## ARTICLE

Check for updates

# Gravity models do not explain, and cannot predict, international migration dynamics

Robert M. Beyer [1✉], Jacob Schewe[1] & Hermann Lotze-Campen [1,2]

The major social and economic impacts of international migration have led to a strong interest in better understanding the drivers of cross-border movement. Quantitative models have sought to explain global migration patterns in terms of economic, social, climatic, and other variables, and future projections of these variables are increasingly being used to forecast international migration flows. An important implicit assumption in the most widely used class of these approaches, so-called gravity models, is that their parameterisation based on panel data enables them to describe the effects of predictor variables on migration flows across both space and time, i.e., that they explain flow variation both across country pairs at a given time and across time for a given country pair. Here we show that this assumption does not hold. Whilst gravity models describe spatial patterns of international migration very well, they fail to capture even basic temporal dynamics, indeed, often worse than even the time-invariant average of the historical flows. We show that standard validation techniques have been unable to detect this important limitation of gravity models due to the different orders of magnitude of migration flows across spatial corridors, on the one hand, and over time, on the other hand. Our analysis suggests that gravity-model-based inferences about the effects that certain variables have had, or will have, on international migration over time may in reality represent statistical artefacts rather than true mechanisms. We argue that future predictions based on gravity models lack statistical support and that, in its current form, this class of models is not suited for informing policy makers about migration trajectories in the coming years and decades.

[1] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany. [2] Humboldt-Universität zu Berlin, Berlin, Germany. ✉email: robert.beyer@pik-potsdam.de

## Introduction

International migration has shaped global socio-political discourses in recent decades unlike any other demographic process and has had lasting effects on domestic political landscapes (Kapur, 2014; Trauner and Turton, 2017; Blinder and Allen, 2020) and international relationships, both between countries of origin and destination and between different countries of destination (Papademetriou and Banulescu-Bogdan, 2016; Martin, 2017). A strong interest in better understanding the drivers of cross-border movements, coupled with an increasing availability of global bilateral migration estimates (Özden et al., 2011; Abel, 2018), has led to the development of quantitative models in which migration flows are estimated as a function of predictors such as country- or country-pair-specific geographic, demographic, political, economic, environmental, linguistic, or cultural variables (e.g., Belot and Ederveen, 2012; Backhaus et al., 2015; Adserà and Pytliková, 2015; Poot et al., 2016; Cai et al., 2016; Cattaneo and Peri, 2016).

Gravity models represent the most widely used, and still rapidly growing (Ramos, 2016), class of such approaches. Originally referring to a characterisation of migration flow in terms of the population sizes of the countries of origin and destination and the geographic distance between them (in analogy to Newton's law of gravitation), the term has since been used more generally for linear or nonlinear regressions that relate migration flows to relevant predictor variables (Beine et al., 2016; Poot et al., 2016; Ramos, 2016). Unknown parameters in the regression function are estimated based on panel data observations from across both countries and time. Gravity models have been used to explain historical migration patterns, and are being used increasingly to seek insights into future global migration flows. For example, gravity models have been used to infer that migration from poor countries will increase as national income levels rise in the coming decades (Docquier, 2018; Rikani and Schewe, 2021), and, in the context of future climatic changes, that migration increases as the result of environmental degradation (Afifi and Warner, 2008; Reuveny and Moore, 2009), more frequent natural disasters (Afifi and Warner, 2008; Ragazzi, 2012; Gröschl and Steinwachs, 2017), higher temperatures (Backhaus et al., 2015; Maurel and Tuccio, 2016; Cai et al., 2016; Cattaneo and Peri, 2016; Helbling and Meierrieks, 2021), or exposure and vulnerability to climate change (Benveniste et al., 2020).

Whilst gravity models have been assumed to map effects "both over time and across countries" (Beine and Parsons, 2015) given the spatio-temporal panel data used to parameterise them, it is important to note that it is entirely possible for a gravity model to match the calibration data very well without describing any temporal patterns correctly. This is because migration flows across countries vary over several orders of magnitude, whereas flows to or from a given country over time generally do not. As a result, a strong performance of a gravity model in terms of capturing spatial patterns can mask a poor performance in terms of mapping temporal trends. This becomes problematic when gravity models that essentially describe only spatial patterns of migration are used to draw inferences about how migration changes through time, both in terms of explaining past and predicting future flows. For example, a spatial correlation between out-migration and some climatic variable, represented in a gravity model, may be used as a statistical basis for predicting how climatic changes will affect out-migration in a certain country over time; however, when correlating the country's historical out-migration and climatic record, the direction of the relationship may actually be opposite to that found at the spatial level (i.e., across countries). This fallacy in the application of gravity models may lead to false conclusions about past or future migration dynamics.

Here, we demonstrate that these caveats are not merely theoretical, but form an important part of the statistical mechanics of gravity models. We consider simple and complex examples of two important categories of gravity models and parameterise them using global panel data. Whilst the models fit the spatio-temporally pooled data very well, they fail to capture even basic temporal trends in the observations. Our analysis reveals that the implicit assumption made in gravity models used in temporal contexts—that migration dynamics follow statistical rules similar to those that describe spatial migration patterns—is generally not justified, and that the former rules are considerably more complex than the latter ones. We conclude that gravity models calibrated based on spatio-temporally pooled data are not suited for predicting future international migration.

## Methods

**Model design and parameterisation.** We considered bilateral gravity models representing two model categories characterised by the presence of origin–destination-specific fixed effects in the one case and their absence in the other. Nicolli and Bettin (2012), Beine and Parsons (2015), Backhaus et al. (2015), Cai et al. (2016) represent examples of the former category, Docquier (2018), Reuveny and Moore (2009), Rikani and Schewe (2021), and Cohen et al. (2008) of the latter. Whilst sharing important similarities in the way they statistically represent migration flows, models from these two categories differ in key aspects with regard to how they map spatial and temporal patterns, to which we return later on. For each of the two categories, we examined a simple and a complex gravity model, based on only a few predictors in the one case and a larger number in the other. This allows us to show that the deficiencies of gravity models emerge already for a simple setup and do not necessarily subside as model complexity increases. First, we considered

$$\text{Model 1}: \log\left(M_{ijt}\right) \sim p_0 + p_1 \cdot \log(P_{it}) + p_2 \cdot \log\left(P_{jt}\right) \\ + p_3 \cdot \log(I_{it}) + p_4 \cdot \log\left(I_{jt}\right) + p_5 \cdot \log\left(D_{ij}\right),$$
(1)

where $M_{ijt}$ denotes the number of migrants that move from country $i$ to country $j$ in year $t$, $P_{it}$ and $I_{it}$ represent the population size and per-capita gross domestic product (GDP) in country $i$ in year $t$, respectively, and $D_{ij}$ denotes the geographic distance between countries $i$ and $j$. The predictors in Eq. (1) are present in almost all gravity models in previous studies and are often used as a basis on top of which the roles of specific additional variables of interest are investigated.

