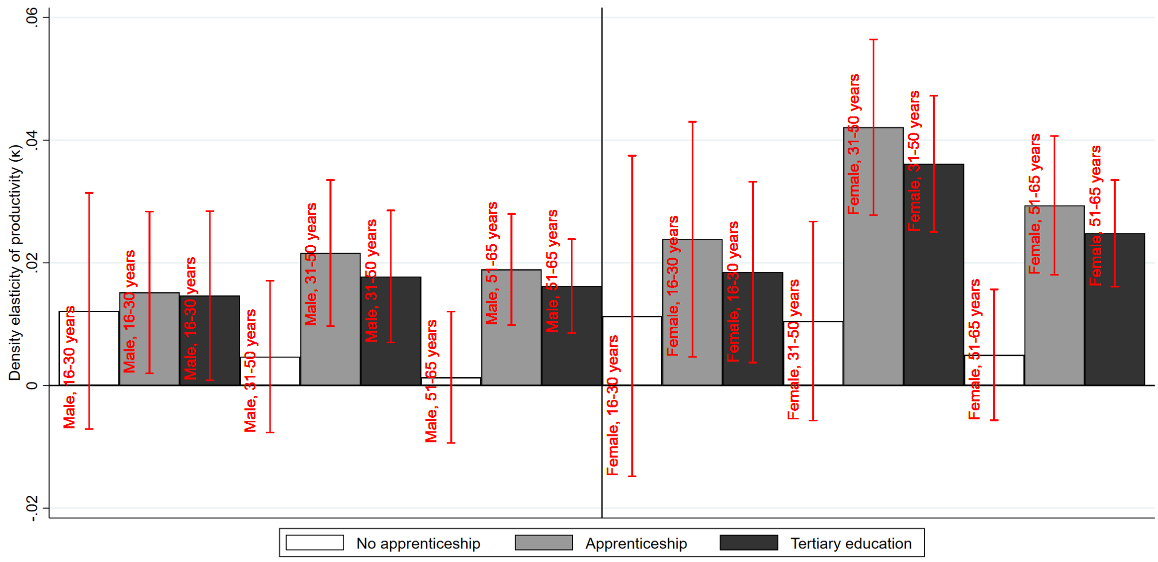
Table A7: Parameter estimates: Average effects by group

	Agglomeration elasticity	Migration resistance	Migration elaticity	Migration iceberg cost	Monetised migration cost
	κ^θ	$\ln \tau^\theta_{ij} \times \gamma^\theta$	γ^θ	τ^θ_{ij}	€1000
Female	0.014***	0.284***	-0.141***	9.669***	-51.783***
	(0.00)	(0.00)	(0.02)	(0.01)	(0.04)
31-50 years	0.010**	0.524***	0.153***	-7.963***	32.498***
	(0.00)	(0.00)	(0.03)	(0.01)	(0.05)
51-65 years	0.003	0.839***	0.148***	-8.177***	37.859***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.06)
Apprenticeship	0.016**	0.427***	0.239***	-24.122***	46.342***
	(0.01)	(0.00)	(0.04)	(0.02)	(0.08)
Tertiary education	0.012*	-0.256***	0.071^{+}	-16.032***	120.518***
	(0.01)	(0.00)	(0.04)	(0.02)	(0.08)
Constant	-0.001	6.436***	0.146***	44.496***	108.045***
	(0.01)	(0.00)	(0.04)	(0.02)	(0.08)
Unit	Group	O-D-group	Group	O-D-group	O-D-group
O-D effects	-	Yes	-	Yes	Yes
Observations	18	355320	18	355320	355320
R^2	.86	.971	.935	.936	.944

Notes: O = origin; D = destination. All explanatory variables are binary indicator variables. Standard errors in parentheses. O-D-group-level regressions weighted by O-D-group flows.

worker groups, with κ^{θ} estimates ranging from close to zero for young male workers to 0.042 for skilled and experienced female workers. In line with skill-biased returns to agglomeration (Baum-Snow and Pavan, 2013), we generally obtain greater κ^{θ} estimates for groups with higher skills. There is also a systematic gender gap in κ^{θ} favouring women, implying a greater gender pay gap in rural areas. Finally, young groups benefit little from agglomeration, suggesting that the productivity advantage associated with urban density materialises through an interaction with experience. An econometric analysis of the conditional variation in κ^{θ}-estimates by group is in Table A7.

Figure A4: Density elasticity of productivity (κ)



Notes: Elasticity estimates from regressions of AKM-adjusted log wages (see section K.1.5) against log density, controlling for zone effects (former East vs. former West Germany) and using 1907 log population density as an instrument. Confidence bands are at the 95% level.

K.2.2 Land share in housing (β)

We use a similar approach as in K.2.1 to identify the exogenous and endogenous determinants of housing costs. We define exogenous housing TFP as $\eta_{i,t} = -\exp(\tilde{a}_g^P + e_{i,t}^P)$, where $\tilde{a}_g^P$ captures zone-specific legacy effects from the division period and $e_{i,t}^P$ is a structural residual. Log-linearisation of Eq. (8) then yields the empirical specification:

$$\ln p_{i,g(i),t} = a_g^P + \beta \ln\left(\frac{X_{i,t}}{\bar{T}_i}\right) + e_{i,g(i),t}^P, \tag{31}$$

where $a_g^P = \beta \ln(1-\alpha)\beta(1-\iota) + \tilde{a}_g^P$ collects all scalars in Eq. (8) and the effects of zone-specific housing TFP. Given set values for α and ι and an estimated value for β, exogenous housing TFP is uniquely identified as $\eta_{i,t} = ((1-\alpha)\beta(1-\iota))^{\beta}(X_{i,t}/\bar{T}_i)^{\beta}/p_{i,t}$. To address the concern that contemporary productivity shocks may be correlated with output and housing

Table A8: Output density elasticity of housing cost

	(1) Log housing costs (region-year-specific)	(2) Log housing costs (region-year-specific)
Log output density (β)	0.189*** (0.02)	
Log employment density		0.195*** (0.02)
Zone effects	Yes	Yes
Observations	1,551	1,551
R^2	.317	.299

Notes: Units of observation are labour market region-year cells. *Housing costs* is the annualised house price index inferred from micro data as described in the data section K.1.5. We use the 1907 log population density as an instrument for log of output density and log employment density. Zone effects distinguish between former East and West Germany. Standard errors in parentheses clustered on labour market areas.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TFP, we use the deep lag of population density as an instrument for output density and cluster standard errors on regions.

In Column (1) in Table A8 we obtain an estimate of the output elasticity of housing costs β of 0.189. Note that because in our framework productivity varies across locations, there is a density-induced demand-side effect on wages in addition to the supply-side effect of employment density on housing costs that arises because of inelastically supplied land (see Appendix J.1 for a formal derivation). Thus, unlike Combes et al. (2019) who model the cost of agglomeration as dependent on population and land area, we have output density $X_{i,t}/\bar{T}_i = (\sum_\theta L^\theta_{i,t}\varphi^\theta_{i,t})/\bar{T}_i$ on the right-hand side of the structural specification. For comparison, we also estimate the employment density elasticity in Column (2), which takes the value of 0.195. This value is between the average in the literature of 0.15 reported by Ahlfeldt and Pietrostefani (2019) and the predicted value of 0.25 for a country with the urban density of Germany (2,800 residents per km^2, see Demographia (2019)) according to the rule of thumb suggested by Ahlfeldt and Pietrostefani (2019). The value is towards the lower bound of the 0.2-0.27 range reported for France by Combes et al. (2019), which is consistent with France having a higher urban density (3,100 residents per km^2) than Germany. Notice that the estimated density elasticity of housing expenditure $(1-\alpha)\frac{\partial \ln p_{i,t}}{\partial \ln (L_{i,t}/\bar{T})} = 0.066$ substantially exceeds our κ^θ-estimates for all groups, which is necessary for a well-behaved solution for the SSE. Note that our estimate of β implies a housing supply elasticity $(1-\beta)/\beta$ of about 4.3, which is close to existing structural estimates (Epple et al., 2010).

K.2.3 Migration elasticity (γ^θ)

The standard approach to the identification of the migration elasticity is to regress relative (to internal migration) log migration flows against bilateral differences in log wages at migration origins and destinations, controlling for leading relative log migration flows (Artuç et al., 2010).

To motivate a similar estimating equation, we start from Eq. (10) and derive the difference in migration propensity between stayers and movers in the special case of no sequential move as (see section J.3.5 for a derivation):

$$\begin{aligned} \ln \chi^{\theta}_{ij|i,t} - \ln \chi^{\theta}_{ii|i,t} = \gamma^{\theta}\left(\ln m^{\theta}_{ij} - \ln m^{\theta}_{ii}\right) + \gamma^{\theta}\left(\ln B^{\theta}_{ij,t+1} - \ln B^{\theta}_{ii,t+1}\right) \\ + \gamma^{\theta}\left(\ln \mathcal{V}^{\theta}_{j,t+1} - \ln \mathcal{V}^{\theta}_{i,t+1}\right) \end{aligned} \tag{32}$$

with $\ln \mathcal{V}^{\theta}_{j,t+1}$ the infinite sum of indirect utilities. This sum can be re-written as a sum of utility in period $t+1$ and the present value of future utilities in the subsequent periods:

$$\begin{aligned} \ln \mathcal{V}^{\theta}_{j,t+1} = & \underbrace{\ln\left(\frac{(1-\iota)\, w^{\theta}_{j,t+1} A^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,t+1}}\right)}_{\text{utility in period t+1}} \\ & + \underbrace{\sum_{s=t+2}^{\infty}\left(\frac{1}{1+p}\right)^{s-(t+1)} \ln\left(\frac{(1-\iota)\, w^{\theta}_{j,s} A^{\theta}_{j,s}}{p^{1-\alpha}_{j,s}}\right) + \frac{\ln B^{\theta}_{jj,t+2}}{\rho\,(1+\rho)}}_{\text{present value of future utilities in the subsequent periods}} \end{aligned}$$

The infinite sum of indirect utilities in the next period simply corresponds to the present value of future utilities from $t+1$, discounted by one period:

$$\ln \mathcal{V}^{\theta}_{j,t+2} = (1+\rho)\sum_{s=t+2}^{\infty}\left(\frac{1}{1+p}\right)^{s-(t+1)} \ln\left(\frac{(1-\iota)\, w^{\theta}_{j,s} A^{\theta}_{j,s}}{p^{1-\alpha}_{j,s}}\right) + \frac{\ln B^{\theta}_{jj,t+3}}{\rho\,(1+\rho)}$$

so that we have

$$\ln \mathcal{V}^{\theta}_{j,t+1} = \ln\left(\frac{(1-\iota)\, w^{\theta}_{j,t+1} A^{\theta}_{j,t+1}}{p^{1-\alpha}_{j,t+1}}\right) + \frac{1}{1+\rho}\ln \mathcal{V}^{\theta}_{j,t+2}. \tag{33}$$

Moreover, $\ln \mathcal{V}^{\theta}_{j,t+2}$ is a determinant of migration probabilities in period $t+2$:

$$\begin{aligned} \ln \chi^{\theta}_{ij|i,t+1} - \ln \chi^{\theta}_{ii|i,t+1} = \gamma^{\theta}\left(\ln m^{\theta}_{ij} - \ln m^{\theta}_{ii}\right) + \gamma^{\theta}\left(\ln B^{\theta}_{ij,t+2} - \ln B^{\theta}_{ii,t+2}\right) \\ + \gamma^{\theta}\left(\ln \mathcal{V}^{\theta}_{j,t+2} - \ln \mathcal{V}^{\theta}_{i,t+2}\right). \end{aligned} \tag{34}$$

Hence, we can use Eq. (34) in Eqs. (33) and (32) to write current relative migration propensities as a function of bilateral wages next period and relative migration propensities one period forward:

$$\begin{aligned}
&\ln \chi^{\theta}_{ij|i,t} - \ln \chi^{\theta}_{ii|i,t} - \gamma^{\theta}\left[\ln\left((1-\iota)\,w^{\theta}_{j,t+1}\right) - \ln\left((1-\iota)\,w^{\theta}_{i,t+1}\right)\right] \\
&- \frac{1}{1+\rho}(\ln \chi^{\theta}_{ij|i,t+1} - \ln \chi^{\theta}_{ii|i,t+1}) + \left(1-\frac{1}{1+\rho}\right)\gamma^{\theta}\tau^{\theta}_{ij} \\
&= \gamma^{\theta}\left(\ln\left(p^{\alpha-1}_{j,t+1}\right) - \ln\left(p^{\alpha-1}_{i,t+1}\right)\right) + \gamma^{\theta}\left(\ln B^{\theta}_{ij,t+1} - \ln B^{\theta}_{ii,t+1}\right) \\
&- \frac{\gamma^{\theta}}{1+\rho}\left(\ln B^{\theta}_{ij,t+2} - \ln B^{\theta}_{ii,t+2}\right) + \frac{\gamma^{\theta}(\ln B^{\theta}_{jj,t+2} - \ln B^{\theta}_{ii,t+2})}{\rho\,(1+\rho)} \\
&- \frac{\gamma^{\theta}(\ln B^{\theta}_{jj,t+3} - \ln B^{\theta}_{ii,t+3})}{\rho\,(1+\rho)^2} + \gamma^{\theta}\left(\ln A^{\theta}_{j,t+1} - \ln A^{\theta}_{i,t+1}\right).
\end{aligned} \tag{35}$$

Following Artuç et al. (2010), we estimate our key parameter of interest using GMM. To this end, we collect the terms on the right-hand side of Eq. (35) in a structural residual $\mathcal{E}^{\theta}_{ij,t}$, take α, ι, and ρ as given, and normalise all variables by their geometric within-origin-destination-group mean, which removes time-invariant migration costs τ^{θ}_{ij}. To identify γ^{θ} we make the following identifying assumption:

$$\mathbb{E}(\bar{\mathbf{Z}}^{\theta}_{\mathbf{ij,t}}\bar{\mathcal{E}}^{\theta}_{ij,t}) = 0, \tag{36}$$

where $\bar{\mathbf{Z}}^{\theta}_{\mathbf{ij,t}}$ is a $(n \geq 1)\times 1$ vector of instrumental variables which we require to be uncorrelated with the structural residual and the upper bar indicates normalisation by the geometric mean. Substituting Eq. (35) into Eq. (36) (via $\mathcal{E}^{\theta}_{ij,t}$), we obtain n moment conditions:

$$\begin{aligned}
&\mathbb{E}\left(\bar{\mathbf{Z}}^{\theta}_{\mathbf{ij,t}}\left[\ln \bar{\chi}^{\theta}_{ij|i,t} - \ln \bar{\chi}^{\theta}_{ii|i,t} - \gamma^{\theta}\left(\ln\left(\bar{w}^{\theta}_{j,t+1}\right) - \ln\left(\bar{w}^{\theta}_{i,t+1}\right)\right)\right.\right. \\
&\left.\left.- \frac{1}{1+\rho}(\ln \bar{\chi}^{\theta}_{ij|i,t+1} - \ln \bar{\chi}^{\theta}_{ii|i,t+1})\right]\right) = 0.
\end{aligned} \tag{37}$$

Eq. (37) excludes rents which are in the structural residual $\mathcal{E}^{\theta}_{ij,t}$. This not only makes the estimation equation similar to the literature, it also avoids an endogeneity problem since our model predicts that amenities in the structural residuals (A and B terms) capitalise into rents.

The conventional approach is to estimate Eq. (37) using lagged values of relative migration probabilities and relative wages as instruments for leading relative migration probabilities and relative wages (Artuç et al., 2010; Caliendo et al., 2019b). This approach addresses the concern that *contemporaneous* shocks that affect wages and leading migration decisions may also affect components of the structural residual term. A remaining concern is that if there is serial correlation in the instrumented variables and the structural residuals, the identifying assumption will be violated.

Against this background, we consider it worth exploiting an alternative source of identifying variation that is specific to the German context. It is rather uncontroversial that after unification former East Germany had a lower fundamental labour productivity owing to an inferior production technology (Burda and Hunt, 2001). Over time, western technology diffused to the eastern parts of Germany, reducing the differences in fundamental productivity inherited from the cold-war era. To capture the convergence in fundamental labour productivity between the formerly separated parts of the country, we compute for each year, group, and bilateral route a regional relative wage, where we replace the observed wages for labour markets i and j with the average wage in zone $r, s \in East, West$ a labour market falls in. We then use lags of these relative zone wages $\left(\ln\left(\bar{w}^{\theta}_{s,t}\right) - \ln\left(\bar{w}^{\theta}_{r,t}\right)\right)$ as sole (excluded) instruments for the identification of the parameters of interest. Effectively, this approach restricts the identifying variation to changes in cross-border differences in wages over time.

The GMM estimation results are in Table A9. With the canonical instrumental variables we estimate a migration elasticity of about 0.1 (Column 1), which is significantly below the implied value of 0.5 for year-on-year variation reported by Caliendo et al. (2019b). With our preferred identification using the zone wage gap, we estimate a migration elasticity of 0.4 (Column 2), which is closer to the literature. A cause for concern is that the discount parameter is either very large (Column 2) or negative and, hence, theory-inconsistent (Column 1). This is in line with the notion in the literature that the identification of these parameters with the state-of-the-art estimation strategy is weak (Artuç et al., 2010). Hence, we repeat the estimation, setting the discount parameter to our preferred value of 0.11 taken from the literature (Moore and Viscusi, 1988). Once we do this, reassuringly, the migration elasticity estimates using both sets of instruments are close (Columns 3 and 4). Our preferred estimate of the migration elasticity of 0.295 (Column 4) is moderately smaller than the 0.5-estimate for the U.S. by Caliendo et al. (2019b), which implies that workers in Germany are, on average, somewhat less responsive to migration incentives than in the U.S.

