Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Forecasts Reveal a Concentration Paradox in UK International Student Mobility

Ruth Neville

ruth.neville@ucl.ac.uk

University College London, Centre for Advanced Spatial Analysis

Francisco Rowe

University of Liverpool https://orcid.org/0000-0003-4137-0246

Emilio Zagheni

Max Planck Institute for Demographic Research https://orcid.org/0000-0002-7660-8368

Article

Keywords:

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8337360/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Machine Learning Forecasts Reveal a Concentration Paradox in UK International Student Mobility

Ruth Neville[1], Francisco Rowe[2], Emilio Zagheni[3]

5th June 2026

[1] Centre for Advanced Spatial Analysis (CASA), University College London, London, UK
[2] Department of Geography and Planning, University of Liverpool, Liverpool, UK
[3]Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany
*Corresponding author: ruth.neville@ucl.ac.uk

Abstract

International student mobility (ISM) underpins the financial resilience, soft power, and knowledge production of major destination countries, yet its future trajectory is increasingly uncertain. In the UK, recent policy shifts and geopolitical tensions have coincided with a slowdown in international student growth and mounting concerns over sectoral stability. We use administrative data from the Universities and Colleges Admissions Service (UCAS) covering successful undergraduate applications from 86 origin countries between 2010 and 2024, linked with economic and demographic projections, to forecast international applications to the UK through to 2030. A Poisson-based XGBoost model is trained and evaluated against ARIMA and negative binomial gravity benchmarks on an out-of-sample period spanning Brexit and the COVID-19 pandemic. The machine learning model delivers consistently lower forecast errors and stronger correspondence with observed flows, but gains are modest in the presence of structural shocks. Forecasts indicate that applications from Mainland China and Hong Kong, India, and the European Union will together comprise over 60% of international demand by 2030, despite stagnation or decline in absolute numbers from all three blocs. We term this configuration a 'concentration paradox', in which reliance on a narrowing set of origins intensifies even as inflows soften, increasing the vulnerability of the UK higher education system to policy and geopolitical disruption.

1 Introduction

International student mobility (ISM) is a major component of both high-skilled mobility and the international knowledge economy. With long-term implications for both origin and destination countries, ISM contributes to the development of human capital, supports brain circulation, and generates economic, social, and cultural exchanges. In 2022, the global number of internationally mobile students was estimated at 6.9 million – with a 176% increase since 2002 (UNESCO Institute for Statistics (UIS), 2024). International students bring knowledge, skills, and social capital across borders, strengthening academic and research capacity of host institutions while maintaining transnational connections to origin countries (Saxenian, 2005; Knight, 2012; Jon, 2013). However, benefits are often unevenly distributed, with high-income countries positioned as the primary recipients and lower-income countries often facing concerns about brain drain (Docquier and Rapoport, 2012; Robertson, 2013; Czaika and Parsons, 2017).

International students have increasingly become an instrument of soft power, allowing states to project cultural and political influence through academic exchange (Altbach and Knight, 2007). This underscores the strategic importance of international students as a pipeline for high-skilled migrants, prompting many high-income countries to integrate student recruitment more explicitly into education, diplomacy, and migration policy frameworks (Lomer, 2017). ISM is often characterised as a form of transitional mobility, occupying a space between temporary and permanent migration; an estimated 15–35% of international students remain in their host country after graduation, complicating simplistic binaries between short-term and long-term migration (She and Wotherspoon, 2013). Consequently, student mobility has become increasingly entangled in political debates over immigration control (Robertson, 2013; Migration Observatory, 2025), creating tensions between efforts to expand

international student recruitment and the broader trend toward restrictive migration policies across advanced economies.

The geography of ISM also reflects broader inequalities. While much attention focuses on the total volume of international students, the composition of flows and who are supplying them carries consequences for institutional dependence, political relationships, and exposure to external shocks. ISM exhibits a similar relationship to developments as other forms of migration, in line with migration transition theory (de Haas, 2010). Weber and van Mol (2023) extend this framework to student mobility, demonstrating that ISM follows a U-shaped curve in which countries initially send more students abroad as they develop, but eventually become attractive destinations and begin to attract students. High-income Anglophone countries, such as the United States, the United Kingdom, and Australia, have therefore tended to be dominant hosts, and the Global North is generally perceived as the standard assumed referent of international student flows (Spivak, 2009). However, some criticise this perception, with King and Sondhi (2018) suggesting these assumptions are built on histories of colonisation, and are an over-simplification given evidence of significant representation of 'high-income' countries amongst cohorts of international students in the UK system.

Within this global system, the UK is a central actor. It is the second-leading destination of international students, attracting around 10–14% of the global market share in recent years, and international students comprise roughly a quarter of the UK student population (Cuibus and Walsh, 2024). Recent research from Jisc (2025) shows that the UK has educated more world leaders than anywhere else in the world, having educated 50 heads of state between 1990 and 2020. International students are crucial to the financial sustainability of UK higher education institutions (HEIs), to the economy more broadly, and to the UK's soft power position. Recent estimates suggest that international students generated a net economic benefit of over £35 billion in 2021/22, while tuition fee income from international students has become increasingly important for cross-subsidising teaching and research (Bolton et al., 2023). Given this reliance, understanding future trajectories of international student demand is critical for institutional and national policy planning.

Concurrently, recent geopolitical shifts have introduced new uncertainties, challenging the stability of ISM to the UK. The UK's withdrawal from the European Union (EU) in 2021 structurally altered the conditions under which EU students access UK higher education. EU applicants lost home-fee status and access to the student loans system, facing higher tuition fees, visa requirements, and more restricted post-study work options (Bolton and Lewis, 2021). Applications from the EU subsequently collapsed, with a decline of around two-thirds recorded in 2021/22, as we have explored in previous work (Neville et al., 2024b) and as further evidenced by other studies (Amuedo-Dorantes and Romiti, 2024). Meanwhile, applications from Mainland China, the UK's largest source of students and the largest global sender (UNESCO Institute for Statistics, 2023), have shown signs of stagnation and decline, potentially amid changing geopolitical relations and expanding domestic HE capacity. Policy tightening under successive UK governments, including restrictions on dependents and proposals to curtail post-study work visas (UK Home Office, 2023a; Universities UK, 2024), has further raised concerns about the UK's attractiveness as a study destination. 2024 saw reductions in sponsored study visa applications and forecasts of financial deficits in many UK institutions (Lapworth, 2024).

Reliable ISM forecasts are therefore essential for institutional resilience and sector-wide stability. Yet, forecasting human mobility is inherently complex. Migration flows respond to an oscillating interplay of socio-political, economic, and institutional factors that are sensitive to external shocks (National Research Council, 2000). Recent sector-wide projections overestimated international student growth. In 2024, undergraduate applicant growth was forecasted at 5.8%, while actual growth was only 1.3%, leaving many UK institutions facing shortfalls, leading to misallocation of resources and unanticipated financial pressures (Office for Students, 2024). These discrepancies highlight the limitations of current forecasting approaches and the need for more robust, adaptive methods.