Our second model provides an example of such a more complex model. In addition to the above variables, it includes additional economic as well as social, political, educational, demographic, linguistic, historical, geographical, and health-related predictors:

$$\text{Model 2}: \log\left(M_{ijt}\right) \sim p_0 + p_1 \cdot \log(P_{it}) + p_2 \cdot \log\left(P_{jt}\right) \\ + p_3 \cdot \log\left(I_{it}\right) + p_4 \cdot \log\left(I_{jt}\right) \\ + p_5 \cdot U_{it} + p_6 \cdot U_{jt} + p_7 \cdot E_{it} \\ + p_8 \cdot E_{jt} + p_9 \cdot H_{it} + p_{10} \cdot H_{jt} \\ + p_{11} \cdot S_{it} + p_{12} \cdot S_{jt} + p_{15} \cdot Y_{it} \\ + p_{16} \cdot B_{jt} + p_{17} \cdot C_{ij} + p_{18} \cdot L_{ij} \\ + p_{19} \cdot \log\left(D_{ij}\right)$$
(2)

Here, $U_{it}$ represents the unemployment rate in country $i$ in year $t$, $E_{it}$ represents education measured in expected years of

schooling, $H_{it}$ represents general health measured in life expectancy at birth, $S_{it}$ represents political stability, $Y_{it}$ represents the proportion of the population aged 20–35 (associated with increased mobility), and $B_{it}$ represents immigration restrictiveness. With the exception of $Y_i$ and $B_j$, which are likely relevant only in the country of origin and destination, respectively, we included all of the aforementioned variables from both the country of origin and destination. The origin-destination-specific and time-invariant variables $C_{ij}$ and $L_{ij}$ take the value 1 if countries $i$ and $j$ share a colonial link and an official language, respectively, and 0 otherwise. Unlike $M_{ijt}$, $P_{it}$, $I_{it}$, and $D_{ij}$, the additional variables in Model 2 are not approximately log-normally distributed and were therefore not logarithmised.

Analogous to Models 1 and 2, respectively, we additionally considered the fixed-effects models

$$\text{Model 3}: \log\left(M_{ijt}\right) \sim p_1 \cdot \log(P_{it}) + p_2 \cdot \log\left(P_{jt}\right) \\ + p_3 \cdot \log(I_{it}) + p_4 \cdot \log\left(I_{jt}\right) + c_{ij} \quad (3)$$

and

$$\text{Model 4}: \log\left(M_{ijt}\right) \sim p_1 \cdot \log(P_{it}) + p_2 \cdot \log\left(P_{jt}\right) \\ + p_3 \cdot \log\left(I_{it}\right) + p_4 \cdot \log\left(I_{jt}\right) \\ + p_5 \cdot U_{it} + p_6 \cdot U_{jt} + p_7 \cdot E_{it} \\ + p_8 \cdot E_{jt} + p_9 \cdot H_{it} + p_{10} \cdot H_{jt} \\ + p_{11} \cdot S_{it} + p_{12} \cdot S_{jt} + p_{13} \cdot R_{it} \\ + p_{14} \cdot R_{jt} + p_{15} \cdot Y_{it} + p_{16} \cdot B_{jt} + c_{ij} \quad (4)$$

in which the $c_{ij}$ are constants depending on the country of origin and destination. They are meant to capture the effects of time-invariant factors, including the geographic distance between countries and whether they share a common language and colonial history, which in Models 1 and 2 had been assumed to follow a specific functional relationship.

For the four models, we estimated the parameters $p_n$, and $c_{ij}$ in Models 3 and 4, using a least-squares approach based on the following observational data. Historical migration flows, $M_{ijt}$, were obtained from a dataset covering 206 countries of origin, 37 OECD countries of destination, and the period 1995–2019 in annual time steps (OECD, 2021). We selected these data for our main analysis because their annual resolution allows us to assess temporal trends particularly well. Corresponding annual national per-capita GDPs, $I_{it}$, during 1995–2019 were obtained from an updated version of the dataset in James et al. (2012), population sizes, $P_{it}$, and unemployment rates, $U_{it}$, were obtained from The World Bank (2021), the expected number of schooling years from UNDP (2021a), life expectancy at birth from UNDP (2021b), the proportion of 20–35 year-olds from Lutz et al. (2018), political stability from the Economist Intelligence Unit (2019), and immigration restrictiveness (available only until 2010) from Helbling et al. (2017). After removing zero flows and missing data, we retained a total of 17,759 and 9166 data points for the parameterisation of Models 1 and 3 and of Models 2 and 4, respectively.

To demonstrate that our findings are not merely the result of the annual temporal resolution of the above data, we repeated our analyses using a global dataset of bilateral migration flows in 5-year intervals (Abel and Cohen, 2019). For comparability with our main analysis, we subset these data to the same country pairs and the time period as in the annual data. Corresponding predictor variables were computed as 5-year averages of the relevant annual data. For this setup, a total of 5265 and 3957 data points were available for the parameterisation of Models 1 and 3 and of Model 2 and 4, respectively.

The estimated parameter values for the different setups are provided in Supplementary Table 1.

**Model comparison against spatio-temporally pooled data**. Using the estimated parameters, we computed the $R^2$ value of each model to assess the proportion of the variance in the observed data that the model explains,

$$R^2 \overset{\text{def}}{=} 1 - \frac{\sum_{i,j,t} \left(\log\left(M_{ijt}\right) - \log\left(\widehat{M}_{ijt}\right)\right)^2}{\sum_{i,j,t} \left(\log\left(M_{ijt}\right) - \overline{\log(M)}\right)^2}, \quad (5)$$

where $M_{ijt}$ and $\widehat{M}_{ijt}$, respectively, denote the observed and the (model-specific) simulated migration flow from country $i$ to country $j$ at time $t$, and where $\overline{\log(M)}$ denotes the mean of the logarithmised observed flows across all $i,j,t$.

The different orders of magnitude of the spatial, compared to the temporal, variation in the observed migration flows make it possible, in principle, for a gravity model to achieve a high $R^2$ value despite not capturing temporal patterns, as noted above. We can largely isolate the models' performances in terms of describing temporal, rather than spatial, signals (see Supplementary Note 1) by comparing the available observed and simulated relative changes in migration over time,

$$\Delta M_{ijt} \overset{\text{def}}{=} \frac{M_{ijt} - M_{ij(t-1)}}{M_{ij(t-1)}} \quad \text{and} \quad \Delta\widehat{M}_{ijt} \overset{\text{def}}{=} \frac{\widehat{M}_{ijt} - \widehat{M}_{ij(t-1)}}{\widehat{M}_{ij(t-1)}}. \quad (6)$$

Note that $\Delta M_{ijt}$ is the discrete derivative of $\log(M_{ijt})$ with respect to time. Analogous to the above-described comparison of the two models against absolute flow data, we computed $R^2$ values for each model based on the observed and simulated relative changes in flows,

$$\Delta R^2 \overset{\text{def}}{=} 1 - \frac{\sum_{i,j,t} \left(\Delta M_{ijt} - \Delta\widehat{M}_{ijt}\right)^2}{\sum_{i,j,t} \left(\Delta M_{ijt} - \overline{\Delta M}\right)^2}. \quad (7)$$

If a gravity model describes temporal trends in the migration data very well, then $\Delta R^2 \approx 1$, whereas if $\Delta R^2 < 0$, then even the constant (time- and country-invariant) mean of all observed rates of change, denoted $\overline{\Delta M}$, provides an overall better estimate than the modelled data.