To obtain group-specific estimates of the migration elasticity, we build on our preferred specification (Column 4) whose estimation we repeat sequentially, keeping only specific gender, age, and skill groups. Thus, we estimate 2 (gender groups) + 3 (age groups) + 3 (skill groups) = 8 specifications. Compared to the alternative of estimating $2 \times 3 \times 3 = 18 = \Theta$ group-specific models, this approach is less susceptible to producing outlier estimates while still allowing for sizable heterogeneity. We disaggregate our gender- (γ^g), age- (γ^a), and skill- (γ^s) specific

Table A9: Migration elasticity estimates (uniform)

	(1)	(2)	(3)	(4)
Migration elasticity	0.118***	0.443***	0.255***	0.295***
γ	(0.03)	(0.15)	(0.02)	(0.08)
Discount	-0.274***	0.376***	-	-
parameter ρ	(0.01)	(0.03)	-	-
Parameter ρ	Estimated	Estimated	Set to 0.11	Set to 0.11
IV	Canonical	Regional wage gap	Canonical	Regional wage gap

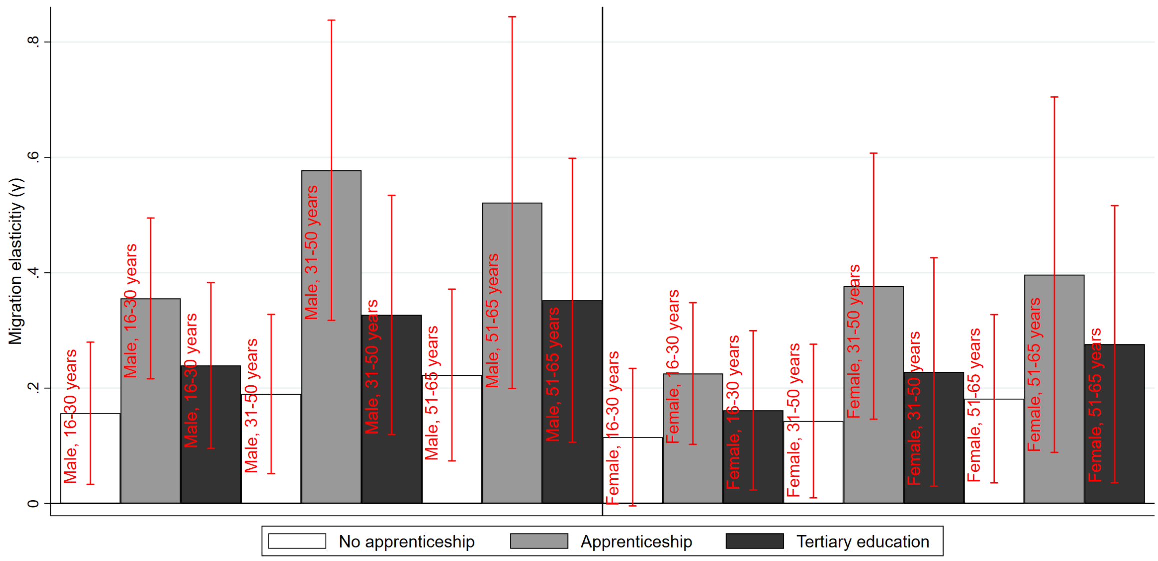
Notes: GMM estimation. Unit of observation is year-group-region-route (origin-destination pair). Weighting. Canonical instrumental variables are lags 1-3 of relative migration probabilities and relative wages. Regional wage gap instrumental variables are lags 1-3 of the year-group-route-specific difference in the regional average wage, where regions are former East- and West-Germany. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

estimates to group-θ-specific estimates as follows:

$$\gamma^{\theta(g,a,s)} = w^{\theta,g}\gamma^g + w^{\theta,a}\gamma^a + w^{\theta,s}\gamma^s,$$

where the weights are defined as the size of a specific θ-group relative to the size of the age-, sex- or skill-group. We obtain standard errors for the resulting $\gamma^{\theta(g,a,s)}$ by means of bootstrapping in 1,000 iterations. The results in Figure A5 reveal sizable heterogeneity in the migration elasticity across groups. In particular, it appears that the migration elasticity is larger for male than for female workers. It is largest for the middle skill and the middle age category.

Figure A5: Migration elasticity estimates (γ) by group



Note: GMM estimates by gender, age, and skill groups, disaggregated to gender-age-skill groups. Bootstrapped standard errors in 1,000 iterations.

K.2.4 Migration cost (τ_{ij}^{θ})

A log-linearised version of Eq. (10) provides the micro foundations for a non-parametric reduced-form migration gravity equation:

$$\ln M_{ij,t}^{\theta} = c^{\theta} + O_{i,t}^{\theta} + D_{j,t}^{\theta} + \tilde{m}_{ij}^{\theta} + \tilde{B}_{ij,t}^{\theta}, \tag{38}$$

where empirically we use the group-specific flow of workers leaving region i for region j in year t after moving $\ln(L_{i,t}^{\theta})$ into the origin-year fixed effect

$$O_{i,t}^{\theta} = \ln\left(L_{i,t}^{\theta}\right) + \ln\left(\sum_{n\in J}(B_{in,t}^{\theta}\mathcal{V}_{n,t}^{\theta}m_{in}^{\theta})^{\gamma^{\theta}}\right)$$

that also captures multilateral resistance. $D^{\theta}_{j,t} = \gamma^{\theta} \ln(\mathcal{V}^{\theta}_{j,t})$ is a destination-group-year effect capturing migration pull factors while $\tilde{B}_{ij,t} = \gamma^{\theta} \ln B_{ij,t}$ is a structural residual capturing bilateral amenity, and $\tilde{m}^{\theta}_{ij} = c^{\theta} - \gamma^{\theta} \times \tau^{\theta}_{ij}$ is an origin-destination effect identifying migration resistance up to a group-specific constant c^{θ}. We use the theory-consistent restriction $\tau^{\theta}_{ij,j=i} = 0$, which implies that $c^{\theta} = \tilde{m}_{ij,j=i}$ to identify $\tau^{\theta}_{ij} = \frac{\tilde{m}^{\theta}_{ij,j=1} - \tilde{m}^{\theta}_{ij}}{\gamma^{\theta}}$.

We estimate Eq. (38) using a Poisson Pseudo Maximum Likelihood estimator (Head and Mayer, 2014). The non-parametric nature of Eq. (38) implies that we require no identifying assumption other than that group-specific shocks to bilateral amenity $B^{\theta}_{ij,t}$ are random within origin-destination pairs. For selected origin-destination routes, we do not observe any migration flow throughout our observation period. In these cases, we impute $\tilde{m}^{\theta}_{ij}$ using a group-specific higher-order polynomial regression of $\tilde{m}^{\theta}_{ij}$ against bilateral distance.

In Figure A6, we present the distribution of the estimated migration resistance effects $\gamma^{\theta}\tau^{\theta}_{ij,t} = \tilde{m}^{\theta}_{ij}$ by group and geographic distance. These reduced-form effects control for arbitrary migration push and pull factors and provide first evidence on which groups exhibit the largest resistance to migration, either because they face large migration costs (reflected in a large τ^{θ}_{ij}), or because of limited idiosyncrasy in their location choice (reflected in a large γ^{θ}). Migration resistance increases in distance at a decreasing rate. There is a kink at about 100 km. The differences in migration resistance across groups are also quantitatively important as revealed by the results from a regression of the estimated resistance parameters against categorical group identifier variables presented in Table A7. The migration resistance of old workers (age between 51 and 65 years) is 131% (=(exp(0.839)-1)*100%) larger than that of young workers (aged 16-30). Likewise, women have an about 33% (=(exp(0.284)-1)*100%) higher migration resistance than men. Skilled (apprenticeship) and high-skilled (tertiary education) workers' migration resistance is about 53% (=(exp(0.0.427)-1)*100%) higher and 23% (=(exp(-0.256)-1)*100%) lower than for unskilled workers (no apprenticeship).

K.2.5 Bilateral amenity (B^{θ}_{ij})

From Eq. (38), it is straightforward to recover $B^{\theta}_{ij} = \left(\tilde{B}^{\theta}_{ij}\right)^{\frac{1}{\gamma^{\theta}}}$. In a theory-consistent manner, we rationalise zero-migration flows with origin-destination-group-year cells by setting $B^{\theta}_{ij} = 0$.

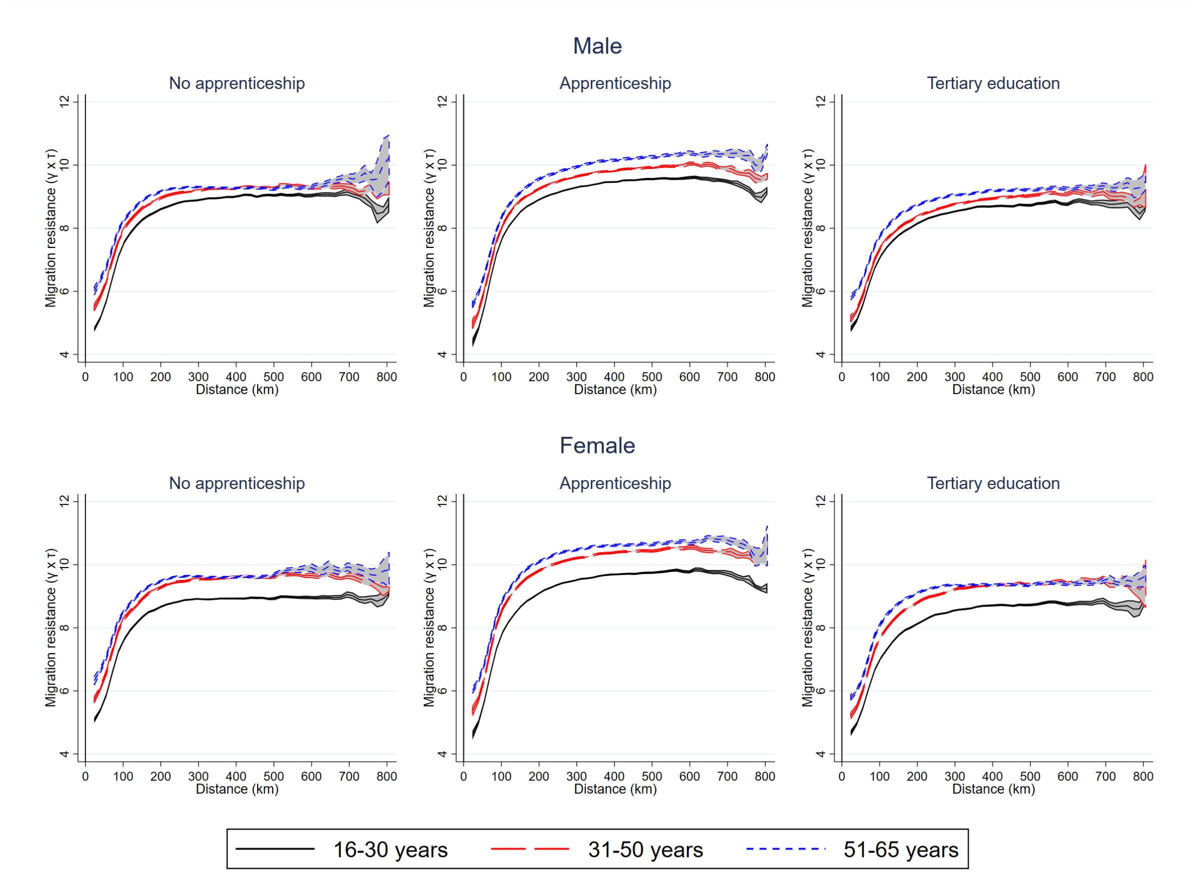
K.3 Structural fundamentals

This section complements Section D.3 in the main paper. We show how to invert fundamental labour and housing productivity and introduce the dynamic solver used to invert QoL.

K.3.1 Fundamental labour productivity

We invert fundamental labour productivity $\psi^{\theta}_{i,t}$ using observed data on mix-adjusted wages $w^{\theta}_{i,t}$, employment $L^{\theta}_{i,t}$, land area $\bar{T}_i$, our estimate of the density elasticity of productivity κ^{θ} and the first-order condition of labour demand using Eq. (6) as follows:

Figure A6: Migration resistance by group and distance



Notes: Migration resistance identified as origin-destination-group effects from panel PPML estimation of a migration gravity model controlling for origin-year-group and destination-year-group effects. Confidence bands are at the 95% level.

$$\psi_{i,t}^{\theta} = w_{i,t}^{\theta} \left(\frac{L_{i,t}}{\bar{T}_i} \right)^{-\kappa^{\theta}}.$$

K.3.2 Fundamental housing productivity

We invert fundamental housing productivity $\eta_{i,t}$ using observed data on mix-adjusted housing rents $p_{i,t}$, employment $L_{i,t}^{\theta}$, wages $w_{i,t}^{\theta}$, land area $\bar{T}_{i,t}^{\theta}$, our estimate of the land share β and housing market clearing using Eq. (8) as follows:

$$\eta_{i,t} = \left(\frac{(1-\alpha)\beta(1-\iota)\sum_{\theta} w_{i,t}^{\theta} L_{i,t}^{\theta}}{p_{i,t}^{\frac{1}{\beta}} \bar{T}_i} \right)^{\beta}.$$

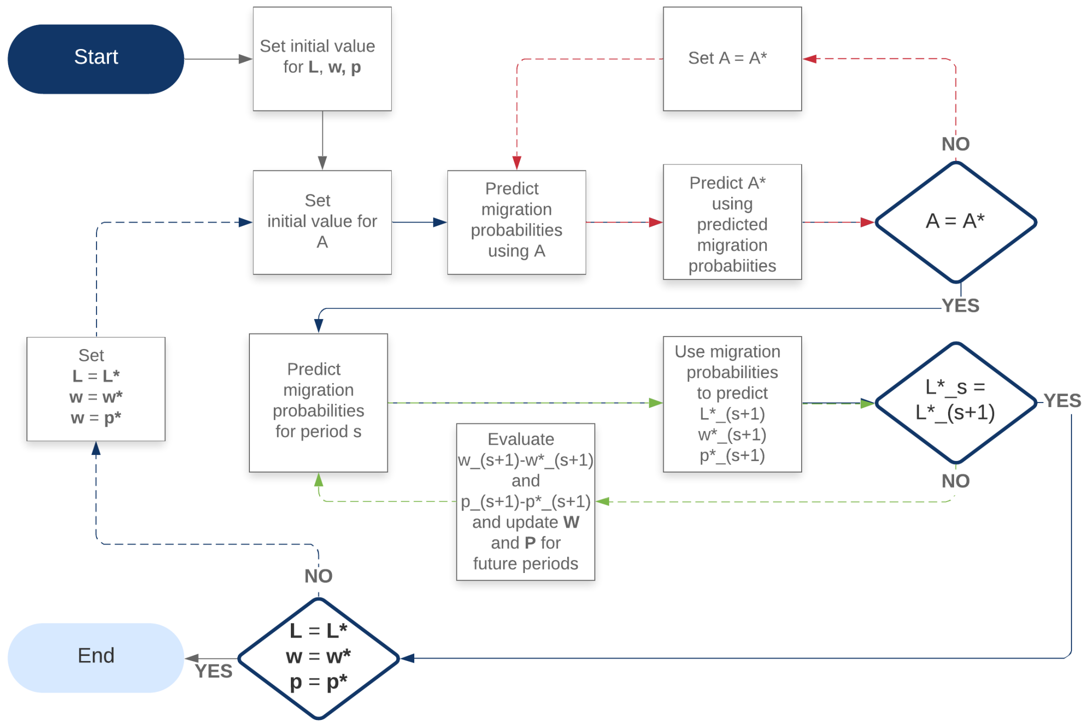
K.3.3 Quality of life ($A^{\theta}_{i,t}$)

The dynamic solver introduced in Section D.3 is a nested dynamic programming algorithm which operates according to the procedure outlined in the programming flowchart in Figure A7. Intuitively, there is an iterative fixed-point algorithm (FP) that solves for $\bar{A}^{\theta}_{i,t}$ for given guesses of $\mathbf{L^{\theta}_{i,t}}$, a dynamic programming algorithm (DP) that delivers $\mathbf{L^{\theta}_{i,t}}$ for guessed values of $\bar{A}^{\theta}_{i,t}$ and an outer loop (OL) that forwards the inputs of the former to the latter and vice versa until guesses and solutions are consistent. We introduce the three building blocks of the nested structure in more detail below.

Fixed-point programming algorithm (FP). We use a Newton algorithm to obtain numerical solutions for QoL $\bar{A}^{\theta}_{i,t}$ which we treat as an unobserved structural fundamental. The algorithm finds a numerical solution of $\bar{A}^{\theta}_{i,t}$ using Eq. (10). It uses the following inputs: Observed data on migration probabilities $\chi^{\theta}_{ij,t}$, values of the structural parameters τ^{θ}_{ij} (migration costs), γ^{θ} (migration elasticity), ι (tax rate), B^{θ}_{ij} (bilateral amenity), the employment vector $\mathbf{L^{\theta}_{i,t}}$ which for given fundamental labour productivity $\psi^{\theta}_{i,t}$ and fundamental housing productivity $\eta_{i,t}$ maps to future wages $w^{\theta}_{i,t+s}$ and rents $p_{i,t+s}$. The iterative procedure starts from uniform guesses $\bar{A}^{\theta}_{i,t}{}^{f=1} = 1$. Given the inputs, Eq. (10) delivers predicted migration probabilities $\hat{\chi}^{\theta}_{ij,t}$ and a multiplicative adjustment factor $\sum_j \hat{\chi}^{\theta}_{ij,t} / \sum_j \chi^{\theta}_{ij,t}$ which we apply to $\bar{A}^{\theta}_{i,t}{}^{f=1}$ before moving into the next iteration $f = 2$. The procedure ends when the adjustment factor approaches one. The FP consists of the processes connected by the red lines in the programming flow chart in Figure A7. Note that in QSMs with static expectations, where data are rationalised assuming a SSE, the FP algorithm alone would suffice to invert quality of life.

Dynamic programming algorithm (DP). Exploiting the dynamic structure of the model, the DP forecasts $\mathbf{L^{\theta}_{i,t}}$ using the following inputs: structural parameters $\{\alpha, \beta, \rho, \iota, \gamma^{\theta}, \zeta^{\theta}, \kappa^{\theta}, B^{\theta}_{ij,t}, \tau^{\theta}_{ij}\}$; inverted labour productivity ψ^{θ}_{ij} and housing productivity $\eta_{i,t}$; observed employment $L^{\theta}_{i,t}$ and land area $\bar{T}_i$; guessed values of $\bar{A}^{\theta}_{i,t}$ and $\mathbf{L^{\theta}_{i,t}}$, which map into vectors of guessed wages $\mathbf{w^{\theta}_{i,}}$ and rents $\mathbf{p_{i,t}}$ via the first-order condition of labour demand (Eq. (6)) and housing market clearing (Eq. (8)). The DP begins the iterative procedure in iteration $s = 0$ where it uses the above inputs to forecast migration probabilities $\chi^{\theta}_{ij|i,t+s}$ using Eq. (10). The labour supply Eq. (12) then delivers employment $L^{\theta}_{i,t+s+1}$ in the next period. $L^{\theta}_{i,t+s+1}$ maps to wages $w^{\theta}_{i,t+s+1}$ via the first-order condition of labour demand (Eq. (6)). $L^{\theta}_{i,t+s+1}$ and $w^{\theta}_{i,t+s+1}$ give regional output $X_{i,t+s+1} = \sum_{\theta} w^{\theta}_{i,t+s+1} L^{\theta}_{i,t+s+1}$ which maps into rents $p_{i,t+s+1}$ via housing market clearing (Eq. (8)). Unless the dynamic solver has converged to the dynamic equilibrium, the forecasts of $L^{\theta}_{i,t+s+1}$, $w^{\theta}_{i,t+s+1}$, and $p_{i,t+s+1}$ will not equate to the respective (s+1)-th elements in the vector of guessed employment $\mathbf{L^{\theta}_{i,t}}_{(1,s+1)}$, wages $\mathbf{w^{\theta}_{i,t}}_{(1,s+1)}$, and rents $\mathbf{p_{i,t}}_{(1,s+1)}$. Hence, we adjust wage and rent guesses concerning future periods $v > s + 1$ by the multiplicative adjustment factors $w^{\theta}_{i,t+s+1} / \mathbf{w^{\theta}_{i,t}}_{(1,s+1)}$ and $p_{i,t+s+1} / \mathbf{w_{i,t}}_{(1,s+1)}$. This way, the dynamic solver "learns" from mismatches between guessed and predicted values in every iteration of the DP in every iteration of the OL as opposed to only once per iteration of the OL. This greatly enhances the speed of the solver. Then, the iterative procedure starts over again and continues until in iteration S employment is stationary ($L^{\theta}_{i,t+s} = L^{\theta}_{i,t+s+1}$). The DP consists of the processes connected by the green lines in the programming flow chart in Figure A7.