Existing forecasting of ISM in the UK is completed using time-series and gravity-type models developed by sector bodies and consultancy firms. Time-series models, such as ARIMA, extrapolate past trends and are widely used due to their simplicity and interpretability (Bijak et al., 2019; Azose and Raftery, 2015). Gravity models, grounded in migration theory, incorporate economic, demographic, and geographic factors and are well established in modelling human and student mobility (Neville et al., 2024a; Weber and van Mol, 2023; Cabanas-Tirapu et al., 2025). More recently, machine learning techniques such as XGBoost have been explored for forecasting student enrolments and have shown some promise (Litmeyer and Hennemann, 2024; Yang et al., 2020). However, academic work applying

such methods to ISM, especially in the UK context, remains limited, and comparative evaluations of traditional and machine learning approaches are still rare.

In this paper, we leverage Universities and Colleges Admissions Service (UCAS) administrative data on successful undergraduate applications from 86 origin countries between 2010 and 2024 to produce disaggregated forecasts of undergraduate international student mobility to the UK through to 2030. We employ an XGBoost model tailored to count data to forecast applications at country level, integrating historical application counts with origin- and destination-specific characteristics. The model's performance is assessed against traditional benchmarks, including ARIMA and gravity-type models, before being used to produce forward-looking forecasts. Substantively, we focus on the emerging risk of an increasingly concentrated system, in which applications from key source countries plateau or decline while their market share, and the UK's dependence on them, continues to grow. We term this a 'concentration paradox' and interpret it as an indicator of systematic vulnerability, where dependence grows even as inflows soften, increasing exposure to policy and geopolitical shocks. A detailed description of data sources, preprocessing steps, and covariate projections used for forecasting is provided in the Supplementary Information.

We address the following research questions:

1. How accurately can machine learning methods predict recent trends in undergraduate international student applications to the UK, relative to traditional forecasting approaches?
2. What future trajectories are projected for undergraduate applications from key countries and regions of origin to the UK between 2025 and 2030?
3. How do these trajectories reshape the concentration of undergraduate UK international student demand?

The remainder of the paper is structured as follows. Section 2 outlines existing evidence on trends, drivers, and forecasting of international student mobility. Section 3 presents the results, focusing on projected trajectories for major origins and the changing composition of demand. Section 4 discusses the implications for UK higher education and migration policy, and Section 5 concludes with reflections on system vulnerability and future research directions. Section 6 describes the data and methods used, including the UCAS data, XGBoost framework, benchmark models, and evaluation strategy.

2 Background

2.1 Global Trends and the UK as a Case Study

ISM has intensified with globalisation, as countries compete to attract international talent (Altbach and Knight, 2007). Key destinations such as the UK have traditionally implemented policies to strategically expand international enrolments, shaping a competitive market for global talent (Lomer, 2017). English-speaking destinations such as the United States, the UK, and Australia have emerged as dominant hosts, supported by the global reach of English language provision, histories of colonialism, and cultural linkage (Altbach and Knight, 2007; OECD, 2023; Cuibus and Walsh, 2024). According to 2023 OECD data, these three countries collectively host 42% of international students, within which the UK hosts ~14%. Mainland China and India have traditionally been the primary sending countries, constituting 21% and 10% of all international students in the OECD in 2021 respectively (OECD, 2023).

The UK is an important case to consider because it combines high international student dependence with a period of intense policy and geopolitical disruption. UCAS (2019) data show that, prior to Brexit, EU countries contributed the largest share of international undergraduates, while Mainland China rapidly expanded its share to become the single largest source. Post-Brexit, applications from the EU have fallen sharply, and growth from Mainland China has slowed (Migration Observatory, 2025). Concurrently, under the 2021 Conservative government, the UK government's International Education Strategy identified new priority markets, including India, Indonesia, Saudi Arabia, Vietnam, and Nigeria, with explicit targets to increase international exports by 2030 (Department of Education and Department for International Trade, 2021). Although these shifts suggest a reconfiguration of the UK's international student base, the policy landscape is one of continuous flux. Recent debates

have pivoted toward curbing international student numbers, with policy discourse increasingly focused on preventing 'non-genuine' applications and tightening post-study opportunities (UK Home Office, 2024).

2.2 Drivers and Theories of International Student Mobility

A large body of work has identified the drivers of ISM in terms of 'push' and 'pull' factors. Notable 'push' factors that influence flows of internationally mobile students are often related to the relative opportunities available in the country of origin compared to studying overseas. As highlighted earlier, richer countries have tended to be more attractive destinations (Wei, 2013; Caruso and De Wit, 2014). Elsewhere we have illustrated how, within the context of the UK, countries with higher levels of GDP per capita tend to send more international students, but the influence of this varies (Neville et al., 2024a),

Network and systems approaches emphasise how historic connections and social ties shape these flows. Colonial linkages, common language, and pre-existing migrant communities reduce barriers to mobility and lower informational and financial costs (Massey et al., 1993; Beine et al., 2014; Neville et al., 2024a). Common language, particularly English, serves as a key factor in ISM, with English proficiency significantly facilitating entry into UK HEIs (Lewer and Van den Berg, 2008).

Human development dynamics also play a role. Migration transition theory argues that migration intensity initially increases with development, as improved resources, capabilities, and aspirations facilitate mobility, before stabilising or declining as countries become attractive destinations themselves (de Haas, 2010). In the context of ISM, Weber and van Mol (2023) show that as countries develop, they transition from sending to increasingly hosting students. This suggests that rapidly developing countries in South Asia and sub-Saharan Africa may experience rising outbound ISM in the medium term, but may also invest in domestic HE capacity that dampens overseas demand.

These frameworks highlight the complexity of ISM drivers: student flows are shaped by intertwined economic, social, political, and institutional forces, which can shift rapidly in response to shocks or policy changes. Forecasting ISM therefore requires tools that can accommodate nonlinearities and interactions across multiple dimensions. While existing theories explain why international student mobility changes, far less work has been able to anticipate how and when these changes occur.

2.3 Forecasting International Student Mobility

Forecasting ISM, and migration more broadly, is difficult as it is among the most volatile demographic processes (National Research Council, 2000). Disruptive events such as economic crises, conflicts, pandemics, and sudden policy changes can produce abrupt shifts in flows and alter long-standing corridors (Disney et al., 2015). Yet, there is strong demand for forecasts to support capacity planning, budgeting, and policy design in higher education systems.

In the UK, a number of institution and sector-level forecasts have been produced, often using time-series models such as ARIMA or gravity-type models that extrapolate from historical patterns (British Council, 2024). While these tools have proved attractive due to their simplicity and grounding in migration theory, projections have proved overly-optimistic. Over-optimistic projections through the early 2020s have contributed to under-anticipated financial deficits when actual growth in international students slowed (Office for Students, 2024).