**Model comparison against bilateral data**. The previous two types of model evaluation based on spatio-temporally pooled data can provide very general indications of the models' performance in terms of describing absolute flows and temporal change rates; however, these statistical summaries may be of limited informative value with regard to the flows over time between any specific pair of countries. For the four models, we therefore also compared, separately for each country of origin $i$ and destination $j$, the modelled and observed migration flows over time, $\{M_{ijt}\}_t$ and $\left\{\widehat{M}_{ijt}\right\}_t$. We assessed how well each model explains the migration flows from a country $i$ to a country $j$ by means of the $R^2$ between the modelled and observed time-series,

$$R_{ij}^2 \overset{\text{def}}{=} 1 - \frac{\sum_t \left(M_{ijt} - \widehat{M}_{ijt}\right)^2}{\sum_t \left(M_{ijt} - \overline{M}_{ij}\right)^2}, \quad (8)$$

where $\overline{M}_{ij}$ denotes the temporal mean of the observed migration flows from country $i$ to country $j$. A gravity model explains the flows from country $i$ to country $j$ very well if $R_{ij}^2 \approx 1$. If $R_{ij}^2 = 0$, then the modelled flows are constant and equal to the mean of the

historical flows between countries $i$ and $j$ across the study period. For $R_{ij}^2 < 0$, the model provides a worse estimate than the mean of the historical flows. We summarised the performances of our four models with respect to bilateral migration flows in terms of the distribution of the $R_{ij}^2$ across countries $i,j$.

Analogous to our assessment based on the spatio-temporally pooled rates of change in flows through time, we also examined how well the models explain the temporal dynamics of a given country pair by computing the $R^2$ between the relevant modelled and observed time-series,

$$\Delta R_{ij}^2 \overset{\text{def}}{=} 1 - \frac{\sum_t \left( \Delta M_{ijt} - \Delta \widehat{M}_{ijt} \right)^2}{\sum_t \left( \Delta M_{ijt} - \overline{\Delta M_{ij}} \right)^2}. \tag{9}$$

If $\Delta R_{ij}^2 = 0$, then the slope of the modelled data is constant and equal to the mean slope of the observed migration flows between countries $i$ and $j$ across the study period. Note that $\Delta R_{ij}^2$ (in contrast to $R_{ij}^2$ in Eq. (8)) is indifferent to whether the modelled and observed absolute flows are similar, as it only accounts for relative changes in the flow through time. If $\Delta R_{ij}^2 = 0$, then these changes are captured less well than by the mean of the observations, $\overline{\Delta M_{ij}}$.

**Comparison against bilateral models**. The role of the parameters $p_1,...,p_4$ in the Models 1 and 3 (Eqs. (1) and (3)) is to describe how changes in the four time-dependent predictor variables (population size and per-capita GDP in the country of origin and destination) affect migration flows. We can estimate the same parameters separately for each combination of country of origin $i$ and destination $j$, and compare these estimates to the ones we had obtained based on the spatio-temporally pooled data. This allows us to test whether the migration dynamics between any given two countries follow similar statistical rules as those established in the pooled analysis. For each pair $i,j$, we considered

$$\begin{aligned}
\log\left( M_{ijt} \right) \sim p_o^{(ij)} &+ p_1^{(ij)} \cdot \log\left( P_{it} \right) + p_2^{(ij)} \cdot \log\left( P_{jt} \right) \\
&+ p_3^{(ij)} \cdot \log\left( I_{it} \right) + p_4^{(ij)} \cdot \log\left( I_{jt} \right)
\end{aligned} \tag{10}$$

and estimated, for each $i,j$ with at least six data points available for calibration, the parameters $p_o^{(ij)}, \ldots, p_4^{(ij)}$ based on the annual observational data also used for Eqs. (1) and (3). We used a robust, bisquare weighting approach (Huber and Ronchetti, 2009) for the parameterisation, so as to minimise the sensitivity to potential outliers in the time-series that may confound signals (however, estimates were overall almost identical when using a standard least-squares approach instead). Distributions across all $i,j$ of the estimated parameters $p_1^{(ij)}, \ldots, p_4^{(ij)}$, associated with the four time-dependent predictors, were obtained after weighing each entry by the inverse of the corresponding standard error obtained from the regression, so as to give more weight to stronger signals (however, almost identical distributions were obtained without weighing entries). This allowed us to compare the distributions of the $p_1^{(ij)}, \ldots, p_4^{(ij)}$ to the specific values $p_1,...,p_4$ that we had previously estimated in Eqs. (1) and (3). The relatively short lengths of the country-pair-specific migration time-series $\{M_{ijt}\}_t$ in relation to the number of parameters in Eq. (10) mean that the distributions of the $p_n^{(ij)}$ need to be interpreted cautiously and should be used only to assess whether or not the coefficients estimated based on the spatio-temporally pooled data and those estimated based on country-pair-specific data are roughly in the same order of magnitude. For the same reason, we

did not conduct an analogous analysis for the larger set of time-dependent predictors in the Models 2 and 4.

## Results
All four gravity models matched the spatio-temporally pooled observed data well, explaining $R^2 = 62\%$, 68%, 94%, and 95% of the observed variance in the case of Models 1–4, respectively, based on the annual migration data (Fig. 1). $R^2$ values for the models calibrated based on the 5-year migration data were very similar ($R^2 = 59\%$, 64%, 95%, 96%, respectively; Supplementary Fig. 1). These values are comparable to those reported for gravity models in earlier studies (e.g., Afifi and Warner, 2008; Cohen et al., 2008; Beine and Parsons, 2015; Poot et al., 2016; Cai et al., 2016; Rikani and Schewe, 2021).

All models failed to capture temporal patterns in the observed migration flows. When evaluated in terms of the changes in migration flows over time, the parameterised models did not explain the observed annual data ($\Delta R^2 < 0$ in all cases; Fig. 2), generally providing worse estimates than even the constant value given by the mean rate of change in the observed flows across all countries and points in time. Whilst the modelled data did not match the observations well for any specific range of values, they in particular substantially underestimated the magnitude of both positive and negative deviations (Fig. 2). Results were again almost identical for the models calibrated based on 5-year migration flows (Supplementary Fig. 2).

The four models' strong performance in terms of the spatio-temporally pooled absolute migration flows in Fig. 1 did not extend to the flows between specific countries of origin $i$ and destination $j$. For the Models 1 and 2, calibrated on annual data, the $R^2$ between the relevant modelled and observed time-series, $R_{ij}^2$ (Eq. (8)), was negative for 87% and 93% of country indices $i,j$, respectively (Fig. 3a, b). In these scenarios, the mean of the observations provides a better fit than the modelled data. For Models 3 and 4, $R_{ij}^2$ was negative in 37% and 42% of cases, respectively (Fig. 3c, d); we return to this better performance of the fixed-effects models later on. Virtually identical results were obtained for the models calibrated based on 5-year migration flows (Supplementary Fig. 1). Figure 3e, f illustrates the discrepancy between the modelled and observed annual time-series for some examples.