Figure A7: Dynamic solver



Notes: Programming flowchart that illustrates the procedure of the dynamic solver introduced in Section D.3. Blue lines outline the outer loop. Red lines mark the nested fixed-point algorithm solving for $A^{\theta}_{i,t}$. Green lines mark the nested dynamic programming algorithm that forecasts $\mathbf{L}^{\theta}_{\mathbf{i,t}}$. Bold letters are $(J \times \Theta) \times H$ matrices of for $H = 1{,}000$ periods into the future. Other letters are $(J \times \Theta) \times 1$ vectors for one period. Letters with *-superscripts indicate solved outputs. Other letters indicate guessed inputs. To ease the presentation we omit all indices $\{\theta, i, t\}$ in the flow chart.

Outer loop (OL). The OL indicated by the blue lines in the flow chart in Figure A7 nests the FP and DP algorithms. It feeds the output of the FP ($\bar{A}^{\theta}_{i,t}$) as input into the DP and the output of the DP ($\mathbf{L}^{\theta}_{\mathbf{i,t}}$) as input into the FP. Intuitively, the OL treats the solutions for $\bar{A}^{\theta}_{i,t}$ and $\mathbf{L}^{\theta}_{\mathbf{i,t}}$ as a fixed point that is found in an iterative procedure when the guessed input into the FP is identical to the output of the DP and vice versa.

Before the dynamic solver enters the OL, the first step is to define initial values for $\{\mathbf{L}^{\theta}_{\mathbf{i,t}}, \mathbf{w}^{\theta}_{\mathbf{i,t}}, \mathbf{p}_{\mathbf{i,t}}\}$ which are critical inputs for the FP. This is the first process in Figure A7 after "Start". Since we do not know a priori the number of years S over which the spatial economy converges to a SSE, we begin with a long time horizon of $H = 1{,}000$ years over which agents form their expectations. Note that H exceeds S for all applications of the solver we report in this paper. As initial guesses for the employment vector $\mathbf{L}^{\theta}_{\mathbf{i,t}}{}^{\mathbf{0}}$ we use the values we observe in year t for which the model is being quantified:

$$\mathbf{L}_{\mathbf{i},\mathbf{t}}^{\theta}{}^{\mathbf{0}} = \underbrace{L_{i,t}^{\theta}, L_{i,t}^{\theta}, ..., L_{i,t}^{\theta}}_{H elements}.$$

Given the first-order condition of labour demand (Eq. (6)) and housing market clearing (Eq. (8)), $\mathbf{L}_{\mathbf{i},\mathbf{t}}^{\theta}{}^{0}$ maps directly to $\{\mathbf{w}_{\mathbf{i},\mathbf{t}}^{\theta}{}^{0}, \mathbf{p}_{\mathbf{i},\mathbf{t}}{}^{0}\}$ for given parameters and fundamental labour and housing productivity.

With these inputs, the first iteration $l = 1$ of the OL begins. The next processes until the first decision rule ($A = A^*$), including the feedback loop marked by red lines, constitute the FP algorithm. Once the decision rule is satisfied, the OL forwards the solutions for $\bar{A}_{i,t}^{\theta}{}^{l=1}$ to the DP which is represented by the processes up to the next decision rule ($L_{i,t+s}^{\theta} = L_{i,t+s+1}^{\theta}$), including the green loop. Once this decision rule is satisfied, the OL evaluates whether the values $\{\mathbf{w}_{\mathbf{i},\mathbf{t}}^{\theta}{}^{l=1}, \mathbf{p}_{\mathbf{i},\mathbf{t}}{}^{l=1}\}$ solved by the DP correspond to the guessed inputs into the FP. Until this criterion is satisfied, the OL updates the guesses and the procedure starts over gain.

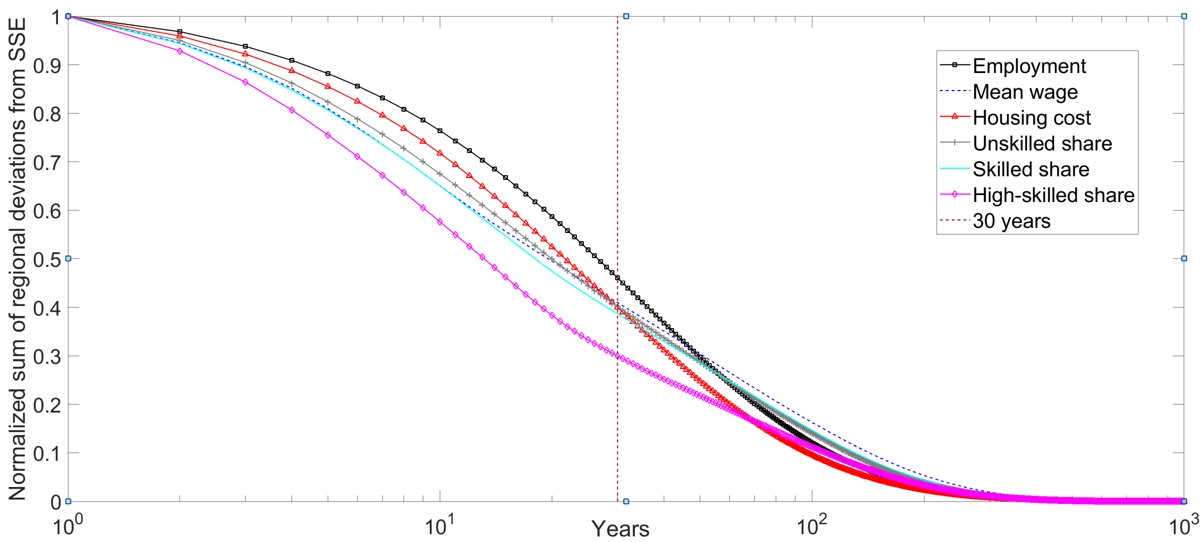
Once the OL converges in iteration $\mathcal{L}$, we crop $\mathbf{L}_{\mathbf{i},\mathbf{t}}^{\theta}{}^{\mathcal{L}}, \mathbf{w}_{\mathbf{i},\mathbf{t}}^{\theta}{}^{\mathcal{L}}, \mathbf{p}_{\mathbf{i},\mathbf{t}}{}^{\mathcal{L}}$ to $S^{\mathcal{L}}$ elements delivered by the DP in the last iteration of the OL. $\bar{A}_{i,t}^{\theta}{}^{\mathcal{L}}$ represent the solution to unobserved QoL. Hence, the model is fully quantified.

K.4 Transition into the stationary spatial equilibrium

This section complements Section D.4 in the main paper.

Figure A8 summarises how the spatial economy converges from the TSE to the SSE using the sum of absolute deviations between TSE and SSE values across region-groups as a benchmark. Depending on the outcome, about 55%-70% of the spatial convergence occurs after 30 years.

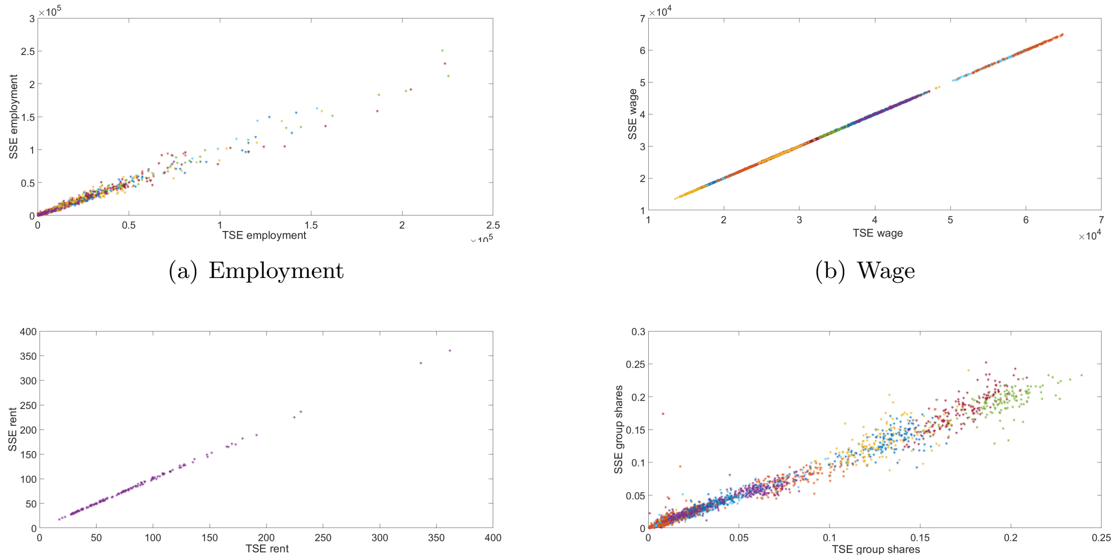
Figure A8: Spatial convergence



Notes: All trends show sum of absolute deviations from SSE values in an outcome across group-regions. 2017 starting values. Model-based forecasts.

Figure A9 scatters the SSE values in selected outcomes against the TSE values. Wages and rents are relatively closely aligned. While there are subtle differences in employment and skill shares, the correlations are still strong.

Figure A9: SSE vs TSE values



(a) Employment

(b) Wage

(c) Rent

(d) High-skilled share

Note: Unit of observation is region-group in panels a), b), d) and regions in c). Ratio of model-based forecasts (SSE) over observed data that are perfectly rationalized by the model (TSE).

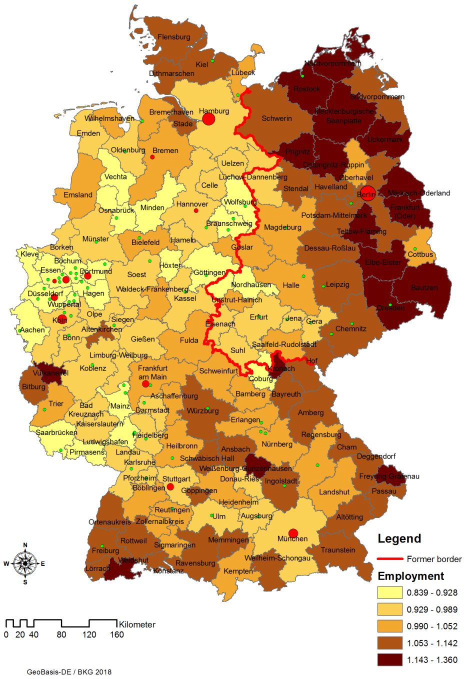
Figure A10 maps the ratios of SSE values over TSE values in selected outcomes at the regional level. As the economy converges to the SSE, the eastern states gain population at the expense of the western states. As the population increases, congestion on housing markets leads to rising rents. In contrast, there is no obvious spatial pattern in the change in skill composition and wages.

K.5 Overidentification

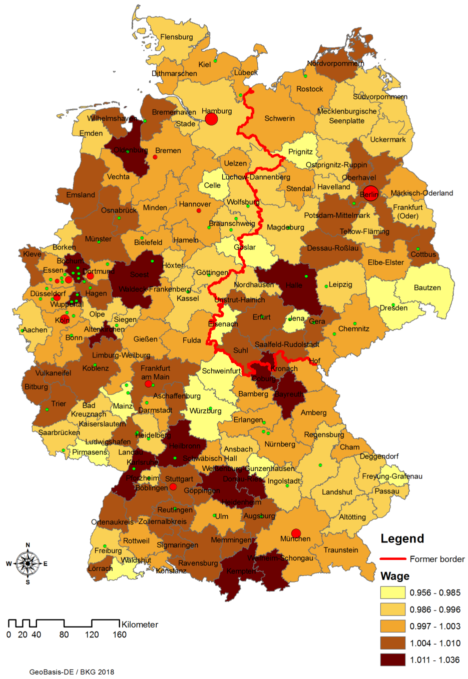
This section complements Section D.5 in the main paper. We correlate some of the structural parameters and structural fundamentals obtained from the model quantification with observable characteristics not used in the quantification of the model. Previewing our results, we find that observable characteristics correlate with model-derived fundamental labour productivity, fundamental housing productivity, and migration costs in an intuitively plausible manner. Moreover, the model forecasts changes in employment over time for the transition from the TSE to the SSE that are closely correlated with employment changes observed in data.

Labour productivity. In Table A10, Column (1), we regress fundamental labour productivity $\psi^{\theta}_{i,t=2017}$, inverted as described in Section K.3.1, against a set of dummy variables denoting some German supra-regions. We control for group fixed effects to net out composition effects.

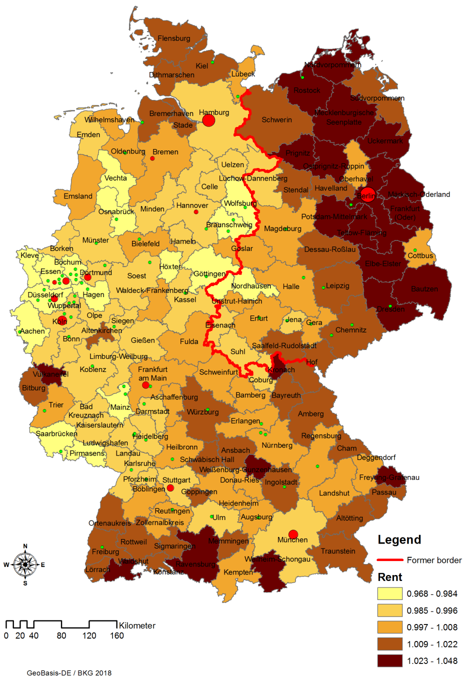
Figure A10: Ratio of SSE over TSE values



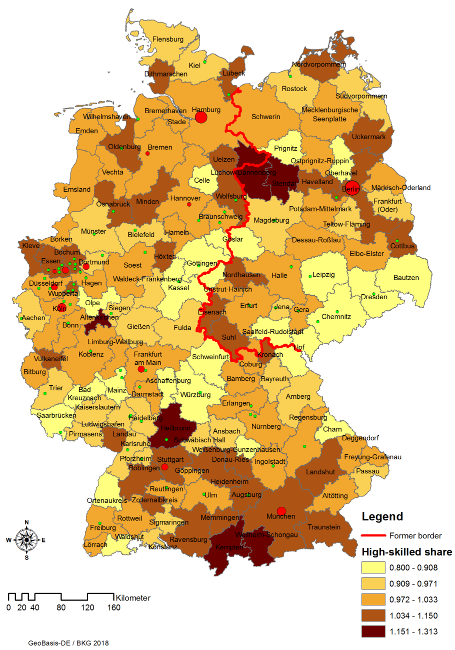
(a) Employment



(b) Wage



(c) Rent



(d) High-skilled share

Note: Unit of observation is 141 labour market areas as defined by Kosfeld and Werner (2012). Ratio of model-based forecasts (SSE) over observed data that are perfectly rationalized by the model (TSE).

We find that fundamental labour productivity is about 7% smaller in the eastern states, likely a legacy of the Cold War era. Fundamental productivity is somewhat higher, on average, near the Alps. In keeping with intuition, a casual inspection of fundamental productivity across regions reveals a greater productivity at peripheral regions where the local economies are dominated by global companies such as Volkswagen in Wolfsburg (see Figure A11). Adding industry sector shares in Column (2) reveals that part of the east-west gap is attributable to industry composition. In keeping with intuition, regions with a high share of tradable services tend to be more productive.

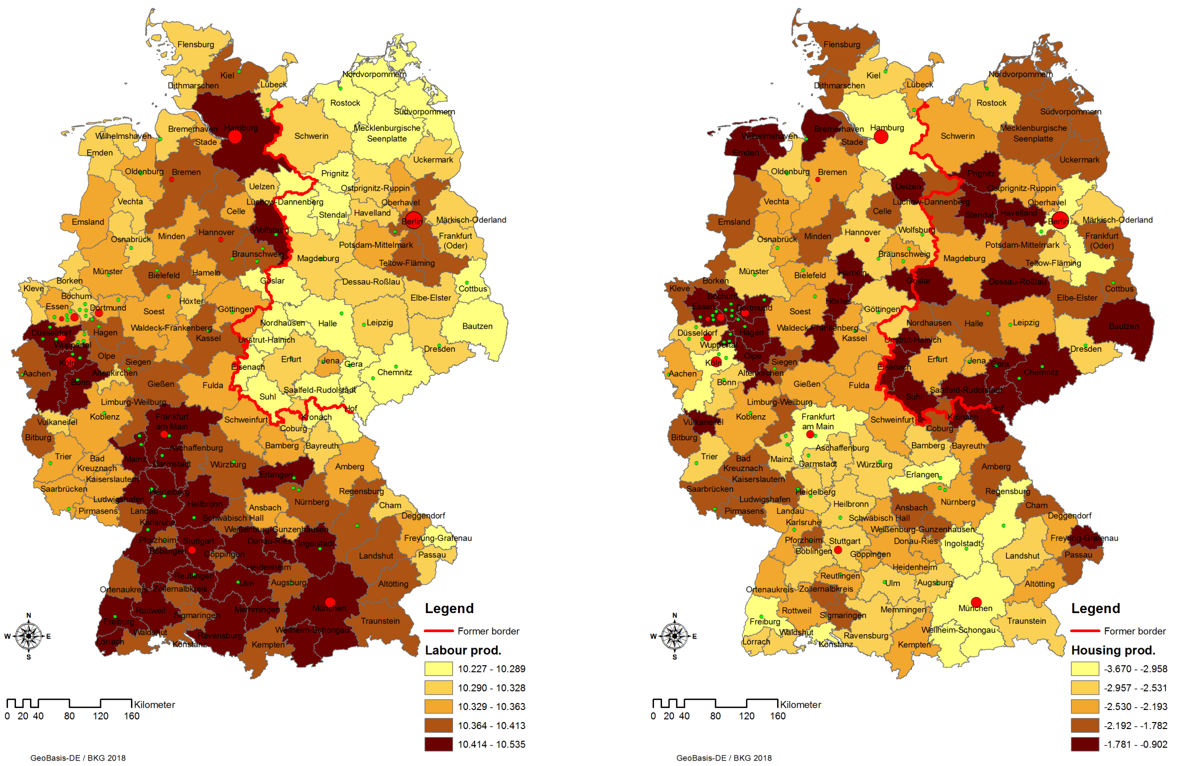
Table A10: Fundamental productivity

	(1) Labour productivity	(2) Labour productivity	(3) Housing productivity	(4) Housing productivity
	$\psi^{\theta}_{i,t=2017}$	$\psi^{\theta}_{i,t=2017}$	$\eta_{i,t=2017}$	$\eta_{i,t=2017}$
East (0,1)	-0.067***	-0.039***	0.263***	-0.072
	(0.01)	(0.01)	(0.09)	(0.10)
Alps (0,1)	0.040***	0.051***	-0.517***	-0.639***
	(0.01)	(0.02)	(0.14)	(0.19)
Coast (0,1)	-0.021***	-0.002	0.007	-0.065
	(0.01)	(0.01)	(0.15)	(0.12)
Agricultural share (%)		0.002		0.019
		(0.00)		(0.04)
Construction (%)		-0.001		-0.029
		(0.00)		(0.03)
Tradable services (%)		0.005***		-0.083***
		(0.00)		(0.01)
Manufacturing (%)		0.002***		-0.007
		(0.00)		(0.01)
Constant	10.316***	10.220***	-2.278***	-1.045***
	(0.00)	(0.02)	(0.05)	(0.39)
Group effects	Yes	Yes	-	-
Observations	2,538	2,538	141	141
R^2	.982	.985	.0794	.413

Notes: Unit of observation is region-groups in (1) and regions in (2). (1) indicates binary indicator variables. Standard errors in parentheses.