Academic research on forecasting ISM remains relatively sparse, and even more so with regards to the employment of machine learning. However, there is some work emerging. (Litmeyer and Hennemann, 2024) use explainable machine learning techniques, including XGBoost, to forecast first-year student enrolments at German universities and show that such models can outperform gravity models on predictive accuracy while also offering insight into variable importance. Yang et al. (2020) apply machine learning to forecast outbound student mobility from Taiwan, finding superior performance relative to traditional ARIMA and vector autoregression models. However, very few studies combine administrative data with machine learning for ISM forecasting in a major destination like the UK, and fewer still compare their performance to the tools currently used in policy and industry.

3 Results

3.1 Model Performance

Across the out-of-sample evaluation period (2019–2023), the XGBoost model demonstrates higher predictive accuracy than the ARIMA and negative binomial gravity benchmarks. Summary performance statistics across the three approaches are presented in Supplementary Table 3, and country-level out-of-sample predictions for 2019–2023 are shown in Figure 1.

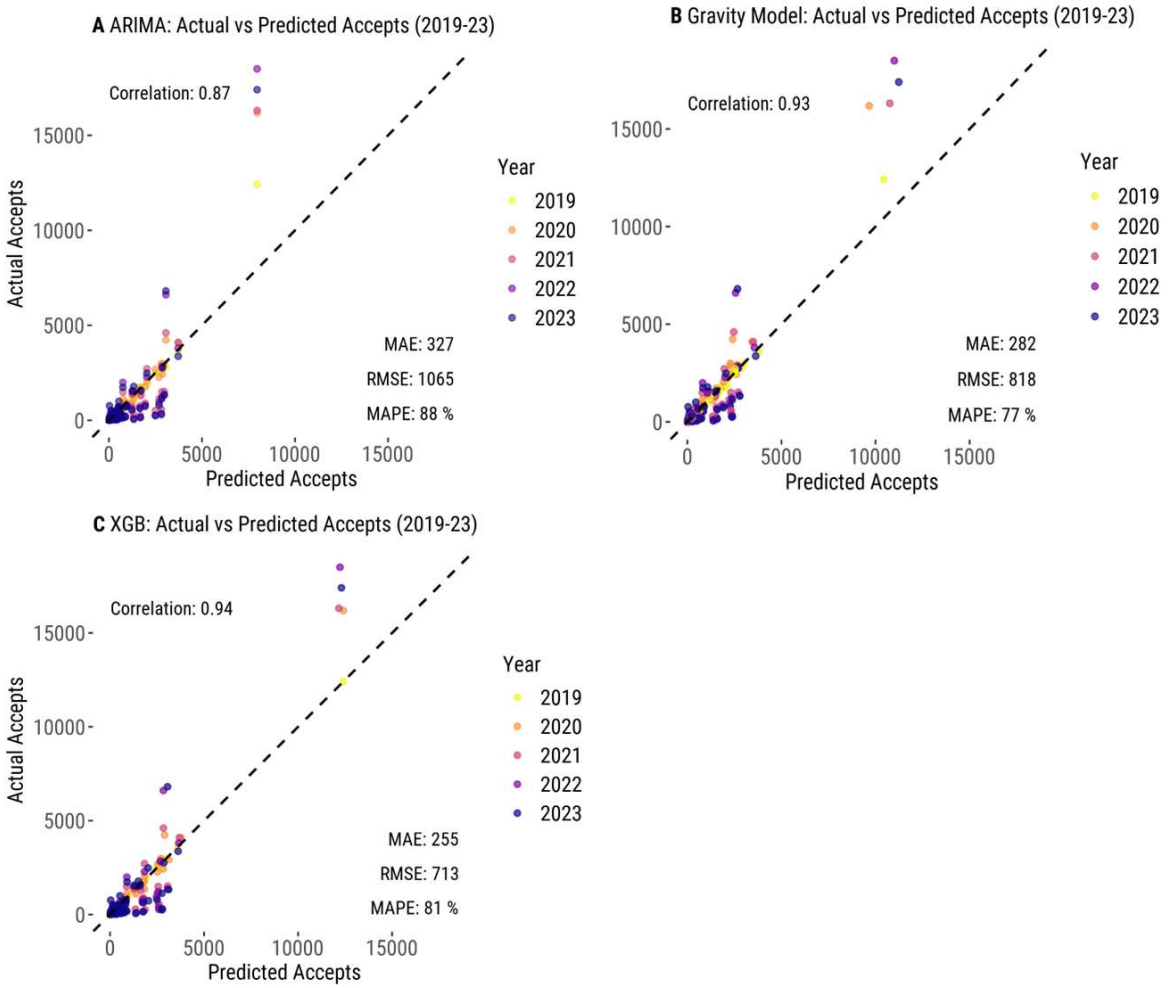

Figure 1: Scatterplots compare predicted versus observed accepted undergraduate applications across origin countries and evaluation years for (A) ARIMA, (B) negative binomial gravity, and (C) XGBoost models. Each point represents an origin country–year observation. Dashed lines indicate the 45-degree line corresponding to perfect prediction. XGBoost achieves the highest correlation and lowest prediction error among the evaluated methods.

As shown in Figure 1 and Supplementary Table 3, the XGBoost model achieves lower MAE and RMSE and stronger correlations with observed application counts, indicating improved capacity to capture non-linear interactions between economic, demographic, and structural drivers. However, these improvements are incremental rather than transformative.

Although consistent, these performance gains remain modest. Predictive accuracy declines during periods affected by major policy developments, suggesting that uncertainty stems not only from methodological constraints but from structural characteristics of the system itself. This pattern is further illustrated in Supplementary Fig. 1, where deviations between predicted and observed values are most