Temporal patterns were captured overall even worse at the bilateral level than in the spatio-temporally pooled scenario in Fig. 2. For Models 1–4, calibrated on annual data, the $R^2$ between the relevant modelled and observed time-series, $\Delta R_{ij}^2$ (Eq. (9)), was negative for 72%, 82%, 75%, and 76% of country pairs $i,j$, respectively (Fig. 4). In these cases, the time-invariant average observed change in the migration flows from country $i$ to country $j$ provides a better estimate than the modelled data. Results were almost identical for the models calibrated based on 5-year migration data (Supplementary Fig. 2).

The regression coefficients of the four time-dependent predictors estimated based on the annual migration flows between specific pairs of countries (Eq. (10)) generally differed substantially from those estimated in Eqs. (1) and (3) based on the spatio-temporally pooled data (Fig. 5). For around half of all country pairs, the sign of the effect of a predictor variable estimated based on bilateral data was different from that estimated based on the spatio-temporally pooled data. In addition, the magnitude of the regression coefficients varied substantially between the different approaches.

## Discussion
**Gravity models describe spatial, not temporal, migration patterns**. All four gravity models considered here performed well
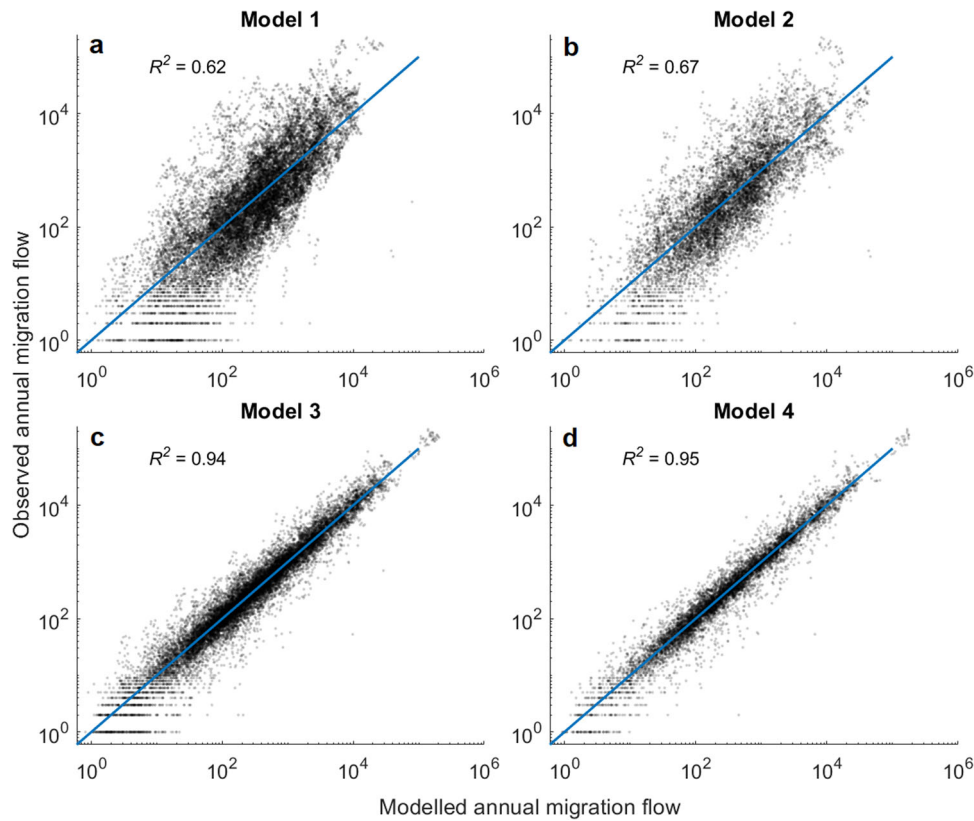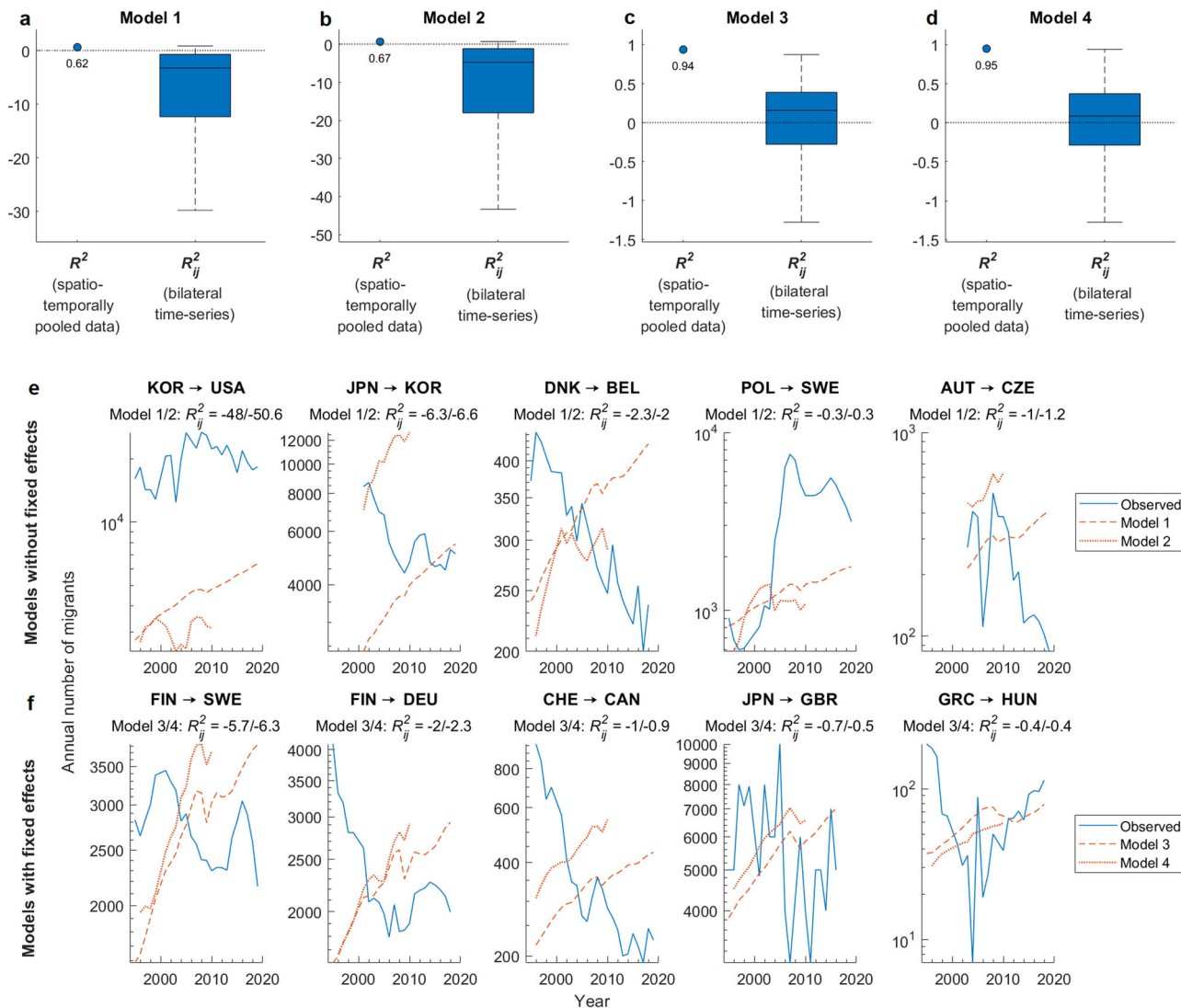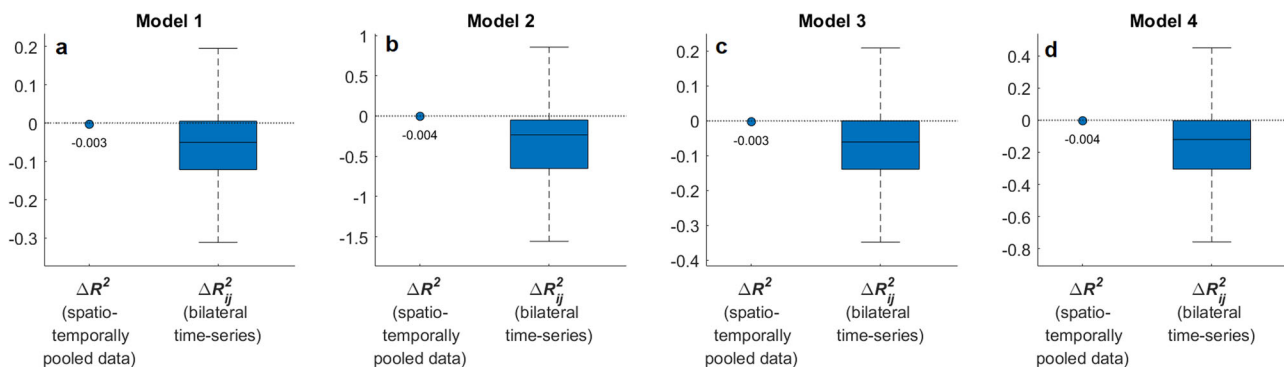
**Fig. 1 Modelled and observed migration flows.** Each marker represents the observed and modelled flow from a country of origin $i$ to a country of destination $j$ in a year $t$. Models 1–4 (**a**–**d**) correspond to Eq. (1)–(4), respectively. $R^2$ values correspond to Eq. (5). Blue lines represent 1:1 relationships.
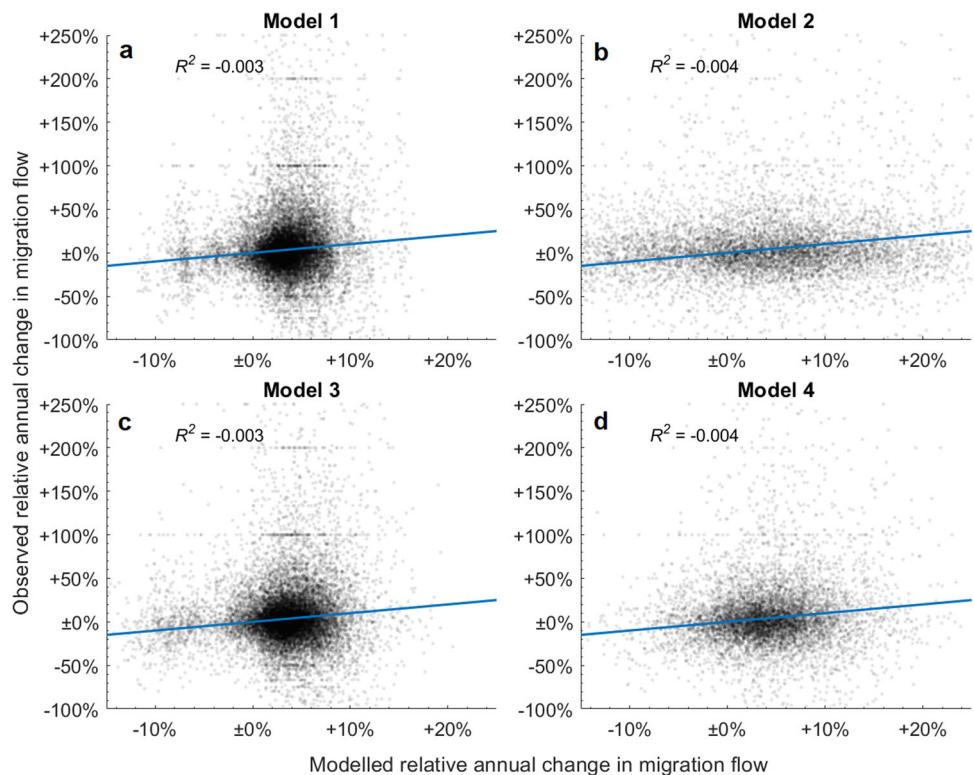