Housing productivity. In Table A10, Column (3), we regress fundamental housing productivity $\eta_{i,t=2017}$, inverted as described in Section K.3.2, against a set of dummy variables denoting the German supra-regions. We find that housing productivity is significantly higher in the eastern states. This is a plausible finding given the country's history. During the division period, former East Germany was governed by a socialist planning regime with an emphasis on the provision of affordable housing. The relatively large quantities of housing provided came at the expense of poor housing quality. Following Germany's unification, favourable tax reliefs to real estate investors led to a construction boom and a rejuvenation of the housing stock (Flockton, 1998). Hence, it is plausible that as of 2017, there is a greater supply of housing services for given levels of geographic land area and demand. Likewise, it is plausible, that there is a negative housing productivity effect near the Alps as mountainous areas are more difficult to develop. Adding industry shares in Column (4) reveals a negative correlation between tradable services and housing productivity. One interpretation that would be in line with anecdotal international evidence is that places with high labour productivity tend to develop restrictive planning systems to protect amenities that are valued by the high-skilled (as, for example, in some Californian cities). Since tradable services are concentrated in cities in the western states (e.g. Frankfurt, Munich, Dusseldorf), the east-west gap is reduced close to zero conditional on the industry controls.

Figure A11: Fundamental labour and housing productivity



(a) Labour productivity $\psi_{i,t=2017}$

(b) Housing productivity $\eta_{i,t=2017}$

Note: Unit of observation is 141 labour market areas as defined by Kosfeld and Werner (2012). Values inverted as described in Sections K.3.1 and K.3.2. Group-region productivities aggregated to regions using TSE sector shares.

Migration costs. In Table A11, we correlate our parameter estimates capturing migration costs with measures of migration distance which, intuitively, should be positively correlated. Indeed we find that migration resistance increases in distance at an elasticity of 1.74 (Column 1). With a negative sign, this parameter corresponds to the distance elasticity of migration flows that is frequently estimated by reduced-form gravity models. Tombe and Zhu (2019) and Imbert and Papp (2019) estimate similar elasticities for China and India. Bryan and Morten (2019) report a distance elasticity of migration of 0.7 for Indonesia. In our model, migration costs are monitored by the origin-destination-group-specific iceberg migration cost parameter τ_{ij}^{θ}. This parameter increases in distance at an elasticity of 0.245 (Column 3).

Table A11: Migration costs vs. geographic and social distance

	Migration resistance	Migration resistance	Migration iceberg cost	Migration iceberg cost
	$\tau_{ij}^{\theta} \times \gamma^{\theta}$	$\tau_{ij}^{\theta} \times \gamma^{\theta}$	τ_{ij}^{θ}	τ_{ij}^{θ}
Ln geographic distance	1.740***	0.815***	0.245***	0.148***
	(0.01)	(0.03)	(0.00)	(0.00)
Ln social distance		0.647***		0.068***
		(0.02)		(0.00)
Group effects	Yes	Yes	Yes	Yes
IV	-	Yes	-	Yes
Observations	355,320	355,320	355,320	355,320
R^2	.808	.904	.933	.86

Notes: Unit of observation is origin-destination-group. Ln social distance is the inverse of the log of the Facebook social connectedness index (Bailey et al., 2018). Log historic dialect (Falck et al., 2012) similarity is used as an instrumental variable for log social distance where indicated. Standard errors in parentheses. Regressions weighted by O-D-group flows.

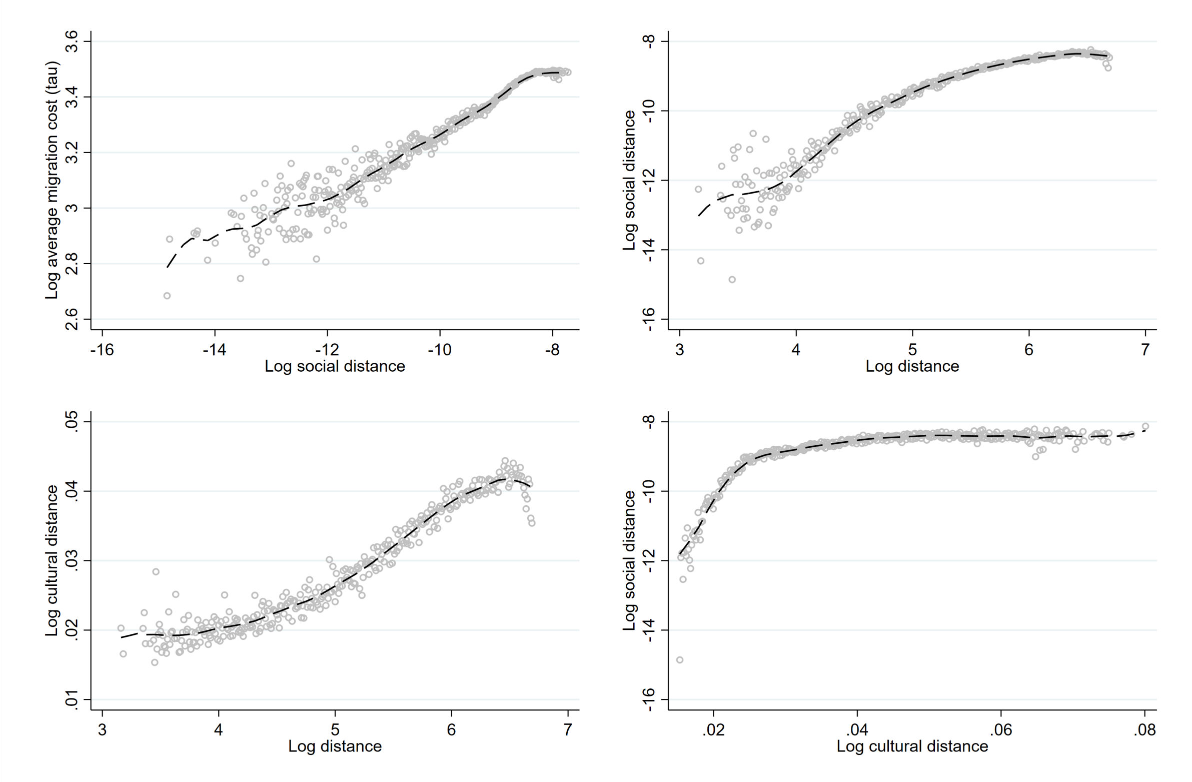
To shed some light on the mechanisms through which the geographic distance effect operates, we utilise a social distance measure defined as the inverse of an index that summarises how connected Facebook users in two regions are (see Appendix K.1.7 for further details on the social connectedness index). Figure A12 shows an approximately log-linear relationship between social distance and our estimated migration cost parameters τ_{ij}^{θ}, suggesting that social ties may reduce the cost of rebuilding social capital at a migration destination. As expected, social distance is positively correlated with geographic distance, a well-known feature of social networks (Bailey et al., 2018). The geographic proximity effect is also visible in a measure of cultural distance which is the inverse of historic dialect similarity (Falck et al., 2012). Hence, it is no wonder that social distance and cultural distance are also positively correlated, implying that regions that are closely connected today usually have had cultural ties in the past.

When adding log social distance as an additional covariate in Columns (2) and (4) in Table A11, we use log cultural distance as an instrumental variable to address reverse causality from migration cost to social connectedness. We find that migration costs increase significantly in social distance, controlling for geographic distance. Moreover, adding social distance, reduces the geographic distance effect by 53% (Column 2) and 40% (Column 4), suggesting that the cost of rebuilding social capital may be an important component of migration costs.

To summarise how the relationship between migration costs and social distance varies by group, we first regress the estimated migration cost parameter τ_{ij}^{θ} on log social distance and geographical distance by group θ. We estimate the average difference in the estimated coefficients of log social distance between gender, sex and skill groups in a second-step regression reported in Table A12. Our preferred instrumental variable results reveal that the elasticity of migration costs with respect to social distance is relatively large for middle-skilled and high-skilled workers.

Employment (out-of-sample). Our data set contains all critical variables for the inversion of the model from 2007 onward. To compare the TSE to SSE transition path forecast by the

Figure A12: Migration cost vs. bilateral distance measures



Note: Social distance is the inverse of the Facebook social connectedness index (Bailey et al., 2018). Cultural distance is the inverse of historic dialect similarity (Falck et al., 2012). Geographic distance is the great-circle distance. Observations are grouped into 10^{-6} on the x-axis. All distance and cost measures computed for pairs of German local labour markets.

dynamic solver to data, we invert the model from a 2007 TSE and regress the model-based employment forecast on values observed in the data in Table A13. This is a demanding out-of-sample over-identification test as we expect all fundamentals to be affected by exogenous shocks, hence the within-region correlation over time is necessarily noisy. We expect a positive correlation to the extent that these shocks are orthogonal to the TSE deviations from the SSE since the model can predict mean reversion and the causal effects of known exogenous shocks, but not the occurrence of future events.

Yet, the within-region elasticity of forecast employment with respect to observed employment is precisely estimated at 0.775 (t-stat $>$ 25). Weighting by employment, the estimated elasticity increases to 0.852 (t-stat $>$ 25). Hence, the model successfully captures a mean reversion tendency that is a feature of the data, in particular for the larger labour markets. Consistent with a less favourable signal-to-noise ratio the correlation is weaker at the group-region level where cell sizes are much smaller. Nevertheless, if we weight by the size of the region-group cells the elasticity, at 0.493 is still positive and precisely estimated (t-stat $>$ 45).

Table A12: Migration cost against social distance (by group)

	OLS	2SLS
Female	0.004	0.004
	(0.00)	(0.01)
31-50 years	-0.015***	-0.000
	(0.00)	(0.01)
51-65 years	-0.036***	-0.003
	(0.01)	(0.00)
Apprenticeship	0.071***	0.055***
	(0.00)	(0.01)
Tertiary education	0.051***	0.049***
	(0.00)	(0.01)
Constant	0.084***	0.023**
	(0.01)	(0.01)
Observations	18	18
R^2	.971	.885

Notes: The units of observation are labour market region pairs. The dependent variable is the estimated coefficient of log social distance from separate regressions of the estimated bilateral migration costs τ_{ij}^{θ} on log geographical and log social distance for each θ-group. In the 2SLS specification, cultural distance is used as an instrumental variable for social distance. Robust standard errors are shown in parentheses.

Table A13: Employment: Model-based forecast vs. data

	Ln employment (2007-2017 in data)	Ln employment (2007-2017 in data)	Ln employment (2007-2017 in data)	Ln employment (2007-2017 in data)
	$L_{i,t}^{\theta}$	$L_{i,t}^{\theta}$	$L_{i,t}^{\theta}$	$L_{i,t}^{\theta}$
Ln employment (2007-2017, forecast from 2007 TSE)	0.775***	0.852***	0.091***	0.493***
	(0.03)	(0.03)	(0.01)	(0.01)
Unit	Region-year	Region-year	Region-group-year	Region-group-year
Time effects	Year	Year	Year-group	Year-group
Location effects	Region	Region	Region-group	Region-group
Weights	-	Region emp.	-	Region-group emp.
Observations	1551	1551	27918	27918
R^2	1	1	.998	.999

Notes: Employment forecast from a 2007 TSE using the dynamic solver. Year and year-group effects capture the effects in changes in the size of the workforce that occur in the data but not in the model-based forecast. Standard errors in parentheses.

L Measuring quality of life

This section complements Section E in the main paper.

L.1 Spatial variation in quality of life

Table A14 presents estimates of the elasticity of RR-QoL with respect to DSM-QoL by worker group that correspond to the log-linear slopes plotted in Figure 5. On average log-point increase in DSM-QoL is associated with a 0.27-log-point increase in RR-QoL. The estimated log slope tends to be somewhat larger for male, old, and skilled workers, but remains below 0.5 for all groups. On average, the DSM-QoL explains almost 60% of the variation in the RR-QoL, with some variation across groups.

Table A14: Elasticity of Rosen-Roback QoL with respect to dynamic model QoL

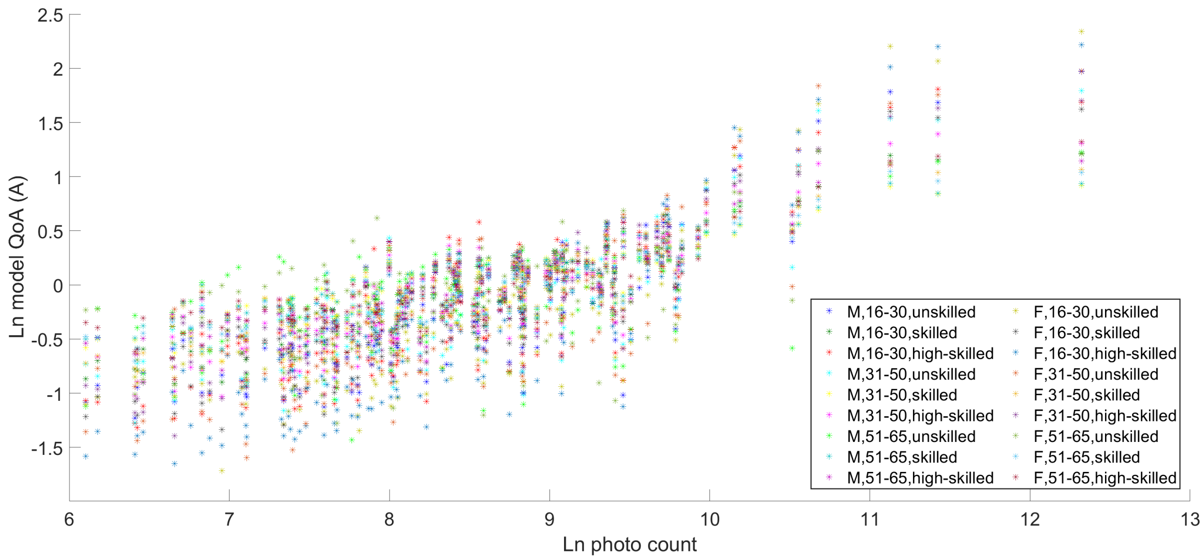
Gender	Age	Unskilled	Skilled	High-skilled	Mean
Male	16-30 years	0.21***	0.32***	0.22***	0.25
Male	31-50 years	0.22***	0.44***	0.27***	0.31
Male	51-65 years	0.29***	0.45***	0.37***	0.37
Female	16-30 years	0.16***	0.20***	0.17***	0.18
Female	31-50 years	0.16***	0.31***	0.21***	0.23
Female	51-65 years	0.19***	0.38***	0.37***	0.32
	Mean	0.21	0.35	0.27	0.27

Notes: Point estimates from group-specific region-level regressions of ln RR-QoL ($\ln \mathcal{A}_i^\theta$) against ln DSM QoL ($\ln \bar{A}_i^\theta$). All estimates are significant at the 1% level. The last column and row present unweighted row and column means. QoL inverted from 2017 data.

L.2 Determinants of quality of life

Figure A13 shows the correlation between DSM-QoL and the big data amenity index introduced in Appendix D.1 by group. Consistent with the pooled regression results in Table 2, there is a positive correlation between both measures. Moreover, the correlation is similarly well defined across groups. Hence, Figure A13 substantiates the notion that social-media-based big data may serve as a proxy for QoL.

Figure A13: Quality of life (DSM) vs. big data (photos) amenity



Notes: 2017 values. Unit of observation is region-group. Model-based amenity inverts QoL from a TSE assuming that agents have perfect foresight. Big data amenity is the number of geo-tagged photos shared in social media (Flickr and Picasa).

M Policy evaluation

This section complements Section F in the main paper.

M.1 Social welfare

In this section, we derive the social welfare measure introduced in Section F.1. We start with a social welfare function that allows for inequality aversion in a general form. Following Atkinson (1970), we assume

$$\mathcal{W} = \frac{1}{1-\varepsilon} \sum_i \sum_\theta \left(\mathcal{R}^\theta_{i,t}\right)^{1-\varepsilon} \frac{L^\theta_i}{\bar{L}} \tag{39}$$

for both the baseline ($*$) and the counterfactual (c) spatial equilibrium. The degree of inequality aversion is measured by $0 \leq \varepsilon \neq 1$.[13]

It is instructive to transform Eq. (39) into a scale-dependent part $\mathcal{R}$ and a scale-independent part that penalises for inequality $1-\mathcal{I}$. The former is simply the weighted average of location-group utility that for the baseline and the counterfactual is respectively given by:

$$\mathcal{R}^* = \sum_i \sum_\theta \mathcal{R}^{\theta}_{i|i}{}^{*} \frac{L^{\theta *}_i}{\bar{L}} \tag{40}$$

$$\mathcal{R}^c = \sum_i \sum_\theta \widehat{\mathcal{R}^\theta_i} \mathcal{R}^{\theta}_{i|i}{}^{*} \frac{L^{\theta *}_i}{\bar{L}}. \tag{41}$$

Using the "exact hat algebra" approach by Dekle et al. (2007), we express group-region utility in the counterfactual measured at the migration origin as $\widehat{\mathcal{R}^\theta_i} \mathcal{R}^{\theta}_{i|i}{}^{*}$. This way, we account for changes in expected utility and migration costs which enter into $\widehat{\mathcal{R}^\theta_i}$.

To derive the inequality measure $\mathcal{I}$, we search for the *equally distributed equivalent utility* $\mathcal{U}_{EDE}$ (a hypothetical average level of expected lifetime utility across individuals) that leads to the same level of welfare as with the actual distribution of expected lifetime utilities. Eq. (39) implies that

$$\mathcal{W}(\mathcal{R}_{EDE}) = \frac{1}{1-\varepsilon} (\mathcal{R}_{EDE})^{1-\varepsilon}, \tag{42}$$

such that we can solve for $\mathcal{R}_{EDE}$ by equalising Eqs. (39) and (42). This yields

$$\mathcal{R}_{EDE} = \left[\sum_i \sum_\theta \left(\mathcal{R}^\theta_{i|i}\right)^{1-\varepsilon} \frac{L^\theta_i}{\bar{L}} \right]^{\frac{1}{1-\varepsilon}}.$$

Using Atkinson's inequality measure

$$\mathcal{I} = 1 - \frac{\mathcal{R}_{EDE}}{\mathcal{R}} \qquad \in [0,1], \tag{43}$$

[13] We obtain log-utility as a special case for $\epsilon = 1$.

we obtain

$$\mathcal{I}^* = 1 - \Big[\sum_i \sum_\theta \Big(\frac{\mathcal{R}^{\theta}_{i|i}{}^*}{\mathcal{R}^*}\Big)^{(1-\varepsilon)} \frac{L^{\theta *}_i}{\bar{L}}\Big]^{\frac{1}{1-\varepsilon}} \tag{44}$$

$$\mathcal{I}^c = 1 - \Big[\sum_i \sum_\theta \Big(\frac{\widehat{\mathcal{R}^{\theta}_i}\mathcal{R}^{\theta}_{i|i}{}^*}{\mathcal{R}^c}\Big)^{(1-\varepsilon)} \frac{L^{\theta *}_i}{\bar{L}}\Big]^{\frac{1}{1-\varepsilon}} \tag{45}$$

for both the baseline and the counterfactual case, respectively. These derivations allow us to reformulate Eq. (39) as $\mathcal{W} = \mathcal{R}\,(1-\mathcal{I})$ and express changes in social welfare according to Eq. (18).