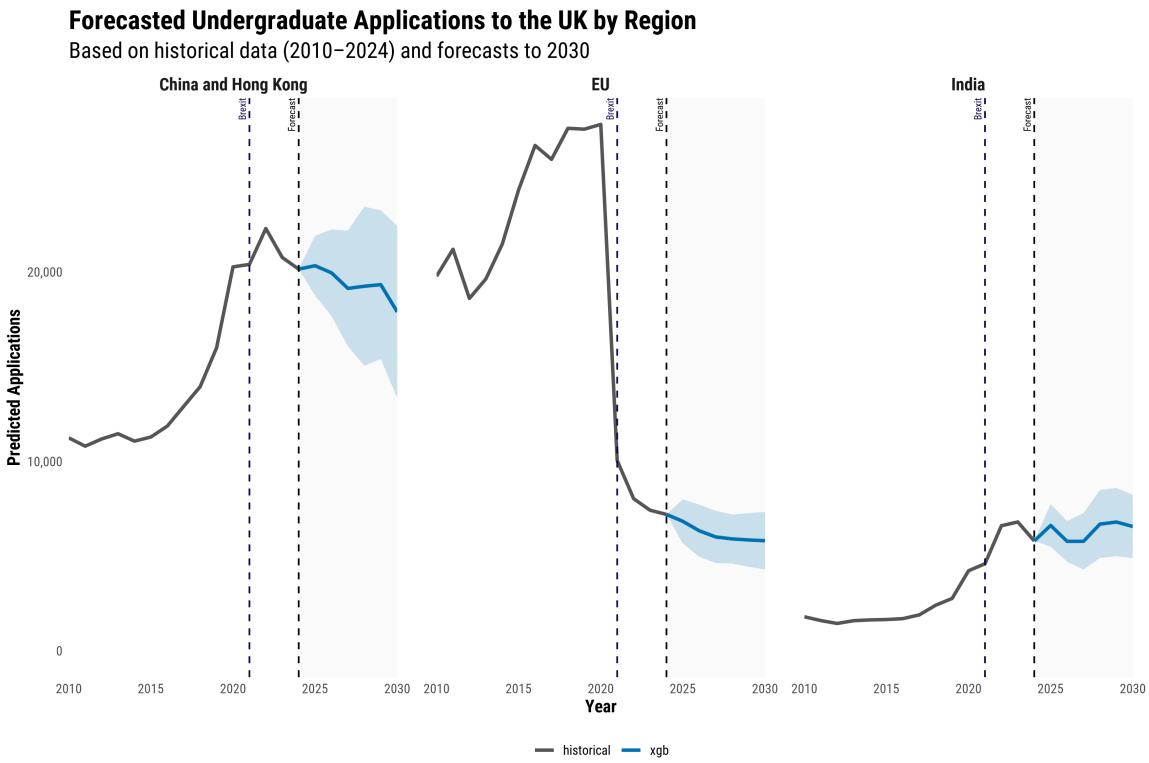

Figure 2: Forecasted undergraduate applications to UK universities from Mainland China and Hong Kong, the EU, and India, 2010–2030. Lines represent historical application counts (2010–2024) and forecasts generated by an XGBoost model (2025–2030). Vertical dashed lines mark the UK's exit from the EU (2020) and the beginning of the forecast period (2024). Shaded areas show 95% prediction intervals based on bootstrapped model uncertainty.

pronounced for origins affected by fee-status changes, visa policy adjustments, or pandemic-related constraints. Sudden shifts in visa regulations, fee status, and geopolitical relations introduce volatility that remains difficult to anticipate for traditional statistical and machine learning models alike.

Taken together, these validation results indicate that XGBoost provides the most reliable basis for generating forecasts, while emphasising that projections remain conditional on the recent structural behaviour of the system. We therefore present forward-looking estimates below for the major sending regions that have shaped the UK's international student intake over the past decade.

3.2 Forecast Trajectories for Key Origins: EU, India, and Mainland China and Hong Kong

Figure 2 presents observed applications (2010–2024) and forecast trajectories (2025–2030) for the three major sending regions: the European Union, India, and Mainland China and Hong Kong. Figure 3 summarises the changing proportional composition of the international applicant pool over the same period. These forecasts reflect concentration in international undergraduate demand, as captured by UCAS acceptance records.

3.2.1 European Union: sharp post-Brexit decline with limited recovery

Applications from EU countries fell substantially after the UK's withdrawal from the European Union and the implementation of new fee and visa regimes. Applications declined from approximately 30,000 in 2020 to around 9,000 in 2024. The model projects a continued gradual decline to around 7,000 by 2030, with narrow prediction intervals suggesting stability in the projected trajectory. This pattern

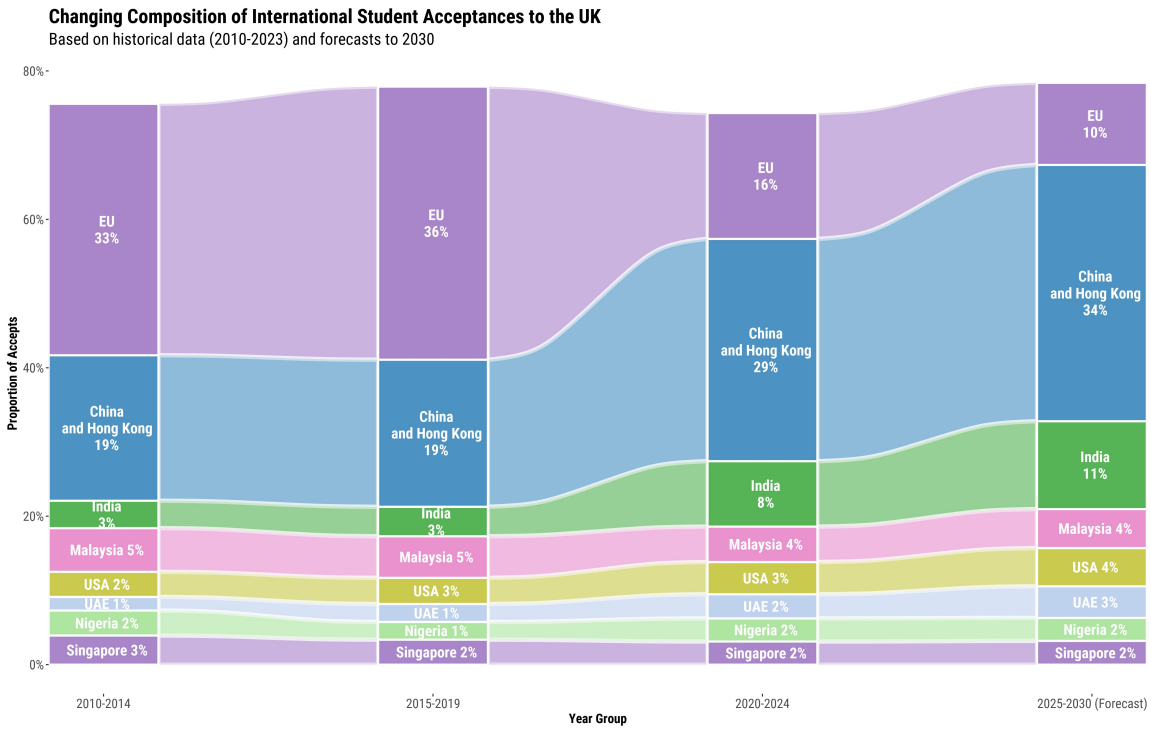

Figure 3: Changing Composition of International Student Acceptances to the UK Stacked alluvial plot showing proportional share of undergraduate applications from top 8 origin countries and regions between 2010 and 2030. Based on historical data (2010-2024) and forecasts to 2030, the figure highlights the severe post-Brexit decline in EU applications and the growing reliance on India and Mainland China. Mainland China and Hong Kong are forecast to account for one third of all undergraduate acceptances through 2025-2030, while the EU share is projected to fall to just 10%.

reflects a structural regime change rather than cyclical fluctuation, with limited evidence of spontaneous recovery under current policy settings.

Proportionally, the EU share of total international applications declines from around 36% in the pre-Brexit period to 16% in 2020–2024 and is projected to fall further to around 10% by 2025–2030.

3.2.2 India: stabilisation following a period of rapid growth

Applications from India increased rapidly in the early 2010s and immediate post-Brexit period, rising to around 6,800 in 2023 before declining slightly in 2024. The model forecasts modest short-term recovery followed by stabilisation rather than continued steep growth. This suggests a shift from recruitment-led expansion to policy-sensitive stabilisation. Despite a plateau in application counts, India's proportional share increases from approximately 8% in 2020–2024 to 11% in 2025–2030, shaped by declines and stagnation elsewhere in the system.