**Fig. 2 Modelled and observed temporal changes in migration flows.** Each marker represents the observed and modelled (based on Models 1–4; **a**–**d**) relative annual change in the number of migrants from a country of origin $i$ to a country of destination $j$ between a year $t$ and the following year $t+1$ (Eq. (6)). $R^2$ values correspond to Eq. (7). For visual clarity, axes were capped. Blue lines represent 1:1 relationships.

**Fig. 3 Modelled and observed migration flows at the bilateral level.** In panels **a**–**d**, the $R^2$ values represented by markers correspond to the comparisons shown in Fig. 1a–d, respectively. Boxplots (not showing outliers, which are as small as −663, −1.2 × 10^6, −7.7, and −89 for Models 1–4, respectively) represent the set of $R_{ij}^2$ values obtained from the comparison of modelled and observed annual data over time for each specific country of origin $i$ and destination $j$ (Eq. (8)). Five examples of these scenarios are shown for the models **e** without and **f** with fixed effects. Blue lines represent observed migration flows, orange lines represent modelled data. By design of the fixed-effects models, flows simulated by the Models 3 and 4 intersect the observed time-series at least once.



**Fig. 4 Modelled and observed temporal changes in migration flows at the bilateral level.** In panels **a**–**d**, the $\Delta R^2$ values represented by markers correspond to the comparisons shown in Fig. 2a–d, respectively. Boxplots (not showing outliers, which are as small as −11, −3.0 × 10^8, −12, and −3.1 × 10^8 for Models 1–4, respectively) represent the set of $R^2$ values obtained from the comparison of modelled and observed annual data over time for each specific combination of origin and destination countries (Eq. (9)).

when being compared to the spatio-temporally pooled data (Fig. 1), but poorly in terms of describing changes in migration over time (Figs. 2–4). This discrepancy is explained by the fact that spatial patterns vary over several orders of magnitude whereas temporal patterns do not, so that the structure of the spatio-temporally pooled data is in reality dominated by the spatial variation in migration flows, and it is this type of variation that is captured by the models. We can illustrate this point by considering slightly modified versions of Model 1 and of the two fixed-effects models, designed to map no temporal patterns of migration flows whatsoever, and by showing that these behave very similarly to the original models and perform equally well in terms of the evaluation in Fig. 1 and the $R^2$ in Eq. (5). Instead of Model 1, consider the following regression in which all temporal variation has been removed

$$
\log\!\left(\overline{M}_{ij}\right) \sim p_0 + p_1 \cdot \log\!\left(\overline{P}_i\right) + p_2 \cdot \log\!\left(\overline{P}_j\right)
$$
$$
+ p_3 \cdot \log\!\left(\overline{I}_i\right) + p_4 \cdot \log\!\left(\overline{I}_j\right) - p_5 \cdot \log\!\left(D_{ij}\right) \tag{11}
$$

where the bar notation denotes the mean over all available points in time. We parameterised Eq. (11) using annual migration and predictor data analogously to the approach used for Eq. (1), this time based on 970 data points representing temporal averages of the relevant predictand and predictor variables. Both the model's $R^2$ and the estimated regression coefficients are very similar to the ones previously obtained for Model 1 (Fig. 6a and Supplementary Table 1). In other words, whether the observed data include temporal variation or not has almost no effect on the resulting model; the much higher spatial variation dominates the parameterisation process and determines the model's $R^2$ in Eq. (5).

The strong performance of the fixed-effects Models 3 and 4 in Fig. 1c, d can also be achieved without describing temporal patterns. Consider the following model consisting exclusively of origin-destination-specific fixed effects,

$$
\log\!\left(M_{ijt}\right) \sim c_{ij}. \tag{12}
$$

We fitted this model based on the spatio-temporally pooled annual observations in the same way as Eq. (2). The model explained $R^2 = 93\%$ of the variation in the observed data (compared to $R^2 = 94\%$, 95% for Models 3, 4, respectively) (Fig. 6b), whilst, by design, simulating no temporal variation whatsoever. This demonstrates that the high $R^2$ of Models 3 and 4 in Fig. 1c, d is likewise due to fact that the models capture the spatial variation in the observed migration flows. Indeed, the fixed effects guarantee this property, and therefore a high $R^2$, a priori, i.e., independently of the time-dependent elements of the regression.
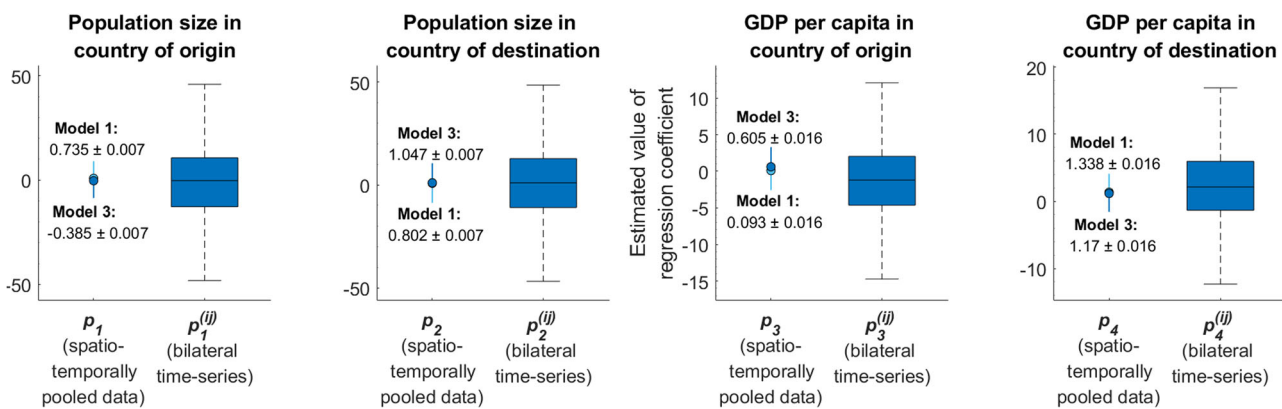


Fig. 5 Regression coefficients based on spatio-temporally pooled data and bilateral time-series. Light and dark blue markers represent the regression coefficients in Models 1 and 3 (Eqs. (1) and (3)), respectively, estimated based on the spatio-temporally pooled annual observations (cf. Supplementary Table 1). Boxplots represent the distributions of the analogous coefficients estimated based on the migration time-series between specific countries of origin and destination (Eq. (10)). Outliers, ranging between [−3730, 5746] for $p_1$, [−9799, 5363] for $p_2$, [−336, 386] for $p_3$, and [−260, 1164] for $p_4$, are not shown.
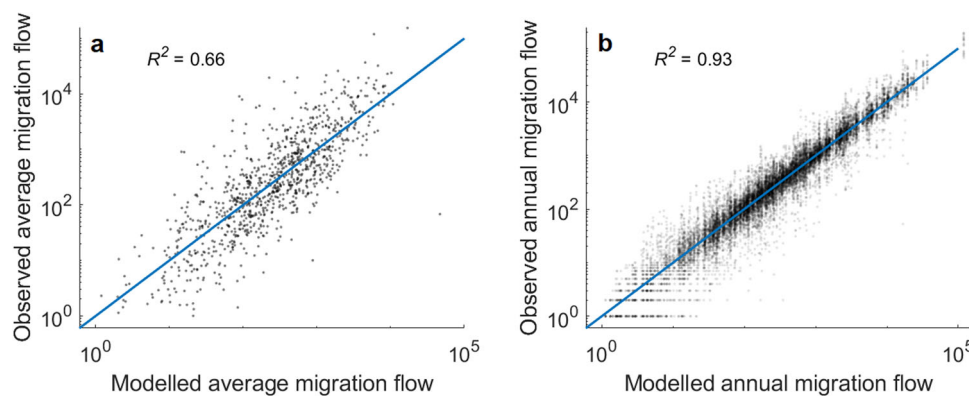


Fig. 6 Spatial patterns dominate gravity model behaviour. a Comparison of modelled and observed time-averaged migration flows. Each marker represents the mean flow from a country of origin $i$ to a country of destination $j$ across all available years. b Comparison of observed annual migration flows against a time-invariant, fixed-effects-only model. The lack of temporal variation in the modelled data is visible in the form of vertical patterns, representing country pairs for which the modelled flow is constant over time. Blue lines represent 1:1 relationships.

Origin-destination-specific fixed effects also explain the overall better performances of the Models 3 and 4 in terms of $R_{ij}^2$, compared to those of Models 1 and 2 (Fig. 2a–d). For the model in Eq. (12), $R_{ij}^2$ would be 0 for all $i,j$ by design. Unlike for Models 1 and 2, we would therefore expect $R_{ij}^2$ to be non-negative for at least half of all country pairs $i,j$ in the cases of Models 3 and 4.