M.2 Instrumental variable estimates of air pollution effects

This section complements Section F.2 in the main paper. We discuss our wind-adjusted coal exposure instrumental variables in greater detail and provide a discussion of the relevance and the validity of the instrumental variables as well as the underlying mechanisms.

Wind-adjusted coal exposure. To generate exogenous variation in pollution levels, we follow Deryugina et al. (2019) and Heblich et al. (2020) and exploit that the diffusion of air pollution is shaped by winds and that, historically, coal deposits attracted high-polluting industries (for example steal mills) and power plants. We define wind-induced coal exposure E for region i as follows:

$$E_i^{\mathcal{E}} = \frac{\sum_{i\neq j} \frac{CC_j^{\mathcal{E}}}{WD_{ij}}}{\sum_{i\neq j} \frac{CC_j^{\mathcal{E}}}{D_{ij}}}, \tag{46}$$

where $\mathcal{E} = \{\text{black coal, brown coal}\}$, CC_j is the percentage of the geographic area of region j with coal deposits, D_{ij} is the crow-flight distance between region i and region j and WD_{ij} is the wind-adjusted distance defined as follows:

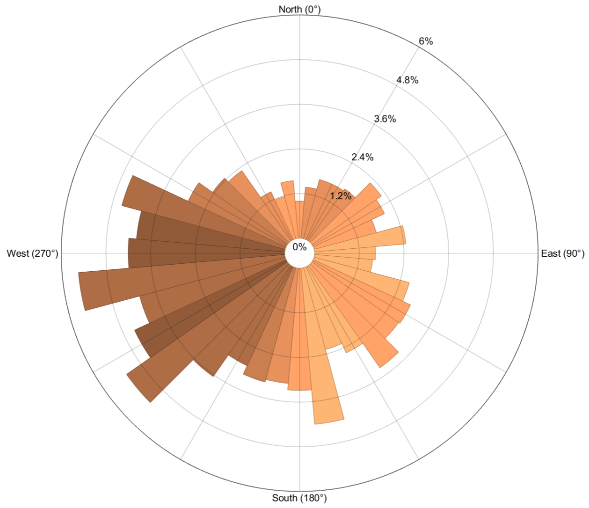
$$WD_{ij} = \frac{D_{ij}}{\frac{w_{i,r(ij)}}{\frac{1}{R}\sum_{s=r}^{R} w_{s,i(ij)}}},$$

where $w_{r,i(ij)} = \frac{W_{r,i(ij)}}{\sum_{s=r}^{R} W_{s,i(ij)}}$ and $W_{r,i(ij)}$ is the frequency of winds blowing from direction $r \in R$.

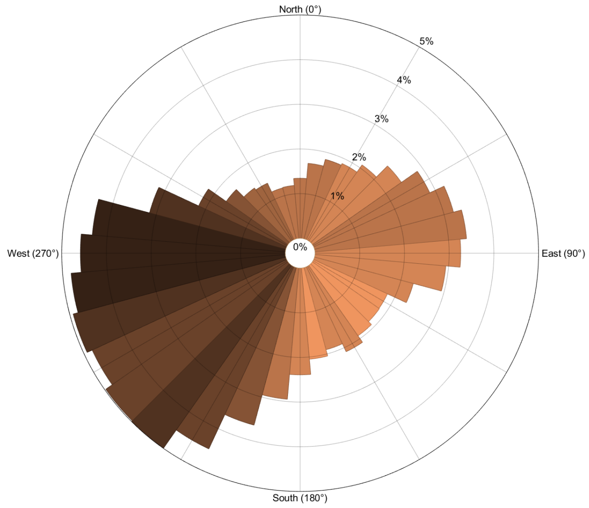
The denominator in Eq. (46) is a *geographical* exposure measure that aggregates CC_i across surrounding regions, weighted by distance. This formulation is closely related to the *market potential* by Harris (1954), which has become a workhorse tool in economic geography, international trade, and urban economics. The fact that we exclude the "self-potential" (for region $i = j$) makes our exposure measure similar to spatial lags used in geographic data science where spatial auto-correlation is viewed as a typical manifestation of the *First Law of Geography* (Tobler, 1970). The numerator in Eq. (46) is a *meteorological* exposure measure constructed in exactly the same way as the denominator, except that the spatial weights incorporate wind patterns. Intuitively, we scale down geographic distance ($WD_{ij} < D_{ij}$) if winds typically blow from j to i. Likewise, we scale up geographic distance ($WD_{ij} > D_{ij}$) if winds typically blow from i to j. Through the normalisation by the conventional spatial lag, we net out the effects of outcomes that are correlated with CC and auto-correlated in space. Since we exclude region i in the exposure measure $E_i^{\mathcal{E}}$, we also exclude any unobserved variables that determine the QoL and pollution production within the same region. As a result, our exposure measure identifies the air pollution effect from wind-induced variation, exclusively.

We obtain the frequency distribution of winds by direction $r \in R$ for region $i \in J$ from Kasperski (2002). In these data, r is defined in terms of $R = 36$ 10-degree intervals where $r = 0$ if region j is exactly north of region i. Figure A14 illustrates the frequency distribution for the four largest German cities using wind rose diagrams. With this information, it is a matter of simple 2D geometry to compute a radian angle for an ij-route as $\mathit{atan2}(y_j - y_i, x_j - x_i)$ (x and y are coordinates in a projected system) and map it to the wind rose via a standard radian-to-degree conversion.

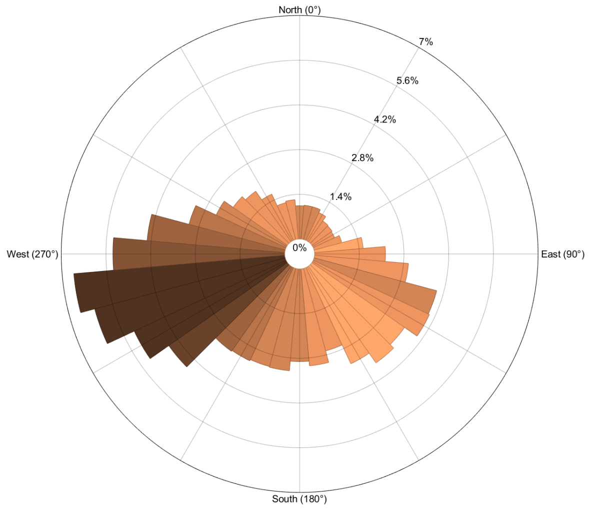
Figure A14: Wind rose diagrams



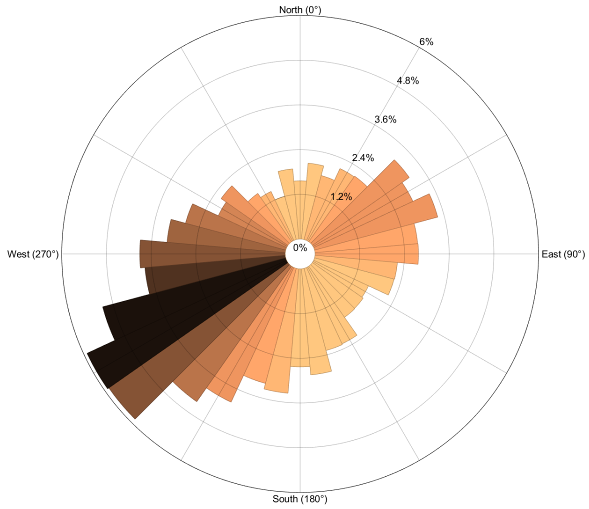
(a) Hamburg



(b) Berlin



(c) Cologne

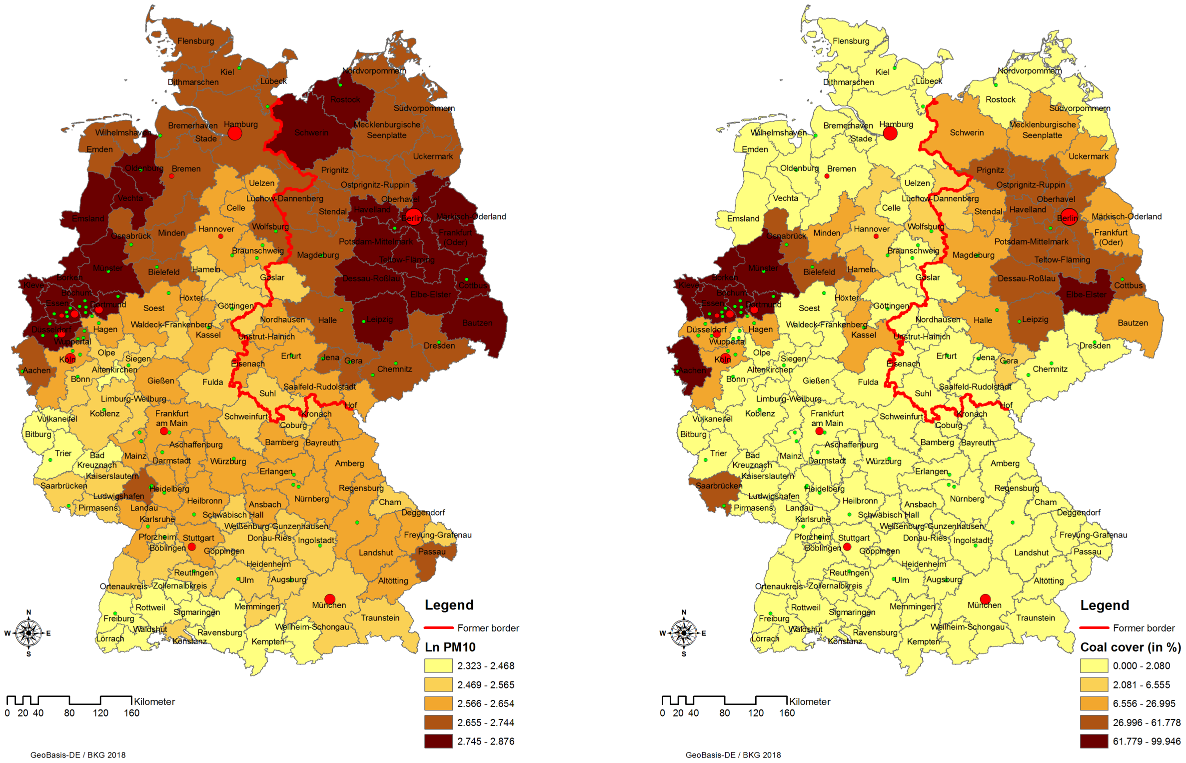


(d) Munich

Note: Own illustration of data provided by Kasperski (2002). Sizes of slices are proportionate to wind frequency. Darker colours indicate stronger winds. Geographic directions (in a polar coordinate system) refer to the directions *from* where winds blow.

Relevance. Panel (a) of Figure A15 illustrates the regional differences in the concentration of particulate matter in Germany. With the exception of a few regions along the former inner-German border, pollution levels are generally higher in East Germany. Within West Germany, higher concentration levels are recorded around the Ruhr Valley in North-Rhine Westphalia as well as in parts of the North. Except for *Ludwigshafen*, where chemical industry is located, and *Passau*, pollution levels are considerably lower in South Germany.

Figure A15: Coal cover vs. pollution

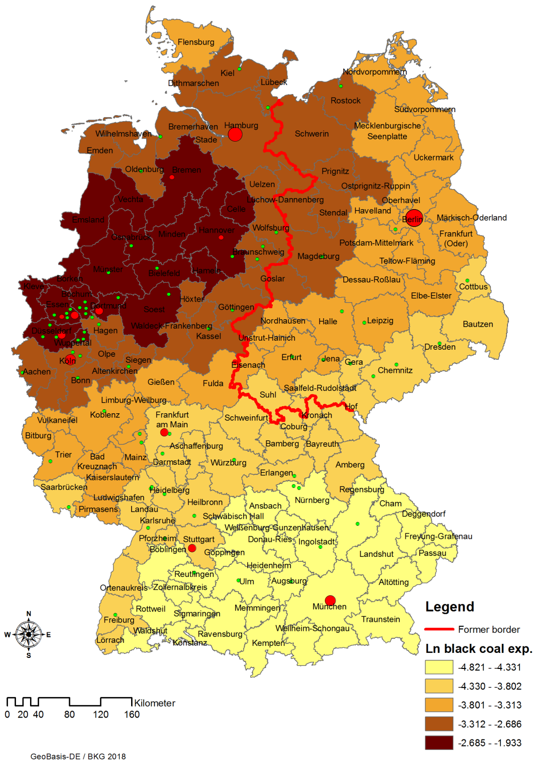


(a) PM^{10} concentration

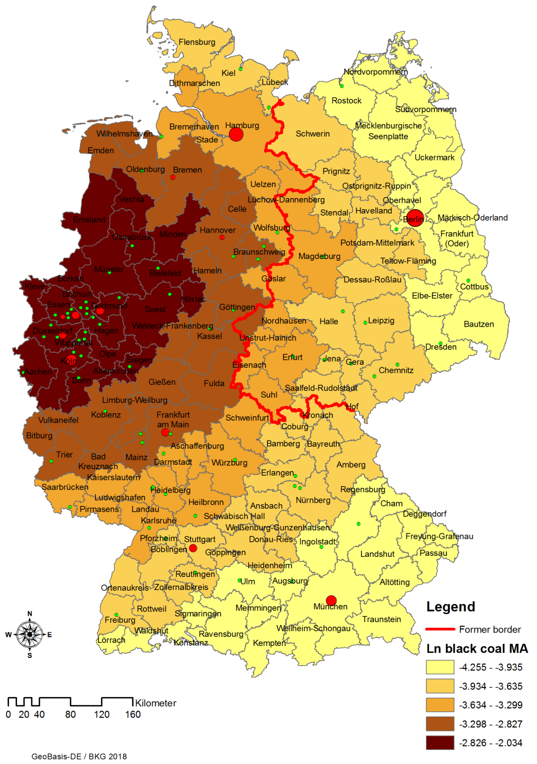
(b) Coal deposits (back and brown)

Note: Unit of observation are 141 labour market regions defined by Kosfeld and Werner (2012). Coal exposure is the wind-adjusted-distance-weighted aggregated of coal deposits in surrounding regions $j \neq i$.

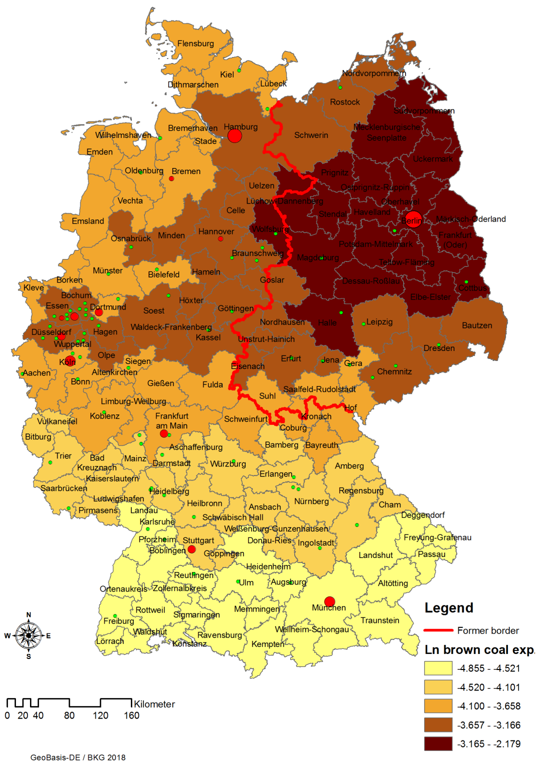
–Figure A15 – continued from previous page



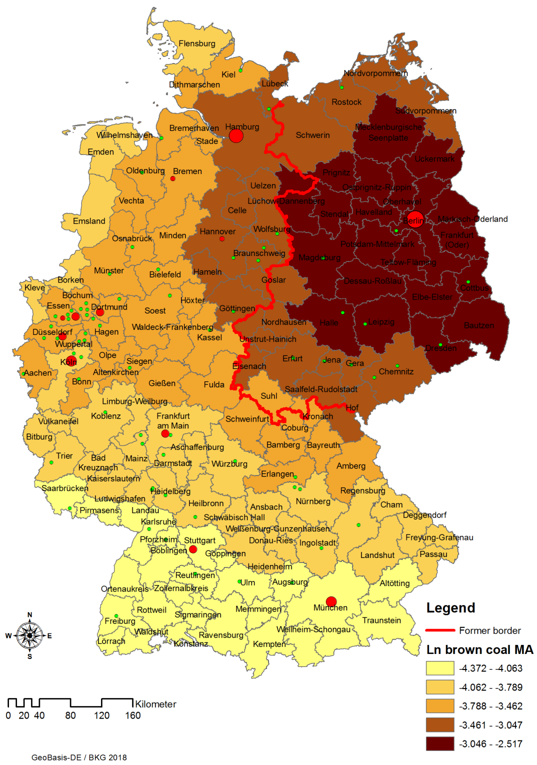
(c) Meteorological black coal exposure



(d) Geographical black coal exposure



(e) Meteorological brown coal exposure



(f) Geographical brown coal exposure

Note: Unit of observation are 141 labour market regions defined by Kosfeld and Werner (2012). Meteorological coal exposure is the numerator in Eq. (46). Geographical coal exposure is the denominator in Eq. (46). Intuitively, both exposure measures are distance-weighted aggregated coal deposits in surrounding regions $j \neq i$ using inverse distance weights. For the meteorological exposure measures, geographic distances are adjusted for wind directions .

These differences in the concentration of particulate matter bear a close resemblance to the spatial distribution of brown and black coal fields which is shown in panel (b). Areas in which large-scale extraction of coal has been taking place are clearly visible in North-Rhine Westphalia and the Saarland in West Germany as well as in parts of Saxony, Saxony-Anhalt and Brandenburg in East Germany.