3.2.3 Mainland China & Hong Kong: gradual decline alongside rising proportional share

Mainland China and Hong Kong have remained the largest source of international undergraduate applications to the UK for much of the period. Applications decreased slightly from approximately 20,800 in 2023 to 20,200 in 2024. The model projects a gradual decline to roughly 18,000 applications by 2030. This dynamic aligns with evidence of growing domestic higher education capacity and shifting geopolitical relations.

Despite the projected moderation in absolute counts, the combined proportional share of Mainland China and Hong Kong is forecast to increase from around 29% in 2020–2024 to approximately 33% in 2025–2030. This increase is not driven by renewed expansion, but by sharper relative declines among other origins, reinforcing the broader concentration dynamics observed across the system.

Although presented here as a combined series for interpretability, forecasts are estimated independently for Mainland China and Hong Kong and aggregated post-estimation. Separate origin-level trajectories (Supplementary Fig. S2) indicate differences in scale and volatility between the two systems, with the decline in applications more pronounced for Mainland China than Hong Kong, but the combined trajectory remains dominated by Mainland China due to its substantially larger application volume.

3.3 Changing Composition and the Concentration Paradox

The results indicate a reconfiguration of the UK's international undergraduate student base. Whereas the early 2010s were characterised by a diversified and expanding system, the late 2020s are projected to be defined by increased concentration among three major blocs which together account for more than 60% of projected international applications.

This configuration constitutes a 'concentration paradox': market share increases for some origins even as their application counts stabilise or decline. From a migration-systems perspective, this suggests that future uncertainty is increasingly concentrated rather than evenly distributed. As a small number of corridors account for a growing share of inflows, the stability of the system becomes more reliant on the political, economic, and educational dynamics of these specific origins. This highlights that the risks inherent in forecasting are less a function of model specification and more a consequence of the structural properties and geopolitical contours shaping international student mobility.

4 Discussion

This study provides a comprehensive application of machine learning methods to forecasting international student applications to the UK. By combining high-frequency administrative data from UCAS with economic, demographic, and geographical covariates, and evaluating performance against traditional forecasting approaches, it offers methodological and substantive contributions to both ISM and migration forecasting literature.

Methodologically, the analysis demonstrated that Poisson-based XGBoost models can improve upon widely used ARIMA and gravity type models in predicting recent mobility and ISM specifically. The machine learning approach captures non-linear relationships and interactions, such as how changes in GDP per capita and population interact with lagged application counts, in ways that are difficult

to specify in parametric models. These gains, while incremental, are consistent across error metrics and robust in an out-of-sample evaluation period that includes substantial disruption (Brexit and COVID-19).

At the same time, the results reinforce the limits of purely data-driven forecasting when faced with structural shocks and policy discontinuities. The model performs less well for origins where policy interventions or conflicts generated abrupt changes in mobility that were not present in the training data. This is not unique to machine learning; it is a core challenge in migration forecasting more broadly. Nonetheless, the approach offers a more flexible and scalable tool for near-term planning, particularly when used alongside scenario-based analysis that explicitly considers policy changes.

Substantively, the forecasts highlight a trend towards increased concentration in the UK's international student base. Mainland China and Hong Kong, India, and the EU are projected to account for over 60% of undergraduate applications by 2030. Yet, two of these three are experiencing structural decline or stagnation, and the third (India) faces potential policy-induced constraints. This concentration paradox increases systematic vulnerability: the UK becomes dependent on fewer source countries whilst relationships and policy landscape imposes greater political and economic strain.

These findings have direct implications for UK higher education and migration governance. International students have long been treated as a stable revenue stream within institutional and governmental planning, but our results suggest that this assumption is increasingly untenable. Policy changes — including restrictions on dependants (UK Home Office, 2023b), tightening of post-study work pathways (Migration Advisory Committee, 2024), and proposals for an international student levy (Centre for Policy Studies, 2023) — may achieve short-term political objectives around net migration but risk undermining the financial foundations of the sector. The Office for Students has warned that around 40% of English universities face deficits post-2024, partly driven by declining international student numbers (Lapworth, 2024; Office for Students, 2024). This pressure is reflected in widespread redundancies and programme closures across the sector (Adams, 2024). At the same time, the UK Home Office has placed greater onus on universities to recruit 'genuine' students, with several institutions temporarily pausing applications from Pakistan and Bangladesh (Financial Times, 2025). These forecasts also are situated within the context of the publication of the UK's recent International Education Strategy (Department for Education and Department for Business and Trade, 2026), which emphasises sustainable recruitment and global competitiveness without imposing specific number targets for onshore student numbers, prioritising growth in education exports to £40bn per year, but through a move towards transnational education and recruitment of 'high quality' students. Forecasts such as those presented here can help quantify the longer-term implications of these policy trajectories and contextualise short-term declines within a system of growing dependence.

The patterns we document are not unique to the UK. International higher education systems globally are becoming exposed to geopolitical instability and shifting migration regimes. In the United States, deteriorating relations with China and a more hostile migration policy environment during the first Trump administration led to declines in student numbers (Rose-Redwood and Rose-Redwood, 2017; Laws and Ammigan, 2020). During the COVID-19 pandemic, anti-immigrant sentiment and discrimination against Asian populations were associated in falls in Chinese student enrolments in key destinations (Rowe et al., 2021). These examples illustrate how student mobility can serve as a barometer of broader political dynamics and how higher education is embedded within, rather than insulated from, migration politics.

From a migration systems perspective, the concentration paradox observed in UK ISM raises questions about system resilience. If a small number of corridors underpin the financial viability of HEIs, shocks affecting those corridors can propagate rapidly through the system. Brexit's impact on EU student flows offers a cautionary example: a long-established migration corridor effectively collapsed within a few application cycles, reshaping the composition of international cohorts and leaving institutions scrambling to diversify recruitment.

Our forecasts suggest that, in the absence of deliberate diversification efforts or policy changes, the UK is on a path towards similar vulnerabilities with respect to Mainland China and Hong Kong and India. This does not imply that these relationships will necessarily deteriorate, but it highlights the strategic risk of over-reliance. Building resilience may require a mix of actions: diversifying origin markets; rethinking funding models to reduce dependence on international fees; and designing migration policies that balance political constraints with the need for stable, predictable student inflows.

The study has several limitations. First, while the XGBoost model offers improved predictive

accuracy, its interpretability remains more limited than that of traditional regression models. Variable importance measures and partial dependence plots can offer insight into drivers, but the model is primarily used here as a forecasting tool rather than a causal inference framework. Institutions and policymakers who prioritise transparency may still prefer simpler models for certain applications.

Second, the evaluation period includes Brexit and COVID-19, making it inherently challenging to predict. The model's performance in calmer periods might differ, and its accuracy under future shocks cannot be guaranteed. Third, important covariates such as exchange rates, detailed visa policies, diaspora community size, and institutional prestige are not included due to data availability, particularly for the forecast horizon. Incorporating such variables could improve accuracy but would require consistent, prospective data.