**Temporal migration patterns of are more complex than spatial patterns**. Temporal dynamics not only follow different statistical rules than spatial migration patterns—shown by the discrepancy in our models' performances with respect to these two aspects—but more complex ones. For example, Model 1 (and the related model in Eq. (11)) imply that, other things being equal, a higher population in a country of origin is associated with a larger migration flow from that country. This intuitive relationship would be fully expected across the several orders of magnitude covered by national population sizes. However, for any specific country, in which population size over time does not vary across orders of magnitude, a change in population size need not generally result in a corresponding change in out-migration; instead, the latter is affected by many other factors that may easily outweigh the effect of changes in population size. This complexity of temporal migration dynamics is illustrated in Fig. 5, highlighting the non-uniformity of the responses of bilateral flow time-series to the four time-dependent predictor variables. Whilst a fixed set of coefficients for these predictors is sufficient to explain a substantial proportion of the spatial variation in migration (as seen in Eq. (11)), a set of fixed values is not able to accommodate the temporal variation.

The poor performances of the fixed-effects Models 3 and 4 in Figs. 2c, d and 4c, d further illustrate the above point. By design, the models' fixed effects already fully capture the spatial variation in the observed flows; thus, the coefficients in Eqs. (3) and (4) corresponding to the time-dependent predictors can in principle be selected to specifically describe the temporal variation. However, this does not succeed either: both on the spatio-temporally pooled level (Fig. 2c, d) and at the bilateral level for the large majority of country pairs (Fig. 4c, d), Models 3 and 4 fail to capture temporal patterns. If the migration responses to changes in population size or per-capita GDP in the country of origin or destination were consistent across country pairs, then the models would perform well in all evaluations in our analysis. The fact that responses are not consistent explains the models' poor performances.

**Gravity models cannot predict future migration**. In principle, available forecasts of country-specific population sizes $P_{it}$ (e.g., UN, 2019) and per-capita GDPs $I_{it}$ (e.g., The World Bank, 2020) could readily be applied to the parameterised regressions in Eqs. (1) and (2) to generate predictions of future international migration. However, our results demonstrate that such predictions would not be statistically supported, and most likely represent worse temporal extrapolations than even the (time-invariant) mean of the historical flows. Understanding how flows vary across time in response to changes in relevant predictor variables represents a fundamental requirement for predicting future international migration. Our analysis shows that gravity models generally do not possess this quality; their ability to explain the variation in flows across countries well is of no avail when it comes to predicting future migration.

Our result that spatial and temporal migration trends do not coincide on the 25-year long period of the observational record (neither for an annual, nor a 5-year, temporal resolution) does not allow us to exclude the possibility that they converge if the temporal scale is chosen sufficiently large, e.g., in the order

of centuries (cf. Clemens, 2014). However, not only would quantitative forecasts of demographic, economic, social, political, or environmental variables considered as relevant predictors of migration be subject to extreme uncertainties at these time scales, but socio-political interest in very-long-term migration scenarios will also be limited. With regard to the nearer future, we argue that the lack of alignment between temporal and spatial migration patterns over the past quarter century strongly challenges the assumption that an alignment will emerge in the coming years or decades, which would be needed in order to justify the use of gravity models for predicting future flows.

Our analysis cannot formally rule out the possibility that gravity models based on additional, or different, predictor variables than the ones in our approach can succeed at describing temporal migration dynamics. However, given that both the simple and the more complex models exhibit identical behaviour in terms of mapping temporal patterns, we see little evidence to expect the results of the model assessments in Figs. 2–4 to be fundamentally different for other gravity models parameterised on spatio-temporally pooled global data. Population size and per-capita GDP in the countries of origin and destination, in particular, have long been cited as four of the most relevant time-dependent drivers of international migration. The fact that migration responses to these established variables differ so substantially at the bilateral level from those estimated in the spatio-temporally pooled case (Fig. 5) leads us to expect similar patterns for other predictors.

Validations in previous studies using gravity models typically do not go beyond comparisons of modelled and observed flows at the spatio-temporally pooled level (analogous to our Fig. 1), making it difficult to readily assess the models' ability to map temporal trends. Here, we mention two case examples of gravity-type approaches used in temporal contexts, in which additional plots hint at the limitations highlighted in our analysis. Haag et al. (1988) model regional utility levels, which are later used to map internal migration flows, as a function of socio-economic predictor variables. Whilst the model performs very well at the spatio-temporally pooled level, this is largely due to the fact that it describes spatial patterns, whereas temporal trends are generally not captured as well (their Fig. 4.26). The nonlinear gravity model in Rikani and Schewe (2021), used to project future global migration trajectories matches spatio-temporally pooled observed flows well; however, plots of country-level migration (their Figs. 5 and 6) show that modelled and observed migration time-series differ considerably. The time between these two publications also illustrates the persistence of the issue.

## Conclusions

Our results demonstrate that a successful calibration of gravity models on the basis of spatio-temporally pooled observed migration flows does not imply, or obviate the need to validate, a model's ability to describe flows between specific pairs of countries in general and changes in migration over time in particular. This is critical as this latter ability represents a fundamental requirement both for understanding past migration dynamics and for generating statistically supported future predictions. The fact that the example gravity models considered in our analysis represented temporal dynamics overall worse than even the time-invariant mean of the historical observations provides strong evidence that predictions of international migration flows based on these models would most likely be highly unreliable. More generally, our results put into question suggestions made based on gravity models in earlier studies regarding the effects that certain economic, social, climatic, or other factors may have on future international migration.

Our analysis reveals that gravity models with and without origin–destination-specific fixed effects fail to capture temporal migration patterns for different, albeit related, reasons. The dominance of the spatial, as opposed to temporal, variation in the observed flow data means that parameters in gravity models without fixed effects, such as our Models 1 and 2, will be selected to match spatial, not temporal, patterns; and these two do not coincide. In principle, fixed effects can compensate for this behaviour by directly capturing spatial patterns, so that the time-dependent predictors in the regression can target temporal patterns; however, this approach is not successful either due to the higher complexity of temporal migration dynamics. A property shared by the two categories of gravity models is that the apparently strong performance in explaining the variation in the spatio-temporally pooled observational data is misleading as only the spatial variation is accurately captured.

Our results demonstrate that gravity models used to seek insights into past or future migration dynamics require first and foremost better evaluation frameworks that allow for a rigorous assessment of the models' ability to map bilateral flows in general and temporal variations in particular. In order for models to perform well in these regards, different statistical approaches that accommodate the higher complexity of temporal, compared to spatial, migration patterns will most likely be necessary. Country-pair-specific time-series of historical migration flows, in particular, provide the most valuable source of information on the responses of migration flows over time to changes in relevant predictors. By pooling observational data across space and time when estimating model parameters, many existing gravity models do not exploit this resource effectively. The lengths of such time-series are currently very limited, especially at the global scale; grouping countries appropriately may allow for circumventing this limitation, and for inferring robust statistical signals. Finally, whilst a set of fixed parameter values may suffice to explain the spatial variation in migration well, a potentially higher degree of stochasticity in the temporal variation may motivate to represent model parameters in terms of probability distributions that can be estimated using Bayesian approaches and may provide an effective way for handling variations in migration flows that are beyond deterministic responses (Bijak, 2011; Azose and Raftery, 2015).