Panel c) shows a region's log meteorological exposure to black coal which is based on the wind-adjusted distances (the numerator of Eq. (46)). The area with the highest concentration contains the Ruhr Valley as well as the regions to the North-East of the former, because winds typically blow from the South-West (as shown in Figure A14). Panel d) reflects the geographical exposure to black coal that is based on adjusted crow-flight distances (the denominator of Eq. (46)). In contrast to panel c), the iso-exposure lines are approximately concentric, with the Ruhr Valley being the nucleus of the gradient. Panels e) and f) show the same exposure measures for brown coal deposits. Intuitively, we identify from wind-induced exposure to coal deposits, exclusively, by using the log-difference between the two exposure measures depicted

Table A15: Quality-of-life determinants

	(1) $\ln PM^{10}$	(2) $\ln PM^{10}$	(3) $\bar{A}_i^{\theta}$
Ln wind-adjusted exposure	0.152***	0.105***	
to black coal	(0.02)	(0.02)	
Ln wind-adjusted exposure	0.077*	0.028	
to brown coal	(0.04)	(0.03)	
Ln pollution concentration (pm10)			-1.935***
			(0.69)
East			-0.221**
			(0.09)
Near Alps (dummy)			-0.452***
			(0.17)
Near coast (dummy)			-0.286***
			(0.10)
Ln crime per capita			0.340***
			(0.11)
Ln area			0.112
			(0.09)
Housing stock destroyed in WWII (%)			0.013***
			(0.00)
Number of Opera houses			0.178***
			(0.03)
Ln water area			0.175**
			(0.08)
First-stage F-statistic			24.532
Group-year effects	Yes	Yes	Yes
Controls	No	Yes	Yes
Observations	27,918	27,918	27,918
R^2	.508	.673	.504

Notes: Unit of observation is region-group-year. Regional pollution is instrumented using the wind-adjusted log exposure to black and brown coal as described in Section M.2 and are determined net of market access. *First-stage F-Statistic* refers to the Kleibergen-Paap rank LM statistic. Standard errors clustered on regions. $^+$ $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in panels (c) and (d) as well as in (e) and (f) as instrumental variables for air pollution.

Columns (1) and (2) of Table A15 provide a closer assessment of the relationship between the concentration of particulate matter and the coal exposure IVs. While the conditional correlation of the brown coal exposure measure depends on the specification, black coal exposure has a positive effect on air pollution levels throughout.

Mechanism. For much of the 19$^{\text{th}}$ and the 20$^{\text{th}}$ century, coal was an essential input for energy-intense industries and coal power plants which co-located with coal deposits due to transport costs that used to be much higher than they are today (Mohammed and Williamson, 2004; Fernihough and O'Rourke, 2021). Path-dependency is a well-documented feature of economic geography (Davis and Weinstein, 2002; Redding et al., 2010; Bleakley and Lin, 2012). Polluting industries and power plants are no exception and have remained in close proximity of coal deposits when transport costs fell, making them an indirect source of air pollution. The advantage of using coal deposits in the exposure measure over polluting establishments is that the former is exogenously determined by geology while the location of the latter is to some extent endogenously determined by local economic conditions.

Column (1) of Table A16 shows that there is a strong relationship between brown and black coal deposits on the one hand and the number of active coal plants on the other hand. This association extends to the geographical coal exposure measures, as shown in Column (2). In Columns (3) and (4) we use the number of employees in energy-intensive heavy industry as the dependent variable. A larger share of areas with coal deposits is associated with a larger number of worker in these sectors (conditional on a region's overall employment level) which supports the hypothesis of collocation of black coal deposits and energy-intensive industry. We find no such relationship in the case of brown coal, which is consistent with the greater predictive power of the black coal exposure measure in the first stage of the IV regressions (see Table A15).

Validity. The use of coal exposure as an instrumental variable hinges on the assumption that there are no other channels through which the former might influence QoL. Arguably, we have ruled out many of the causes for concern by excluding region $i = j$ from the exposure measures. Hence, local disamenity effects of coal power plants, for example related to unpleasant views, will not be captured by our instrumental variables. There is also the concern that the presence of heavy industry in regions with coal deposits led to intensive bombing raids during WWII. It is conceivable that the resulting destruction of the housing stock and of infrastructure led to a permanent reduction in QoL in those regions. We control for a potential war-destruction effect in our IV regressions, but even if our control was imperfect, excluding region $j = i$ in the construction of the coal exposure IVs ensures that the IVs will not capture effects of WWII destruction in region i.

However, one may argue that workers travel across regions for leisure. Hence, WWII destruction or any other legacy effect of nearby coal fields on the attractiveness of nearby regions could be captured by our coal exposure measures. This is why we normalise meteorological

Table A16: Collocation of energy-intensive heavy industries and coal plants with coal deposits

	(1) Number of active coal plants	(2) Number of active coal plants	(3) Employees in energy-intensive heavy industry	(4) Employees in energy-intensive heavy industry
Black coal cover (in %)	1.658***		0.288**	
	(0.51)		(0.14)	
Brown coal cover (in %)	3.226***		-1.133**	
	(0.89)		(0.53)	
Ln black coal exposure		1.063***		0.390***
		(0.33)		(0.10)
Ln brown coal exposure		0.869**		-0.336***
		(0.43)		(0.13)
Ln employment			0.789***	0.740***
			(0.06)	(0.04)
Constant	-1.107***	5.966***	-0.369	0.279
	(0.26)	(1.84)	(0.76)	(0.89)
Observations	141	141	141	141
Pseudo R^2	.099	.089	.650	.717

Notes: Unit of observation is region. Poisson estimation. Sector shares are measured in 2017. Robust standard errors. $^{+}$ $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

exposure by geographic exposure in Eq. (46). We argue that all spatial spillover effects that operate independent of wind directions will be net out by geographic exposure in the denominator. Hence, the instrument defined in Eq. (46) provides identifying variation stemming purely from wind-induced patterns in air pollution which we argue to be exogenous.

M.3 Other applications

This section provides a more detailed discussion of the results summarised in Section F.3. We present several counterfactual exercises, which are all motivated by the frequently expressed concern that the Covid-19 pandemic may negatively affect the attractiveness of large cities due to reduced personal contacts that are crucial for productivity (e.g. knowledge spillovers) or the utility derived from endogenous amenities (e.g. pubs).

We would like to stress that we do not wish to take any stance on the likely effect of Covid-19 on productivity and QoL in cities. The below scenarios are hypothetical thought experiments and, if anything, worst-case scenarios. The reduction in the urban wage and QoL life premia will likely be partial and to some extent temporary. Moreover, even if worse comes to worst, our model predicts that it will take 30 years for 75% of the effects in Table A17 to materialise (see Figure A8).

No agglomeration economies. In this counterfactual, we hold all structural fundamentals constant. Instead, we set the agglomeration elasticity parameter $\kappa^{\theta} = 0$ after we solve for the initial SSE and before we solve for the counterfactual SSE. The results are in panel a) of Table A17. Without agglomeration economies, productivity, wages, and GDP decline in all local labour markets (LLMs). The effects are stronger in the large LLMs which lose about 9% of

Table A17: Counterfactual analysis: Other applications

(a) No agglomeration economies	All LLM	Large LLM	Small LLM
Population	1.000	0.912	1.095
GDP	0.895	0.840	0.976
Average wage	0.895	0.893	0.904
Average rent	0.942	0.969	0.996
High-skilled share	1.000	1.032	0.981
Skilled share	1.000	0.981	1.010
Average utility	0.886	0.903	0.905
(b) No social amenities			
Population	1.000	0.633	1.492
GDP	0.990	0.625	1.521
Average wage	0.990	0.988	1.019
Average rent	0.734	0.907	1.067
High-skilled share	1.000	0.999	1.173
Skilled share	1.000	0.967	0.977
Average utility	0.604	0.601	0.781
(c) Scenarios (a) and (b) combined			
Population	1.000	0.621	1.508
GDP	0.891	0.554	1.379
Average wage	0.891	0.893	0.914
Average rent	0.718	0.889	1.051
High-skilled share	1.000	1.008	1.169
Skilled share	1.000	0.961	0.979
Average utility	0.548	0.551	0.707
(d) Scenario (c) with threefold γ^{θ}			
Population	1.000	0.373	1.848
GDP	0.889	0.339	1.690
Average wage	0.887	0.907	0.915
Average rent	0.627	0.794	1.066
High-skilled share	1.000	0.746	1.342
Skilled share	1.000	1.011	0.943
Average utility	0.642	0.653	0.717

Notes: Results from model-based numerical simulations. Large (small) local labour markets (LLM) have a workplace employment of more (less) than 250k workers. All outcomes except for the last two are given in ratios of counterfactual (SSE) values over initial (SSE) initial values.

their workers to smaller LLMs. The reduction in housing demand owing to decreasing wages leads to lower rents in all LLMs. Due to the shift in demand from the larger towards smaller LLMs the effect is quantitatively small in the smaller LLMs.

No social amenities. We start from the assumption that conditional on controls the big data amenity index captures QoL-effects of endogenous amenities (e.g. cafes, concert halls, pubs) where people engage in activities that generate social media content. Solving Eq. (29) for the log of QoL gives the following specification which we take to the data in group-specific regressions:

$$\ln A^{\theta}_{i,t} = \bar{c}^{P} \tilde{\zeta^{\theta}} \ln P_i + \boldsymbol{\mathcal{X}_i' \tilde{b}^{\theta}} + \tilde{\epsilon_i}^{\theta},$$

where $\bar{c}^P \equiv \tilde{c}^P$, $\tilde{\zeta^{\theta}} = \frac{1}{\zeta^{\theta}}$, $\boldsymbol{\mathcal{X}_i' \tilde{b}^{\theta}} \equiv \sum_n (b^{\theta}_n{}^P \ln \mathcal{X}_{i,n})$ and $\tilde{\epsilon_i}^{\theta} \equiv -\ln \epsilon_i{}^{\theta}$. We include all covariates other than the residualised big data amenity index from Table 2 in $\boldsymbol{\mathcal{X}_i'}$. To evaluate an extreme case in which all amenities captured by the big data amenity conditional on covariates become obsolete, we define the counterfactual change in QoL as $\widehat{A^{\theta}_i} = \left(\frac{P^{Min}}{P_i}\right)^{\zeta^{\theta}}$, where P^{Min} is the smallest value in the distribution of the big data amenity index across regions. Otherwise, the procedure is identical to the one outlined in Section F.

The results are in panel b) of Table A17. As with the reduction in agglomeration economies, the QoL shocks hit the larger LLMs harder which is consistent with large cities offering particularly vibrant cultural, gastronomic, and nightlife amenities. The effects are considerably larger than in the no "agglomeration economies" scenario, with population size predicted to drop by almost 40% in large LLMs. The effect on overall GDP is more moderate, though there is a large drop (increase) for large (small) LLMs. The large migration into small LLMs causes rents to rise in absolute terms, whereas they naturally fall in the large LLMs.

No agglomeration economies and no social amenities. In a third scenario, we explore the joint effect of eliminating productivity and consumption benefits of big cities. As expected, the results in panel c) of Table A17 blend the results from panels a) and b). Large LLMs lose slightly more of their population than in panel b). There are large negative effects on wages in both regions and yet rents increase in the small LLMs due to the shift in demand.

No agglomeration economies and no social amenities, with threefold γ^{θ}. The last scenario in panel d) of Table A17 serves the purpose of illustrating how the frictional nature of our DSM anchors the spatial economy in the presence of a major shock. As discussed in Section C, spatial arbitrage in our model is imperfect unless the migration elasticity γ^{θ} is very large. Larger γ^{θ} necessarily imply lower migration costs $\tau^{\theta}_{ij,j\neq i}$ since $\gamma \times \tau^{\theta}_{ij}$ is jointly identified empirically. Tripling γ^{θ} brings the average across groups close to unity after which the DSM-QoL approaches the RR-QoL (see Figure 6).

In panel d) we invert the model using thrice the estimated value of γ^{θ}. We then make the same changes to κ^{θ} and A^{θ}_i as in scenario c) maintaining the large γ^{θ} values. Expectedly, the larger migration response owing to reduced idiosyncratic attachment amplifies the effects found in scenario c). Large LLMs lose more than 60% of their workers and almost 70% of their GDP. Despite a reduction in wages of about 9%, small LLMs experience rents increasing by 7% due to an increase in employment by about 85%.

The important takeaway is that the effects predicted by our model are not nearly as devastating as predicted by a canonical SSE model. The intuition is that because of idiosyncratic tastes many infra-marginal workers will not leave large LLMs even if the expected group-specific utility is larger in small LLMs. We consider this a realistic feature of our model.

N The role of expectations

In quantifying our DSM, we need to define how agents form expectations. One extreme is to assume perfectly informed agents that anticipate all future changes in goods and factor prices as well as future migration decisions. The other extreme is to assume myopic agents that extrapolate current prices into the infinite future and abstract from future migration decisions. We treat the latter as a special case that can be obtained from the former general case under some seemingly restrictive assumptions (J.3). Other special cases are imaginable in which agents anticipate future prices, but abstract from future migration decisions to varying degrees. As already pointed out by Caliendo et al. (2019b), the general case involves a dimensionality problem when all fundamentals of the model are to be inverted. Within the special cases, the complexity of the computational problem is generally larger, the less restrictive the assumptions regarding expectations are. Our methodological contribution is to develop an approach to the inversion of a QSM under perfect foresight in goods and factor prices. Guided by the stylized fact that the average worker changes the local labour market only once over the employment biography, we assume that workers do not anticipate subsequent moves when making migration decisions in the baseline quantification of our model. The purpose of this section is to compare this baseline case to more or less restrictive special cases that make the quantification of the model computationally less or more burdensome. The main takeaway is that the choice of how residents form expectations has moderate effects on the quantitative predictions of the model. With respect to the key findings in the paper – in particular the greater dispersion in regional QoL in a frictional model – the baseline case delivers the most conservative predictions. Even the fairly restrictive assumption of myopic agents appears to be a reasonable choice in high-dimensional settings.

N.1 General and special cases

We refer to *perfect foresight* as the model-consistent anticipation of all future prices on goods and factor markets. By *sequential moves*, we refer to the anticipation of migration decisions that happen in the future, subsequent to an initial migration decision.

General case with perfect foresight and multiple sequential moves. The general formulation of expected utility and migration probabilities in Section J.3 allows agents to anticipate all future group-region-specific wages and rents and imposes no restriction on the sequence of migration decisions. Two challenges lie in the way of quantifying the model in this general case. First, there is a circularity problem because we can invert $\bar{A}^{\theta}_{i,t}$ for given future wages and rents; however, to forecast future wages and rents, we require $\bar{A}^{\theta}_{i,t}$. We address this problem using the dynamic solver introduced in Section D.3 in the main paper and in Section K.3 in this appendix. Second, with sequential migration, the attractiveness of a potential migration destination j depends on the migration options j offers as a migration origin in subsequent moves. The migration option value $\mathcal{O}_j$ introduced in Eq. (25) depends on migration costs τ_{jm} and the pull factors at m, which include the migration option value $\mathcal{O}_m$. Building on Artuç et al. (2010), Caliendo et al. (2019b) show how to exploit the Bellman principle to implement the general case in a setting where the model does not have to be fully inverted. Intuitively, the migration option value cancels out when the model is solved in first differences. Caliendo et al. (2019b) use this approach to quantify the effects of the China trade shock. Using a similar approach, Caliendo et al. (2019a) evaluate the effects of EU market integration. Similarly, Balboni (2019) analyses the effects of road investments in Vietnam. We adopt a similar approach when estimating the migration elasticity in Section D.2. However, when we invert QoL, we need to solve for a structural fundamental in levels, which requires us to model the migration option value explicitly. Sequential moves imply that migration probabilities depend on migration option values at potential destinations which, in turn, depend on migration option values at subsequent destinations. This creates the dimensionality problem that we discuss in Section N.2.

Special case with perfect foresight and one sequential move. To reduce the dimensionality problem, we can assume that workers, when making their migration decisions, expect to relocate once. The assumption of perfect foresight along with the restriction to one expected sequential move leads to the following definition of the migration option value (see Appendix J.3 for the derivation):

$$\mathcal{O}^{\theta}_{j,t+2} = \frac{1}{1+\rho} \ln \Bigg[\sum_{m \in J} \left(\exp\left(-\tau^{\theta}_{jm}\right) B^{\theta}_{jm,t+2} \mathcal{V}^{\theta}_{m,t+2} \right)^{\gamma^{\theta}} \Bigg]^{\frac{1}{\gamma^{\theta}}} \tag{47}$$

Unlike in the general case in Eq. (11), the migration option value in Eq. (47) no longer depends on the migration option value at m. This restriction solves a dimensionality problem in the inversion of structural fundamentals under perfect foresight that is discussed in Caliendo et al. (2019b). Implicitly, we assume that $\mathcal{O}_m = 0$ in Eq. (25), so that $\mathcal{O}_j$ only depends on future wages, rents, QoL, and migration probabilities in m for which we solve using the dynamic solver. Of course, $\mathcal{O}_j$ includes the option to stay put at j with a high probability. Tracking an entire cohort of West German workers over their entire 40-year employment biographies, we find that almost 90% switch less than twice between local labour markets. Hence, we argue that it is reasonable to assume that workers do not expect more than one additional future relocation when making a migration decision.

Special case with perfect foresight and no sequential move. In fact, the mean number of moves across local labour markets in the data is close to one. Therefore, one might argue that workers, when making migration decisions, expect to stay put at a potential destination. Abstracting from subsequent moves further simplifies the problem as we can set $\mathcal{O}_j = 0$ in Eq. (25), allowing us to abstract from migration option values altogether. This choice is primarily a matter of taste. It does not reduce the dimensionality as the dynamic solver solves for all ingredients that enter the migration option values $\mathcal{O}_j$. Of course, the restriction still lowers the computational burden in high-dimensional settings (large J and Θ since the migration option value needs to be computed for $J \times \theta$ groups in $H = 1000$ periods in each of the iterations of the outer loop of the dynamic solver (see Figure A7).

Special case with static expectations and no sequential move. The assumption that workers correctly anticipate the transition path to the SSE is standard in a large dynamic macroeconomics literature. In the context of a DSM, the implication is that workers, when making migration decisions, process spatially disaggregated information for J interconnected regions in such a way that they correctly anticipate future demand and supply conditions on all goods and factor markets in all regions. One alternative is to assume that workers are myopic and project observed prices into the infinite future. While we assume perfect foresight in the quantification of the model, we acknowledge that we are expecting a lot from the real-world counterparts to the agents in our model and that the question whether economic agents have perfect foresight or are myopic is a philosophical one. Since myopia features as a special case in our model when wages and house prices are expected to stay constant in real terms (see Section J.3), it is easy enough to provide a comparison to evaluate the consequences of this choice. In fact, myopia greatly simplifies the inversion of the model since QoL $\bar{A}$ can be solved using migration probabilities, wages, and rents that are readily observable using a standard numerical solution algorithm. Therefore, this approach is particularly amenable to high-dimensional settings such as in Desmet et al. (2018) and Conte et al. (2020) who cover the entire planet.

N.2 Dimensionality

Imagine a worker who considers moving from region $i \in J$ to region $j \in J$. Let's further assume that the worker expects to relocate once more in the future. From any possible region j, the worker has the option to migrate to region $m \in J$. These are the J options summarised by the option value $\mathcal{O}_j$. Let's now assume that the worker expects to make yet another move from region m to $n \in J$. While staying in j the worker has then a total of $J \times J$ options that will have to be summarised by the option value. More generally, there are J^S options if the worker expects S additional moves subsequent to the initial relocation.