Finally, we focus on undergraduate applications captured via UCAS. Our analysis relies on UCAS accepted applications rather than realised enrolments. While UCAS captures the majority of international undergraduate entrants and provides the most consistent and temporally granular administrative measure available across origin countries and time, conversion rates from acceptance to enrolment may vary across countries, subjects, and contexts. Forecasts should therefore be interpreted as projections of international student demand and confirmed undergraduate placements, rather than guaranteed realised enrolments. At the same time, supplementary comparisons with publicly available HESA undergraduate data suggest that UCAS acceptances provide a reasonable proxy for both aggregate temporal dynamics and the composition of the principal origin countries analysed here (Supplementary Section S5). The weaker UCAS–HESA correspondence for Hong Kong may partly reflect differences between UCAS applicant domicile and HESA non-UK permanent address reporting, particularly following the introduction of the British National Overseas route. In addition, UCAS data cover only undergraduate admissions and do not include all pathways into UK higher education, particularly at postgraduate level. Postgraduate taught and research students are an important part of the UK HE landscape and may follow different trajectories, and changes in their numbers may have distinct institutional and financial implications.

Future research could address these limitations in several ways. Extending the analysis to postgraduate and subject-specific applications would provide a more complete picture of risks and opportunities in the HE sector. Incorporating more detailed policy and economic indicators—including exchange rates, targeted scholarship schemes, and visa rule changes—could help to disentangle the drivers of observed shifts and refine forecasts. Methodologically, exploring alternative machine learning architectures such as recurrent neural networks or attention-based models may offer further gains in accuracy, though with additional complexity. Finally, applying similar methods to other major destinations and to multi-destination systems would facilitate comparative analysis of global ISM patterns and shed light on how shocks and policies ripple through the broader migration system.

5 Conclusion

This paper has used machine learning forecasting to examine the future of international student applications to UK universities in a period of heightened political and economic uncertainty. Leveraging UCAS administrative data from 86 origin countries between 2010 and 2024, and integrating economic and demographic covariates, we applied an XGBoost model to generate disaggregated forecasts of undergraduate applications through to 2030.

Our findings suggest that the UK's international student landscape is undergoing a profound transformation. EU student demand has collapsed following Brexit and shows little sign of recovery in the medium term. Applications from Mainland China and Hong Kong, long the largest source, appear to have peaked and are expected to decline gradually. India, a key growth market, is projected to plateau amid tightening migration policies and domestic volatility. Yet the relative importance of Mainland China and Hong Kong, India, and the EU in the UK's international student base is forecast to increase, leading to a concentration paradox in which reliance on a small set of origins intensifies even as their absolute contributions stagnate or fall.

These dynamics have important implications for the resilience of the UK higher education system and for migration governance more broadly. As institutions and governments increasingly integrate international students into their funding and migration strategies, they must grapple with the reality that ISM is highly sensitive to policy shifts and geopolitical tensions. Forecasts such as those presented here can inform more anticipatory and coordinated approaches, but they cannot substitute for

structural reforms that reduce over-reliance on volatile revenue streams and contest the framing of international students primarily through the lens of migration control.

Ultimately, international student mobility remains central to the UK's higher education landscape and its global position in knowledge production. In an era of heightened uncertainty, proactive diversification of origin markets, stable and transparent migration pathways, and investment in inclusive support for diverse student populations will be key to sustaining the benefits of ISM while mitigating the risks of an increasingly concentrated and politically contested system.

6 Data and Methodology

6.1 Data and preprocessing

We use administrative data from the Universities and Colleges Admissions Service (UCAS), the centralised platform through which the majority of applicants apply for a full-time undergraduate degree at UK higher education institutions. UCAS captures approximately 95% of EU undergraduate entrants and around 60% of non-EU entrants UCAS (2019). We analyse accepted international undergraduate applications ('acceptances') by country of domicile for the period 2010-2024, covering 86 countries of origin. These data were accessed through a formal research partnership with UCAS.

The dependent variable is the annual count of successful undergraduate applications from each origin country. In UCAS terminology, an acceptance refers to an applicant who has firmly accepted an offer of a place at a UK higher education institution through the UCAS admissions system. This stage follows the submission of an application and the receipt of one or more conditional or unconditional offers. An acceptance therefore represents a confirmed admissions placement and a high-commitment stage in the mobility decision process, although final enrolment remains contingent on meeting offer conditions and completing institutional registration. UCAS acceptance data offer several advantages for forecasting: they provide a consistent, origin-disaggregated, and temporally granular administrative series that is available with minimal delay, enabling near-term analysis of international student demand. We therefore use UCAS acceptances as a timely and consistent proxy for undergraduate mobility dynamics, while recognising that they represent admissions outcomes rather than final enrolments.

While not all accepted applicants ultimately enrol, UCAS acceptance data provide the most temporally granular and up-to-date national-level administrative measure of international undergraduate admissions outcomes available. Alternative national datasets, such as Higher Education Statistics Agency (HESA) enrolment data, are typically released with a lag of one to two years, limiting their suitability for near-term forecasting. As the centralised admissions platform for UK undergraduate education, UCAS provides a consistent longitudinal record of international student admissions across origin countries. We therefore interpret accepted applications as an administrative proxy for international student demand and mobility, while recognising that they may not perfectly correspond to final enrolment counts. The forecasting models therefore predict accepted undergraduate placements within the UCAS admissions system rather than realised international student enrolments or visa issuances. Comparisons with publicly available HESA data indicate strong aggregate temporal alignment for EU and non-EU undergraduate entrants, close correspondence in the composition of the principal non-EU origin countries highlighted in the analysis, and strong origin-specific temporal correspondence for most major sending countries. Additional sensitivity checks against HESA all-student enrolment counts show similar temporal alignment for most major origins, while also confirming that broader enrolment-stock measures are less directly comparable to UCAS undergraduate admissions flows, particularly for Hong Kong (Supplementary Section S5).

To capture the drivers of these flows, we augment the UCAS data with origin- and destination-specific covariates drawn from the World Bank, the CEPII Gravity Database, and United Nations World Population Prospects (WPP). These include:

- GDP per capita in origin and destination;
- Population size in origin and destination;
- Indicators for common official language (English) and historical colonial ties;
- Great-circle distance between the origin country and the UK.

These covariates are chosen based on prior evidence on the determinants of ISM (Neville et al., 2024a; Beine et al., 2014; Weber and van Mol, 2023). Economic and demographic variables are aligned with the application years; for the forecast period (2025–2030), we use projected values from the World Bank and WPP, mirroring the information set realistically available to decision-makers at the time of forecasting. The use of projected covariates implies that some known drivers of ISM cannot be included, such as institutional rankings or detailed measures of diaspora size.

We engineer lagged application features to capture temporal dependence, including the lagged count of accepted applications for each origin-year $(t-1)$. Continuous covariates are standardised using z-scores to facilitate model convergence and comparability across features. As the outcome is a count variable and right-skewed, we adopt modelling approaches that are appropriate for count data.