Recent gravity models generally represent statistical rather than mechanistic approaches to describing migration. Whilst the 'original' gravity model may be interpreted mechanistically in analogy to Newton's law of gravitation (Poot et al., 2016), the inclusion of additional terms representing social, political, environmental, or other factors in the regression (like our Eqs. (2) and (4)) has largely removed gravity models from a mechanistic basis, making plausible extrapolation in time all the more challenging. Efforts to design the mathematical structure of regression-based migration models using linear or nonlinear mechanistic processes, which can ideally be independently verified prior to model calibration (cf., for example, the decomposition of the regression formula into separate components related to mobility, origin–destination utility differential, and population sizes in Helbing (2010)), may prove valuable for improving our ability to map international migration dynamics.

## Data availability

Data and code associated with this study are available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/E7NUA).

## References

Abel GJ (2018) Estimates of global bilateral migration flows by gender between 1960 and 2015. Int Migr Rev 52:809–852

Abel GJ, Cohen JE (2019) Bilateral international migration flow estimates for 200 countries. Sci Data 6:82

Adserà A, Pytliková M (2015) The role of language in shaping international migration. Econ J 125:F49–F81

Afifi T, Warner K (2008) The impact of environmental degradation on migration flows across countries. UNU-EHS Working Papers. UNU-EHS

Azose JJ, Raftery AE (2015) Bayesian Probabilistic Projection of International Migration. Demography 52:1627–1650

Backhaus A, Martinez-Zarzoso I, Muris C (2015) Do climate variations explain bilateral migration? A gravity model analysis. IZA J Migr 4:3

Beine M, Parsons C (2015) Climatic factors as determinants of international migration. Scand J Econ 117:723–767

Beine M, Bertoli S, Fernández-Huertas Moraga J (2016) A practitioners' guide to gravity models of international migration. World Econ 39:496–512

Belot M, Ederveen S (2012) Cultural barriers in migration between OECD countries. J Popul Econ 25:1077–1105

Benveniste H, Oppenheimer M, Fleurbaey M (2020) Effect of border policy on exposure and vulnerability to climate change. Proc Natl Acad Sci USA 117:26692–26702

Bijak J (2011) Forecasting International Migration in Europe: A Bayesian View. Springer

Blinder S, Allen W. (2020) UK public opinion toward immigration: Overall attitudes and level of concern. The Migration Observatory. https://migrationobservatory.ox.ac.uk/resources/briefings/uk-public-opinion-toward-immigration-overall-attitudes-and-level-of-concern/

Cai R, Feng S, Oppenheimer M, Pytlikova M (2016) Climate variability and international migration: the importance of the agricultural linkage. J Environ Econ Manag 79:135–151

Cattaneo C, Peri G (2016) The migration response to increasing temperatures. J Dev Econ 122:127–146

Clemens MA (2014) Does development reduce migration? In: Lucas REB (ed) International handbook on migration and economic development. Edward Elgar Publishing, pp. 152–185

Cohen JE, Roig M, Reuman DC, GoGwilt C (2008) International migration beyond gravity: A statistical model for use in population projections. Proceedings of the National Academy of Sciences 105(40):15269–15274. https://doi.org/10.1073/pnas.0808185105

Docquier F (2018) Long-term trends in international migration: lessons from macroeconomic model. Econ Bus Rev 4:3–15

Economist Intelligence Unit (2019) Political stability and absence of violence. www.eiu.com.

Gröschl J, Steinwachs T (2017) Do natural hazards cause international migration? CESifo Econ Stud 63:445–480

Haag G, Munz M, Reiner R, Weidlich W (1988) Federal Republic of Germany. In: Weidlich W, Haag G (eds) Interregional migration: dynamic theory and comparative analysis. Springer, pp. 65–100

Helbing D (2010) Dynamic decision behavior. In: Helbing D (ed) Quantitative sociodynamics: stochastic methods and models of social interaction processes. Springer, pp. 17–45

Helbling M, Meierrieks D (2021) How climate change leads to emigration: conditional and long-run effects. Rev Dev Econ 25:2323–2349

Helbling M, Bjerre L, Römer F, Zobel M (2017) Measuring immigration policies: the IMPIC database. Eur Polit Sci 16:79–98

Huber PJ, Ronchetti EM (2009) Robust statistics. Wiley

James SL, Gubbins P, Murray CJ, Gakidou E (2012) Developing a comprehensive time series of GDP per capita for 210 countries from 1950 to 2015. Popul Health Metr 10:12

Kapur D (2014) Political effects of international migration. Annu Rev Polit Sci 17:479–502

Lutz W, Goujon A, Kc S, Stonawski M, Stilianakis N (2018) Demographic and human capital scenarios for the 21st century: 2018 assessment for 201 countries. Publications Office of the European Union

Martin PL (2017) President Trump and US migration after 100 days. Migr Lett 14:319–328

Maurel M, Tuccio M (2016) Climate instability, urbanisation and international migration. J Dev Stud 52:735–752

Nicolli F, Bettin G (2012) Does climate change foster emigration from less developed countries? Evidence from bilateral data. Working Papers. University of Ferrara, Department of Economics

OECD (2021) OECD international migration database. https://stats.oecd.org/Index.aspx?DataSetCode=MIG

Özden Ç, Parsons CR, Schiff M, Walmsley TL (2011) Where on earth is everybody? World Bank Econ Rev 25:12–56

Papademetriou DG, Banulescu-Bogdan N (2016) Understanding and addressing public anxiety about immigration. Migration Policy Institute

Poot J, Alimi O, Cameron MP, Maré DC (2016) The gravity model of migration: the successful comeback of an ageing superstar in regional science. IZA Discussion paper

Ragazzi P (2012) Climate change and migration: a gravity model approach. University of Ferrara, Department of Economics

Ramos R (2016) Gravity models: a tool for migration analysis. IZA World Labor

Reuveny R, Moore WH (2009) Does environmental degradation influence migration? Emigration to developed countries in the late 1980s and 1990s. Soc Sci Q 90:461–479

Rikani A, Schewe J (2021) Global bilateral migration projections accounting for diasporas, transit and return flows, and poverty constraints. Demogr Res 45:87–140

The World Bank (2020) Global economic prospects, June 2020. World Bank. https://www.worldbank.org/en/publication/global-economic-prospects

The World Bank (2021) World Bank open data. https://data.worldbank.org/

Trauner F, Turton J (2017) "Welcome culture": the emergence and transformation of a public debate on migration. Österr Z Polit 46:33–42

UN (2019) United Nations, Department of Economic and Social Affairs, Population Division: world population prospects 2019. https://population.un.org/wpp/Publications/

UNDP (2021a) United Nations development programme: expected years of schooling (years). http://hdr.undp.org/en/indicators/69706

UNDP (2021b) United Nations development programme: life expectancy at birth (years). http://hdr.undp.org/en/indicators/69206

## Acknowledgements

## Author contributions
RMB designed the project, conducted the analysis, and drafted the manuscript. RMB, JS and HL-C interpreted the results and revised the manuscript.

## Funding

## Competing interests
The authors declare no competing interests.

## Ethical approval
Ethical approval was not applicable.

## Informed consent
Informed consent was not applicable.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1057/s41599-022-01067-x.

**Correspondence** and requests for materials should be addressed to Robert M. Beyer.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.