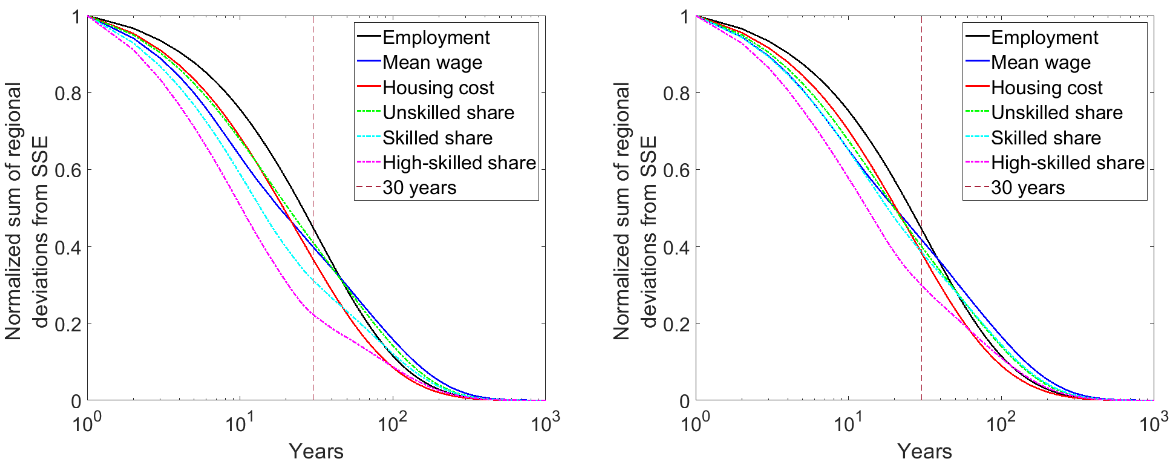
For the quantification of the model, we compute option values for $J \times \Theta$ region-groups and H forecast periods. Hence, the dimensions of the option value are $J \times \theta \times J^S \times H$. At $J = 141$, $\theta = 18$, $H = 1000$, we obtain close to 360 million migration probabilities that enter option values if we set $S = 1$, i.e. worker anticipate one subsequent move. At $S = 2$, this number increases to more than 50 billion, a dimensionality that is not amenable to the three nested iterative solution algorithms forming the dynamic solver introduced in Section D.3.

N.3 Comparison of special cases

In this section, we compare some of the main results of our quantitative analysis for the three special cases introduced in Section N.1.

Transition into the stationary spatial equilibrium. Figure A17 expands on Figure 3 from the main paper by showing the TSE-to-SSE transition path for the two special cases not reported in the main paper. The projections are fairly similar. The main difference to the baseline scenario is that the share of high-skilled workers drops if we account for one sequential move (OV). The restriction to static expectations has similarly minor effects (SP). Pooling across all region-groups, Figure A16 confirms that the speed of convergence from the TSE to SSE is fairly similar in all three special cases.

Figure A16: Spatial convergence: Alternative special cases

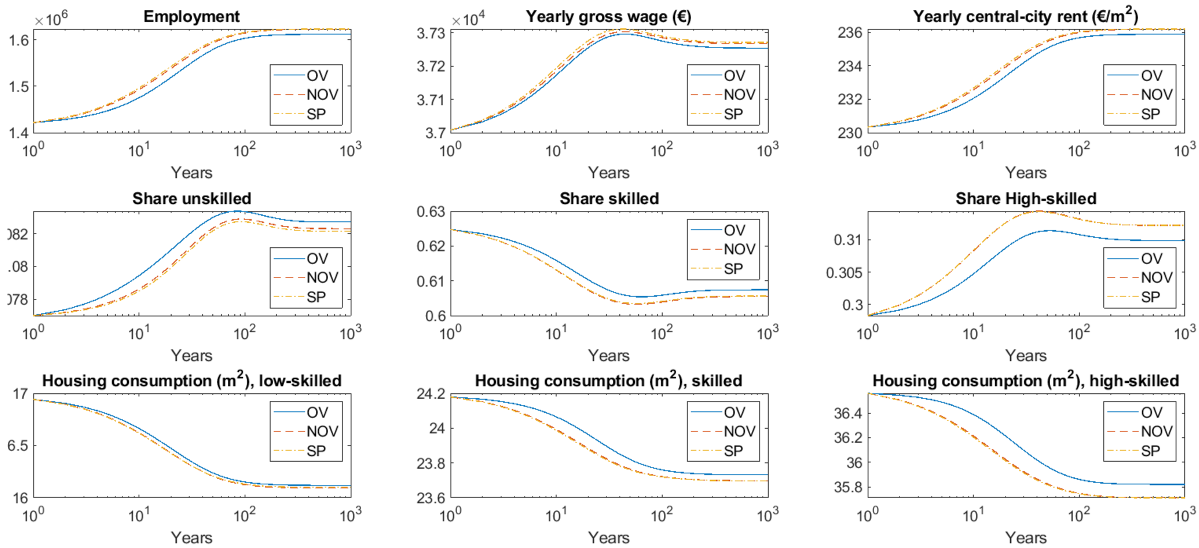


(a) Perfect foresight with one sequential move (OV) (b) Static expectations (SP)

Note: All trends show the sum of absolute deviations from SSE values in an outcome across group-regions. 2017 starting values. Model-based forecasts. OV replicates the results for the special case with one sequential move (includes option value of migration). SP replicates the results for static preferences. The baseline case with perfect foresight and no sequential moves is in Figure A8.

In Table A18, we compare average SSE outcomes (the PF scenario is identical to the SSE in Table 1 in the main paper). For most outcomes, the SSE is virtually identical across all special cases. There are notable differences in the amenity index, which is the population-weighted QoL inverted using the dynamic solver. Allowing for a sequential move, the model rationalises the data by a higher weighted QoL. Note that the unweighted average QoL ($\bar{A}^{\theta}_{i,t}$) across regions is one by definition (see Eq. (3)). The weighted average density is almost identical across the special cases, revealing a similar population distribution. Hence, the generally higher amenity index in the special case with sequential moves reveals that the model rationalises the data with a higher QoL in the densely populated areas, i.e. the urban QoL premium is greater than in the other special cases. The intuition is straightforward. The urban QoL premium is positive in Germany, even when assuming a competitive spatial equilibrium (see Table 2). Allowing for sequential moves implies that workers migrating into high-QoL urbanised regions expect to relocate to, on average, lower-QoL regions with an expected probability > 0. This depreciates the expected utility at the migration destination relative to the special case in which workers expect to stay put. Therefore, the model rationalises the larger migration flows into high-QoL urbanised regions via an even larger QoL.

Figure A17: Transition from TSE into SSE in Berlin: Varying expectations



Notes: Model-based forecasts using the dynamic solver introduced in Section D.3. 2017 starting values. Yearly gross wage, skill shares and housing consumption are weighted by group shares. NOV is the baseline result from Figure 3 in the main paper. OV replicates the results for the special case with one sequential move (includes option value of migration). SP replicates the results for static preferences.

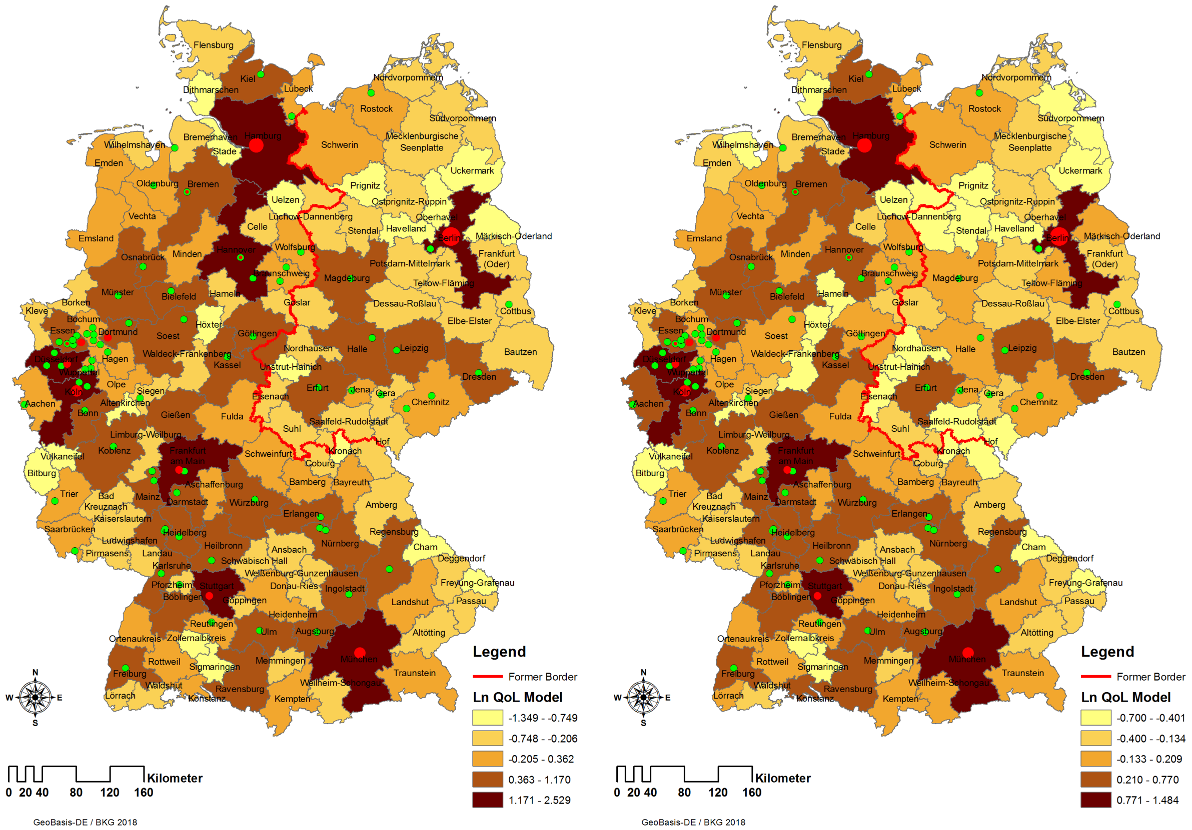
Spatial QoL differentials. Figure A18 replicates the left panel of Figure 4 in the main paper for the special cases of perfect foresight with one sequential move and static expectations. The figure substantiates the conclusions from Table A18. The spatial distribution of QoL in the case of static expectations closely resembles the baseline case. There is even more dispersion in the spatial case with perfect foresight and one sequential move. The region-level correlations in Figure A19 further substantiate this impression. We conclude that our central finding that spatial QoL differences are much larger once mobility frictions are accounted for does not depend on how we model expectations. In fact, if workers anticipate a subsequent move in their initial migration decision, the inverted spatial QoL differences become even larger than reported in the main paper.

Table A18: SSE: Varying expectations

Outcome	NOV	OV	SP
Output in bn.	1.056	1.056	1.056
Amenity index	1.639	2.164	1.553
Weighted average density (emp./km^2)	144.462	143.383	144.437
Amenity index, unskilled	2.272	3.142	2.082
Amenity index, skilled	1.414	1.742	1.363
Amenity index, high-skilled	2.218	3.330	2.043
Weighted density, unskilled	161.940	161.198	161.805
Weighted density, skilled	136.149	135.413	136.134
Weighted density, high-skilled	168.040	165.580	168.020
Yearly wage (€), unskilled	23239	23239	23239
Yearly wage (€), skilled	33722	33721	33722
Yearly wage (€), high-skilled	50773	50742	50773
Yearly housing cost (€/m^2), unskilled	133.601	133.452	133.495
Yearly housing cost (€/m^2), skilled	120.420	120.232	120.420
Yearly housing cost (€/m^2), high-skilled	152.752	151.453	152.748
Housing consumption m^2, unskilled	43.656	43.657	43.684
Housing consumption m^2, skilled	70.489	70.528	70.488
Housing consumption m^2, high-skilled	85.577	86.239	85.578

Notes: All SSE values are model-based forecasts of the dynamic solver. QoL index is normalised within-group measure $\bar{A}^{\theta}_{i,t}$, weighted by group-region employment $L^{\theta}_{i,t}$. OV is the special case with perfect foresight and one sequential move. NOV replicates the results for the special case with perfect foresight without sequential moves (the SSE in Table 1). SP replicates the results for the special case with static expectations.

Figure A18: Spatial variation in quality of life: Varying expectations



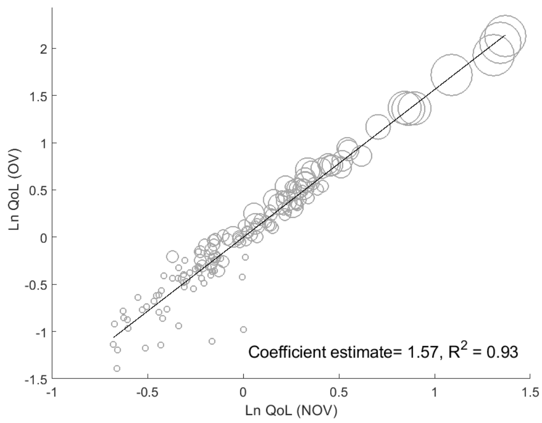
(a) $\bar{A}_i$, perfect foresight, one sequential move (OV)

(b) $\bar{A}_i$, static expectations (SP)

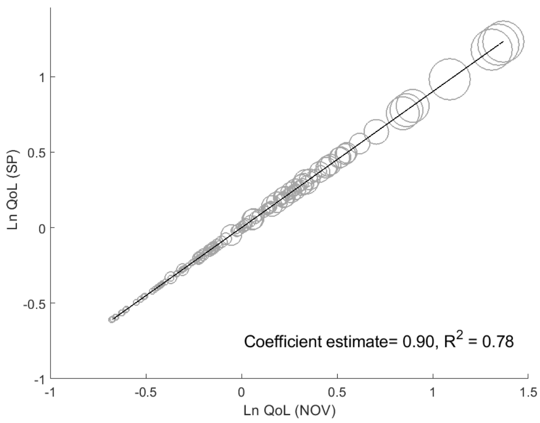
Note: Unit of observation is 141 labour market areas as defined by Kosfeld and Werner (2012). Group adjustment in auxiliary regressions of ln(QoL) against group and region fixed effects, the latter being shown on the maps. The baseline case with perfect foresight and no sequential moves is in Figure 4 in the main paper.

Counterfactual analysis. In Figure A20, we compare some key outcomes of the counterfactual analysis of the effects of a regional pollution reduction in heavily polluted regions from Section F across the three special cases. The counterfactual effects on wages, rents, and indirect utility are very similar, both in levels and trends. This is an important insight as it suggests that the computationally much less demanding special case with static preferences represents a reasonable approximation for high-dimensional settings.

Figure A19: Correlation in quality of life across special cases



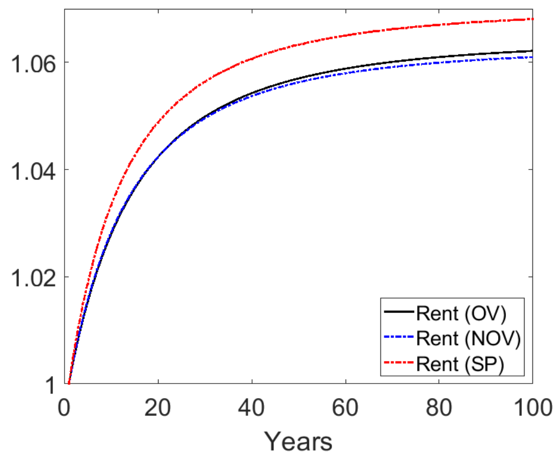
(a) Perfect foresight and sequential move vs. perfect foresight and no sequential moves



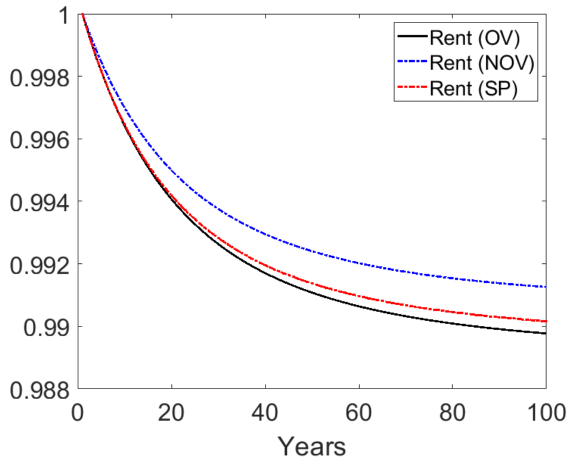
(b) Static expectation vs. perfect foresight and no sequential moves

Note: Unit of observation is group-weighted regional QoL level. OV = perfect foresight, one sequential move, NOV = perfect foresight, no sequential move. SP = static expectations.