6.2 Forecasting models

Primary model: XGBoost

Our primary forecasting model is Extreme Gradient Boosting (XGBoost), a gradient-boosted decision tree algorithm adapted for count data using a Poisson objective function. XGBoost incrementally fits an ensemble of regression trees, with each successive tree trained to correct residual errors from previous trees (James et al., 2021). This makes it well suited to capturing complex, non-linear relationships and interactions among predictors, while offering standard measures of variable importance.

Let y_i denote the number of successful applications from origin country–year i, and x_i the corresponding vector of covariates, including lagged applications, economic and demographic indicators, and structural factors (distance, common language, colonial ties). XGBoost approximates the log of the expected count via a sum of decision trees:

$$\log \hat{y}_i = f(x_i) = \sum_{k=1}^{K} f_k(x_i),$$

where each f_k is a regression tree and K is the number of boosting rounds. Under the Poisson objective, the loss function can be written as:

$$L(\theta) = \sum_{i=1}^{n} [\exp\{f(x_i)\} - y_i f(x_i)] + \lambda \sum_{j} \theta_j^2,$$

where θ denotes the set of tree parameters and λ is a regularisation term that discourages overfitting. Predictions are obtained as

$$\hat{y}_i = \exp\{f(x_i)\}.$$

Further implementation details are provided in the Supplementary Information.

Hyperparameter tuning for XGBoost was performed using grid search with five-fold cross-validation over the training set. We varied the number of boosting rounds, learning rate, maximum tree depth, subsampling rate, and column subsampling rate, selecting the combination that minimised mean absolute error (MAE) on a validation set. The final model used approximately 150–200 boosting rounds, a learning rate of 0.3, and a maximum tree depth of 6, as summarised in Supplementary Table 2.

Benchmark models: ARIMA and negative binomial gravity

To contextualise the added value of machine learning, we implement two benchmark forecasting approaches commonly used in sector planning and migration research.

First, we estimate autoregressive integrated moving average (ARIMA) models for each origin country. ARIMA models extrapolate historical trends in application counts without incorporating external covariates and are widely used in institutional planning. Country-specific orders are selected using information criterion–based optimisation. ARIMA thus serves as a pragmatic baseline for assessing whether more flexible methods provide measurable gains in predictive performance.

Second, we estimate a negative binomial gravity-type model that incorporates time-varying economic and demographic characteristics, as well as structural relationships (distance, common language, colonial ties), with country-level random intercepts to capture unobserved heterogeneity. Gravity models are well established in migration analysis and provide a theory-aligned comparator to the machine learning approach. Complete model equations, estimation details, and comparative assumptions are presented in the Supplementary Information (Section 2).

6.3 Model evaluation and forecasting strategy

Models are trained on data from 2010–2018 and evaluated on an out-of-sample period from 2019–2023, which includes the implementation of Brexit and the COVID-19 pandemic. This period allows us to assess performance under conditions of structural volatility rather than smooth trend continuation.

We evaluate predictive accuracy using three standard error metrics. Mean absolute error (MAE) is defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|,$$

and measures the average magnitude of prediction errors. Root mean squared error (RMSE) is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

which penalises larger errors more heavily. Normalised mean absolute error (NMAE) is defined as:

$$\text{NMAE} = \frac{\text{MAE}}{\bar{y}},$$

where $\bar{y}$ is the mean number of applications for a given country, allowing comparison of relative error across countries with different scales of flows. We also examine Pearson correlations between observed and predicted application counts to assess the strength of association across the out-of-sample period.

Definitions and implementation notes for all evaluation metrics (MAE, RMSE, MAPE, NMAE, and Pearson correlation) are provided in the Supplementary Information (Section 2.3).

To construct prediction intervals, we use a bootstrap procedure, repeatedly resampling residuals and regenerating predictions to approximate the distribution of forecast errors. The 2.5th and 97.5th percentiles of the resulting distribution form the bounds of 95% prediction intervals, providing a measure of uncertainty around the point forecasts.

After evaluating model performance, we retrain the XGBoost model on the full observed period (2010–2024) using the tuned hyperparameters. We then generate forecasts for 2025–2030 using recursive prediction: the forecast for 2025 uses observed lagged applications from 2024, and forecasts for subsequent years treat prior forecasts as lagged inputs. Economic and demographic covariates for 2025–2030 are taken from the projected World Bank and WPP series, ensuring that the model uses information that would realistically be available at the time of forecasting.

All models are estimated independently at the origin–year level. Regional aggregates (e.g., Mainland China and Hong Kong combined; European Union) are constructed post-estimation by summing origin-specific forecasts. Uncertainty intervals for aggregated series are obtained by summing bootstrap predictions across origins within each iteration and then computing empirical quantiles. No pooled or joint estimation is performed across origins. This approach is particularly important for cases such as Mainland China and Hong Kong, which differ in visa regimes, historical ties, and mobility dynamics. Modelling origins independently avoids conflating heterogeneous trajectories, and the aggregated series shown in the main text are purely descriptive.

7 Data Availability

The data that support the findings of this study were provided by the Universities and Colleges Admissions Service (UCAS) under licence and are not publicly available. Access to these data is subject to agreement with UCAS. Data may be made available from the authors upon reasonable request and with permission of UCAS.

8 Code Availability

The code used to process the data, develop the forecasting models, and generate all results is publicly available at https://github.com/ruthneville/ism_uk_forecasts. The repository includes scripts for data preprocessing, model estimation (ARIMA, negative binomial gravity model, and XGBoost),

forecasting, and uncertainty quantification. All analyses were conducted in R using packages including `tidyverse`, `glmmTMB`, and `xgboost`.

All scripts are designed to enable reproduction of the main results, subject to access to the underlying data.

Acknowledgements

This work benefited from exchanges with colleagues at the Max Planck Institute for Demographic Research and the Geographic Data Science Lab, in particular Ugofilippo Basellini and Irena Chen (MPIDR), as well as mentorship and discussion with Alex Singleton (University of Liverpool). This research was supported by grants from the Economic and Social Research Council (ES/P000401/1), and data and support were provided through an industry partnership with the UK Universities and Colleges Admissions Service (UCAS).

References

Adams, R. (2024). More than 40 uk universities planning job cuts amid financial strain. *The Guardian*.

Altbach, P. G. and Knight, J. (2007). The internationalization of higher education: Motivations and realities. *Journal of Studies in International Education*, 11:290–305.

Amuedo-Dorantes, C. and Romiti, A. (2024). International student applications in the uk after brexit. *Journal of Economic Geography*, 24(5):637–662.

Azose, J. J. and Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, pages 1627–1650.

Beine, M., Noël, R., and Ragot, L. (2014). Determinants of the international mobility of students. *Economics of Education Review*, 41:40–54.

Bijak, J., Hilton, J., Silverman, B. W., and Smith, P. W. (2019). Assessing time series models for forecasting international migration: Lessons from the united kingdom. *Journal of Forecasting*, 38(5):470–487.

Bolton, P. and Lewis, J. (2021). The end of free movement and the implications for higher education. Technical Report CBP 9187, House of Commons Library. Accessed: 2025-12-09.

Bolton, P., Lewis, J., and Gower, M. (2023). International students in uk higher education. Technical Report CBP 7976, House of Commons Library. Accessed: 2025-12-09.

British Council (2024). The outlook for international student mobility: amidst a changing global macroeconomic landscape.

Cabanas-Tirapu, O., Danús, L., Moro, E., Sales-Pardo, M., and Guimerà, R. (2025). Human mobility is well described by closed-form gravity-like models learned automatically from data. *Nature Communications*, 16:1336.

Caruso, R. and De Wit, H. (2014). Determinants of mobility of students in europe: Empirical evidence for the period, 1998–2009. *Journal of Studies in International Education*, 19:265–282.

Centre for Policy Studies (2023). Stopping the brain drain: A student visa levy for the uk. Technical report, Centre for Policy Studies.

Cuibus, M. and Walsh, P. (2024). *Student Migration to the UK*. Migration Observatory, Oxford.

Czaika, M. and Parsons, C. R. (2017). The gravity of high-skilled migration policies. *Journal of Economic Geography*, 17(2):371–403.

de Haas, H. (2010). Migration and development: A theoretical perspective. *International Migration Review*, 44(1):227–264.

Department for Education and Department for Business and Trade (2026). The uk international education strategy 2026. https://assets.publishing.service.gov.uk/media/696a6164448fedc1eb4248ef/international-education-strategy-2026.pdf. UK Government policy document.

Department of Education and Department for International Trade (2021). International education strategy: global potential, global growth.

Disney, G., Wiśniowski, A., Forster, J. J., Smith, P. W., and Bijak, J. (2015). Evaluation of existing migration forecasting methods and models: Report examining and assessing the various analytical approaches to migration forecasting. Technical report, ESRC Centre for Population Change, University of Southampton, Southampton, UK.

Docquier, F. and Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of Economic Literature*, 50(3):681–730.

Financial Times (2025). Uk universities restrict recruitment of bangladeshi and pakistani students over visa concerns. *Financial Times*. Accessed: YYYY-MM-DD.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer, New York, 2 edition.

Jisc (2025). Uk tops new index of university destinations for world leaders. https://www.jisc.ac.uk/news/all/uk-tops-new-index-of-university-destinations-for-world-leaders. Accessed: YYYY-MM-DD.

Jon, J.-E. (2013). Realizing internationalization at home: Korean higher education institutions' responses to internationalization. *Journal of Studies in International Education*, 17(4):460–475.

King, R. and Sondhi, G. (2018). International student migration: a comparison of uk and indian students' motivations for studying abroad. *Globalisation, Societies and Education*, 16(2):176–191.

Knight, J. (2012). *Student Mobility and Internationalisation: Trends and Policies*. UNESCO, Paris.

Lapworth, S. (2024). Universities fighting for their future must be bold and must be fast. Office for Students Blog. Available at: https://www.officeforstudents.org.uk/news-blog-and-events/blog/universities-fighting-for-their-future-must-be-bold-and-must-be-fast/. Accessed 18 July 2025.

Laws, K. N. and Ammigan, R. (2020). International students in the trump era: A narrative view. *Journal of International Students*, 10(3):xviii–xxii.

Lewer, J. J. and Van den Berg, H. (2008). A gravity model of immigration. *Economics Letters*, 99(1):164–167.

Litmeyer, M.-L. and Hennemann, S. (2024). Forecasting first-year student mobility using explainable machine learning techniques. *Review of Regional Research*, 44:119–140.

Lomer, S. (2017). Uk policy discourses and international student mobility: The deterrence and subjectification of international students. *Globalisation, Societies and Education*, 16(3):308–323.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J. E. (1993). Theories of international migration: A review and appraisal. *Population and Development Review*, 19(3):431–466.

Migration Advisory Committee (2024). Rapid review of the graduate route. Technical report, Migration Advisory Committee.

Migration Observatory (2025). Student migration to the uk. https://migrationobservatory.ox.ac.uk/resources/briefings/student-migration-to-the-uk/. Accessed: 2025-12-09.

National Research Council (2000). *Beyond Six Billion: Forecasting the World's Population*. National Academy Press, Washington, DC.

Neville, R., Rowe, F., and Singleton, A. (2024a). Nonlinear relationships between human development and international student mobility: The prominent role of employment prospects and cultural and linguistic ties. *Population, Space and Place*, 30:e2610.

Neville, R., Rowe, F., and Singleton, A. (2024b). Unraveling the brexit–covid-19 nexus: Assessing the decline of eu student applications into uk universities. *Data & Policy*, 6:e45.

OECD (2023). *Education at a Glance 2023*. OECD Publishing, Paris.

Office for Students (2024). Financial sustainability of higher education providers in england: 2024. Technical report, Office for Students. Accessed: 2025-12-09.

Robertson, S. (2013). Transnational student-migrants and the state: The education–migration nexus. *Population, Space and Place*, 19(2):180–193.

Rose-Redwood, C.-A. and Rose-Redwood, R. (2017). Rethinking the politics of the international student experience in the age of trump. *Journal of International Students*, 7(3):I–X.

Rowe, F., Mahony, M., Graells-Garrido, E., Rango, M., and Sievers, N. (2021). Using twitter to track immigration sentiment during early stages of the covid-19 pandemic. *Data & Policy*, 3:e9.

Saxenian, A. (2005). From brain drain to brain circulation: Transnational communities and regional upgrading in india and china. *Studies in Comparative International Development*, 40(2):35–61.

She, Q. and Wotherspoon, T. (2013). International student mobility and high-skilled migration: A comparative study of canada, the united states, and the united kingdom. *SpringerPlus*, 2:132.

Spivak, G. C. (2009). Rethinking comparativism. *New Literary History*, 40(3):609–626.

UCAS (2019). 2019 international insights report. Available at: https://www.ucas.com/data-and-analysis/ucas-international-insights. Accessed 18 July 2025.

UK Home Office (2023a). Immigration rules: Changes to appendix student and appendix graduate. Policy statement.

UK Home Office (2023b). Statement of changes to the immigration rules: HC 1160. Accessed: 2025-12-09.

UK Home Office (2024). New package of measures to cut net migration and tackle abuse of the student route. Accessed: 2025-12-09.

UNESCO Institute for Statistics (2023). Global flow of tertiary students. http://uis.unesco.org/en/uis-student-flow. Accessed: 2025-12-09.

UNESCO Institute for Statistics (UIS) (2024). Global flow of tertiary-level students. Available at: https://www.migrationdataportal.org/themes/international-students-trends.

Universities UK (2024). The impact of reforms to the graduate route and dependant rules. Technical report, Universities UK. Policy Brief.

Weber, T. and van Mol, C. (2023). The student migration transition: An empirical investigation into the nexus between development and international student migration. *Comparative Migration Studies*, 11(5).

Wei, H. (2013). An empirical study on the determinants of international student mobility: A global perspective. *Higher Education*, pages 105–122.

Yang, S.-W., Chen, H.-C., Chen, W.-C., and Yang, C.-H. (2020). Forecasting outbound student mobility: A machine learning approach. *PLOS ONE*, pages 1–21.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryinfotracked.pdf
- supplementaryinfo.pdf