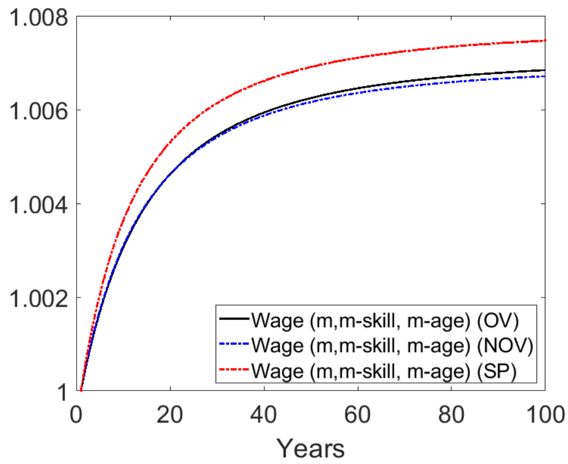
Figure A20: Counterfactual analysis: Comparison across special cases



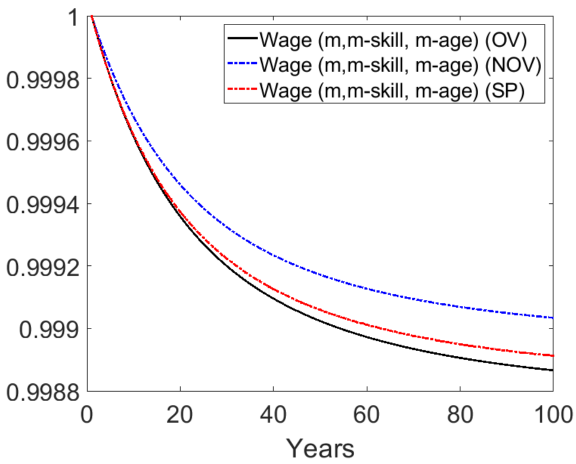
(a) Rent Bochum



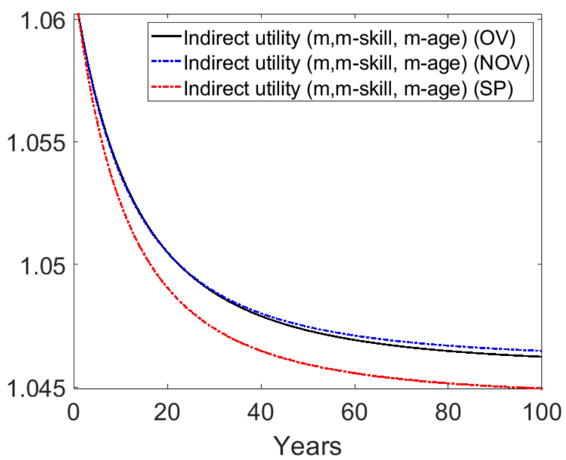
(b) Rent, Munich



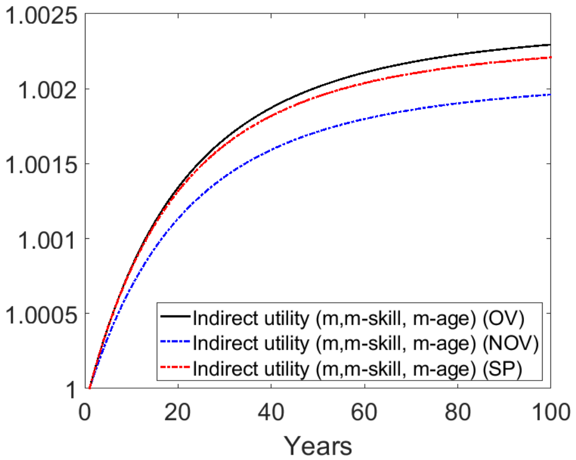
(c) Wage Bochum



(d) Wage Munich



(e) Utility Bochum



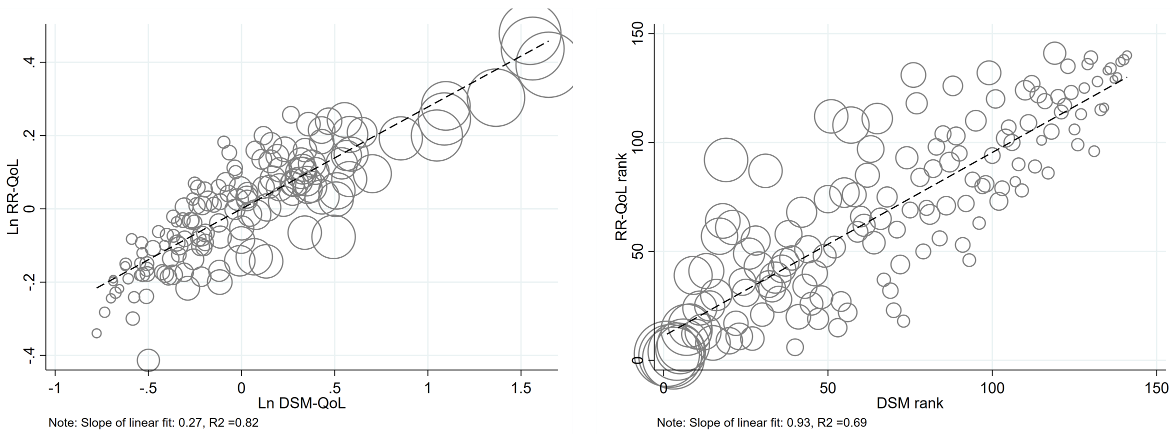
(f) Utility Munich

Note: Counterfactual experiment is a QoL increase induced by a pollution reduction in the most polluted regions. See Figure 8 for an illustration. OV = perfect foresight, one sequential move, NOV = perfect foresight, no sequential move. SP = static expectations.

O Quality-of-life rankings

In Table A19, we provide a ranking of regions according to QoL in the spirit of Blomquist et al. (1988) and Albouy (2011). We use the region-level group-mix adjusted QoL measures displayed in Figure 4. Confirming the evidence presented in Section E, DSM-QoL and RR-QoL are closely correlated at the regional level, in logs and ranks (see also Figure A21).

Figure A21: DSM-QoL vs. RR-QoL



(a) Logs (b) Ranks

Note: Unit of observation is 141 labour market areas as defined by Kosfeld and Werner (2012). Group adjustment in auxilliary regressions of ln quality of life against group and region fixed effects, the latter being shown in figures. Marker size proportionate to employment in local labour market.

Table A19: Quality of life rankings

Labour market	**DSM-QoL Rank**	**DSM-QoL in logs**	**RR-QoL Rank**	**RR-QoL in logs**	**Rank difference**
Berlin	1	1.648	3	0.393	2
München	2	1.563	2	0.438	0
Hamburg	3	1.549	1	0.478	-2
Frankfurt am Main	4	1.366	4	0.303	0
Köln	5	1.097	5	0.281	0
Düsseldorf	6	1.089	7	0.245	1
Stuttgart	7	1.048	14	0.200	7
Hannover	8	0.855	16	0.192	8
Nürnberg	9	0.702	39	0.096	30
Mainz	10	0.647	12	0.209	2
Leipzig	11	0.588	24	0.149	13
Münster	12	0.581	13	0.203	1
Karlsruhe	13	0.578	41	0.081	28
Heidelberg	14	0.563	25	0.146	11
Dresden	15	0.553	8	0.243	-7
Bonn	16	0.538	30	0.124	14
Bremen	17	0.512	57	0.049	40
Ludwigshafen	18	0.507	64	0.026	46
Essen	19	0.494	92	-0.076	73
Freiburg	20	0.467	9	0.239	-11
Bielefeld	21	0.432	61	0.033	40
Ingolstadt	22	0.430	17	0.184	-5
Regensburg	23	0.430	11	0.215	-12
Würzburg	24	0.397	36	0.103	12
Böblingen	25	0.378	31	0.124	6
Koblenz	26	0.374	49	0.060	23
Erlangen	27	0.359	10	0.231	-17
Heilbronn	28	0.355	55	0.055	27
Ravensburg	29	0.350	43	0.076	14
Darmstadt	30	0.341	21	0.160	-9
Dortmund	31	0.339	87	-0.065	56
Ulm	32	0.329	35	0.104	3
Kiel	33	0.328	33	0.112	0
Aachen	34	0.320	38	0.098	4
Augsburg	35	0.320	28	0.131	-7
Gießen	36	0.307	42	0.076	6
Erfurt	37	0.286	45	0.071	8
Kassel	38	0.274	58	0.046	20
Osnabrück	39	0.273	47	0.068	8
Konstanz	40	0.265	6	0.257	-34
Traunstein	41	0.232	20	0.163	-21
Soest	42	0.225	68	0.021	26
Göttingen	43	0.205	34	0.107	-9
Magdeburg	44	0.205	51	0.059	7
Oldenburg	45	0.192	26	0.145	-19
Braunschweig	46	0.168	40	0.082	-6
Rostock	47	0.158	19	0.179	-28
Landshut	48	0.155	29	0.130	-19
Reutlingen	49	0.152	48	0.061	-1
Halle	50	0.138	74	0.004	24

–Table continued from previous page

Labour market	**DSM-QoL Rank**	**DSM-QoL in logs**	**RR-QoL Rank**	**RR-QoL in logs**	**Rank difference**
Bochum	51	0.132	112	-0.143	61
Fulda	52	0.118	52	0.059	0
Bamberg	53	0.117	15	0.200	-38
Aschaffenburg	54	0.109	27	0.135	-27
Saarbrücken	55	0.078	77	-0.016	22
Trier	56	0.074	22	0.160	-34
Chemnitz	57	0.068	108	-0.131	51
Heidenheim	58	0.044	76	-0.015	18
Kempten	59	0.032	59	0.044	0
Lübeck	60	0.029	66	0.023	6
Ortenaukreis	61	0.029	62	0.032	1
Rottweil	62	0.001	85	-0.052	23
Minden	63	-0.001	97	-0.085	34
Wolfsburg	64	-0.007	54	0.057	-10
Hagen	65	-0.009	111	-0.142	46
Schwerin	66	-0.021	65	0.025	-1
Teltow-Fläming	67	-0.032	37	0.100	-30
Schweinfurt	68	-0.035	75	-0.002	7
Jena	69	-0.037	32	0.112	-37
Weilheim-Schongau	70	-0.066	23	0.153	-47
Bayreuth	71	-0.070	60	0.042	-11
Vechta	72	-0.084	44	0.074	-28
Märkisch-Oderland	73	-0.095	18	0.182	-55
Emsland	74	-0.114	93	-0.077	19
Göppingen	75	-0.114	69	0.020	-6
Wuppertal	76	-0.117	131	-0.200	55
Olpe	77	-0.120	118	-0.169	41
Pforzheim	78	-0.122	84	-0.038	6
Lörrach	79	-0.123	50	0.060	-29
Schwäbisch Hall	80	-0.149	70	0.011	-10
Borken	81	-0.179	67	0.021	-14
Kaiserslautern	82	-0.194	88	-0.066	6
Limburg-Weilburg	83	-0.195	98	-0.092	15
Memmingen	84	-0.197	56	0.053	-28
Potsdam-Mittelmark	85	-0.202	104	-0.106	19
Altötting	86	-0.211	71	0.009	-15
Amberg	87	-0.213	91	-0.074	4
Emden	88	-0.216	126	-0.185	38
Siegen	89	-0.220	103	-0.103	14
Frankfurt (Oder)	90	-0.231	95	-0.082	5
Deggendorf	91	-0.234	53	0.059	-38
Landau	92	-0.239	72	0.009	-20
Oberhavel	93	-0.251	46	0.069	-47
Bad Kreuznach	94	-0.253	83	-0.037	-11
Flensburg	95	-0.258	110	-0.138	15
Waldshut	96	-0.261	63	0.027	-33
Cottbus	97	-0.268	80	-0.031	-17
Passau	98	-0.273	81	-0.032	-17
Bautzen	99	-0.289	132	-0.217	33
Ansbach	100	-0.305	94	-0.082	-6

– Table continued from previous page

Labour market	**DSM-QoL Rank**	**DSM-QoL in logs**	**RR-QoL Rank**	**RR-QoL in logs**	**Rank difference**
Goslar	101	-0.307	120	-0.173	19
Coburg	102	-0.308	73	0.005	-29
Nordvorpommern	103	-0.315	79	-0.031	-24
Pirmasens	104	-0.336	102	-0.101	-2
Elbe-Elster	105	-0.336	107	-0.125	2
Kleve	106	-0.346	100	-0.093	-6
Ostprignitz-Ruppin	107	-0.349	82	-0.035	-25
Celle	108	-0.365	90	-0.074	-18
Donau-Ries	109	-0.367	78	-0.021	-31
Dessau-Roßlau	110	-0.369	124	-0.180	14
Mecklenburgische Seenplatte	111	-0.380	109	-0.136	-2
Bremerhaven	112	-0.392	127	-0.186	15
Stade	113	-0.401	89	-0.071	-24
Gera	114	-0.409	122	-0.178	8
Weißenburg-Gunzenhausen	115	-0.414	101	-0.099	-14
Suhl	116	-0.427	119	-0.173	3
Cham	117	-0.446	86	-0.062	-31
Südvorpommern	118	-0.475	105	-0.107	-13
Hof	119	-0.500	141	-0.413	22
Hameln	120	-0.505	121	-0.178	1
Eisenach	121	-0.508	114	-0.150	-7
Saalfeld-Rudolstadt	122	-0.509	117	-0.165	-5
Wilhelmshaven	123	-0.511	135	-0.239	12
Zollernalbkreis	124	-0.530	123	-0.179	-1
Uckermark	125	-0.533	106	-0.121	-19
Nordhausen	126	-0.537	99	-0.093	-27
Lüchow-Dannenberg	127	-0.544	113	-0.148	-14
Altenkirchen	128	-0.547	125	-0.181	-3
Havelland	129	-0.577	136	-0.241	7
Waldeck-Frankenberg	130	-0.584	139	-0.299	9
Dithmarschen	131	-0.589	96	-0.082	-35
Stendal	132	-0.608	128	-0.191	-4
Sigmaringen	133	-0.621	115	-0.150	-18
Bitburg	134	-0.624	116	-0.156	-18
Freyung-Grafenau	135	-0.655	133	-0.218	-2
Höxter	136	-0.676	134	-0.229	-2
Vulkaneifel	137	-0.688	129	-0.192	-8
Kronach	138	-0.689	130	-0.198	-8
Uelzen	139	-0.701	137	-0.244	-2
Unstrut-Hainich	140	-0.735	138	-0.282	-2
Prignitz	141	-0.777	140	-0.340	-1

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Ahlfeldt, Gabriel M., "Urbanity," *CESIfo working paper*, 2013, *4533.*

— **and Elisabetta Pietrostefani**, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

— **, Fabian Bald, Duncan Roth, and Tobias Seidel**, "Qality of life in a dynamic spatial model," *Working Paper*, 2020.

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," *NBER Working Paper*, 2011, *14472.*

— **and Bert Lue**, "Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life," *Journal of Urban Economics*, 2015, *89*, 74–92.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 2010, *100* (3), 1008–1045.

Atkinson, Anthony B., "On the measurement of inequality," *Journal of Economic Theory*, 1970, *2* (3), 244–263.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 2018, *32* (3), 259–280.

— **, Theresa Kuchler, Dominic Russel, Bogdan State, and Johannes Stroebel**, "Social Connectedness in Europe," Technical Report, SocArXiv 2020.

Balboni, Clare, "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities," 2019, *Working paper.*

Baum-Snow, Nathaniel and Ronni Pavan, "Inequality and City Size," *The Review of Economics and Statistics*, 2013, *95* (5), 1535–1548.

Bleakley, Hoyt and Jeffrey Lin, "Portage and Path Dependence," *The Quarterly Journal of Economics*, 4 2012, *127* (2), 587–644.

Blomquist, Glenn C., Mark C. Berger, and John P. Hoehn, "New Estimates of Quality of Life in Urban Areas," *The American Economic Review*, 1988, *78* (1), 89–107.

Boelmann, Barbara and Sandra Schaffner, "Real-Estate Data for Germany (RWI-GEO-RED v1) - Advertisements on the Internet Platform ImmobilienScout24 2007-03/2019," Technical Report, RWI Leibniz-Institut für Wirtschaftsforschung 2019.

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

Burda, Michael C. and Jennifer Hunt, "From reunification to economic integration: Productivity and the labor market in Eastern Germany," *Brookings Papers on Economic Activity*, 2001.

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement," *NBER Working paper*, 2019, *23695.*

— **, Maximiliano Dvorkin, and Fernando Parro**, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Carlino, Gerald A. and Albert Saiz, "Beautiful city: Leisure amenities and urban growth," *Journal of Regional Science*, 2019, *59* (3), 369–408.

Ciccone, Antonio and Robert E. Hall, "Productivity and the Density of Economic Activity," *The American Economic Review*, 1996, *86* (1), 54–70.

Clark, Colin, "Urban Population Densities," *Journal of the Royal Statistical Society. Series A (General)*, 1951, *114* (4), 490–496.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, chapter 5, pp. 247–348.

_ , Gilles Duranton, and Laurent Gobillon, "Spatial wage disparities: Sorting matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

_ , _ , and _ , "The Costs of Agglomeration: House and Land Prices in French Cities," *The Review of Economic Studies*, 2019, *86* (4), 1556–1589.

Conte, Bruno, Klaus Desmet, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "Local Sectoral Specialization in a Warming World," *NBER Working Paper*, 2020, *28163.*

Davis, Donald R. and David E. Weinstein, "Bones, Bombs, and Break Points: The Geography of Economic Activity ," *American Economic Review*, 2002, *92* (5), 1269–1289.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Unbalanced Trade," *American Economic Review*, 2007, *97* (2), 351–355.

Demographia, "Demographia World Urban Areas," 2019.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif, "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, 2019, *109* (12), 4178–4219.

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *Journal of Political Economy*, 2018, *126* (3), 903–983.

Epple, Dennis, Brett Gordon, and Holger Sieg, "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, 2010, *100* (3), 905–924.

Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Südekum, "Dialects, cultural identity, and economic exchange," *Journal of Urban Economics*, 2012, *72* (2-3), 225–239.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–288.

Fernihough, Alan and Kevin H. O'Rourke, "Coal and the European Industrial Revolution," *The Economic Journal*, 11 2021, *131* (635), 1135–1149.

Fitzenberger, Bernd, Aderonke Osikominu, and Robert Völter, "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 2006, *126* (3), 405–436.

Flockton, Chris, "Housing situation and housing policy in east Germany," *German Politics*, 1998, *7* (3), 69–82.

Gaigné, Carl, Hans R.A. Koster, Fabien Moizeau, and Jacques-François Thisse, "Who Lives Where in the City? Amenities, Commuting and Income Sorting," *CEPR Discussion Paper*, 2017, *11958.*

Georgi, Sabine and Peter Barkow, "Wohnimmobilien-Indizes: Vergleich Deutschland-Großbritannien," Technical Report, Zentraler Immobilien Ausschuss e.V. 2010.

Harris, Chauncy D., "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 1954, *44* (4), 315–348.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse,Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, pp. 131–195.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg, "East Side Story: Historical Pollution and Persistent Neighborhood Sorting," *Journal of Political Economy*, 2020, *forthcoming.*

Hohn, Uta, *Die Zerstörung deutscher Städte im Zweiten Weltkrieg : Regionale Unterschiede in der Bilanz der Wohnungstotalschäden und Folgen des Luftkrieges unter bevölkerungsgeographischem Aspekt*, Dortmund: Dortmunder Vertrieb für Bau- u. Planungsliteratur, 1991.

Imbert, Clément and John Papp, "Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India," *Journal of the European Economic Association*, 2019, *18* (2), 927–963.

Kasperski, Michael, "A new wind zone map of Germany," *Journal of Wind Engineering and Industrial Aerodynamics*, 2002, *90* (11), 1271–1287.

Kosfeld, Reinhold and Alexander Werner, "Deutsche Arbeitsmarktregionen-Neuabgrenzung nach den Kreisgebietsreformen 2007-2011," *Raumforschung und Raumordnung*, 2012, *70* (1), 49–64.

Koster, Hans R. A. and Edward W. Pinchbeck, "How do Households Value the Future? Evidence from Property Taxes," *American Economic Journal: Economic Policy*, 2021, *forthcoming.*

Mohammed, Saif I.Shah and Jeffrey G. Williamson, "Freight rates and productivity gains in British tramp shipping 1869 - 1950," *Explorations in Economic History*, 2004, *41* (2), 172–203.

Monras, Joan, "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis," *Journal of Political Economy*, 2020, *128* (8), 3017–3089.

Moore, Michael J. and W. Kip Viscusi, "The quantity-adjusted value of life," *Economic Inquiry*, 1988, *26* (3), 369–388.

OECD, "Taxing Wages 2017," 2017.

Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf, "History and Industry Location: Evidence from German Airports," *The Review of Economics and Statistics*, 2010, *93* (3), 814–831.

Saiz, Albert, Arianna Salazar, and James Bernard, "Crowdsourcing architectural beauty: Online photo frequency predicts building aesthetic ratings," *PLoS ONE*, 2018, *13* (7).

Statistisches Bundesamt, "Einkommens- und Verbrauchsstichprobe Konsumausgaben privater Haushalte," *Fachserie*, 2020, *15* (5).

Tobler, W. R., "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, 1970, *46*, 234–240.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872.