Methodology to estimate international migration in the absence of the International Passenger Survey

21 July 2021

Nicky Rogers and Duncan Elliott, Principal Statistical Methodologists, Methodology Division

Dominic Webber, Mark Bangs, Brendan Georgeson, Migration Statistics Division

Louisa Blackwell, Project Lead Integrated Statistical Design, Social Statistics Transformation, Analysis and Research

CONTENTS **Page No.**

1. Executive Summary

- ONS are delivering, through the Census and Data Collection Transformation Programme (CDCTP), a transformed population, migration and social statistics systems using administrative data. This work will inform the recommendation in 2023 on the future of the census and is a key enabler for delivering the Statistics for the Public Good strategy.
- The ONS Centre for International Migration has been engaged in a programme of work to develop estimates of international migration based on administrative data sources since 2017.
- ONS has long acknowledged that the International Passenger Survey (IPS), which previously underpinned our international migration estimates, has been stretched beyond its original purpose and that we need to consider all available sources to understand international migration.
- Enabled by data-sharing powers in the Digital Economy Act 2017, guided by our data strategy and principles, and building on the ongoing collaboration and data-sharing under way with government departments, we are seizing the opportunity to make use of more data to give us a much richer understanding of how our population is changing.
- The coronavirus (COVID-19) pandemic has led us on an innovative journey to estimate international migration in the absence of the International Passenger Survey (IPS).
- This paper outlines the approaches and methods that have been developed by ONS Methodology and the ONS Centre for International Migration to estimate international migration.
- It is not the intention of ONS to use the IPS as the leading indicator of migration in the future so these approaches will form the basis for international migration estimates in the transformed population statistics system.
- Two methods to estimate international migration were developed independently; a tactical model-based approach to estimate international migration in the absence of the IPS and a strategic Admin-Based Migration Estimates (ABMEs) approach that has focussed on administrative data sources that can tell us about both immigration and emigration; namely Department for Work and Pensions (DWP) Registration and Population Interaction Database (RAPID)[1] and border data from the Home Office exit checks dataset[2].
- We introduce the application of State Space Models to estimate international migration flows for March 2020 and Quarter 2 (April to June) 2020, and our plans to continue using this approach to estimate migration in the last two quarters of 2020. Our intention is to use these models as a statistical framework to estimate international migration going forward. Panel members commented on earlier iterations of these models.
- DWP RAPID data are used to determine signs of "activity" among the population in the UK, and this information is used to infer migrant flows based on when that activity commenced (indicating arrival in the UK) and when it ends (indicating departure). So far, our work has

[1] DWP Registration and Population Interaction Database (RAPID). RAPID provides a single coherent view of citizens' interactions across the breadth of systems in both DWP and HM Revenue and Customs (HMRC) including benefits, employment, self-employment, pensions and in-work benefit.

[2] Home Office border crossing data collect data on travellers arriving and departing the UK. These data are based on administrative data derived from the data matching system and analytical capability built by the Exit Checks Programme, the Initial Status Analysis (ISA). While the Exit Checks Programme closed in May 2016 having delivered its objectives, the ISA remains in place and is delivering results.

used aggregate data which provide counts of the number of weeks of activity for each individual in the dataset.

- Home Office visa and border data provide further insight into indicative estimates of immigration. Initial figures are based on actual travel patterns of non-EEA nationals. Future work will aim to build further understanding of the inclusion of EEA nationals in the data following the UK's exit from the European Union.
- Estimating international migration using non-survey data sources is complex and challenging. Definitionally, 12 months or more need to pass before we can assume if someone who has travelled to or from the UK at the start of that period is a migrant according to the long-established UN international definition[3].
- Users of our statistics have identified the need for timely coherent statistics on the size (or stock) of the population and how it changes over time (flows, both nationally and locally, and by age and sex), as well as relevant migration analysis in a rapidly changing society. There is scope to look beyond the current UN definition of international migration and produce estimates based on different definitions.
- We plan to further develop and use the tactical model-based approach to produce a *provisional* set of estimates with a minimal time lag. We also plan to develop and use the strategic administrative data approach to produce a set of *final* estimates, albeit with a significant time lag. Alongside this, we will continue to explore data sources, and incorporate them into both approaches if required.
- Our ambition is to bring together and therefore narrow the time between the provisional and final estimates if possible. To do this we aim to use modelling as the statistical framework around which we will develop international migration estimates.
- We introduce the admin-based dynamic population model that aims to produce timely provisional England and Wales population estimates that will be at the core of the transformed population and migration system. Timely international migration estimates are critical to the dynamic admin-based population models and will feed into this system.

2. Ask

Members of the Methodological Assurance Review Panel (MARP) are invited to:

- provide feedback on the presentation of the tactical and strategic approaches to estimating international migration set out in Sections 5 and 6
- provide feedback on our proposals to bring together the approaches for future iterations of international migration estimates - Administrative Based Migration Estimates (ABMEs) supported by statistical modelling - outlined in Section 8.
- consider the following questions specific to our approach:
    - Should we pursue this approach for estimating international migration?
    - Are there any other approaches we should be considering?
    - Do you have any specific comments on the strategic and tactical approaches?
    - Do you have any suggestions for bringing together the tactical and strategic approaches?

[3] The UN recommended definition of a long-term international migrant: "A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence".

- Are you aware of similar work?

The panel are not being invited to review the results of the modelling and ABME research or to provide comment on the proposed future population, migration and social statistics system outlined in Section 9.

3. Introduction

ONS has long acknowledged that the International Passenger Survey (IPS), which previously underpinned our international migration estimates, has been stretched beyond its original purpose and that we need to consider all available sources to understand international migration.

Enabled by data-sharing powers in the Digital Economy Act 2017, guided by our data strategy and principles, and building on the ongoing collaboration and data-sharing under way with government departments, we are seizing the opportunity to make use of more data to give us a much richer understanding of how our population is changing. Since 2017, the ONS Centre for International Migration has been engaged in a programme of work to develop estimates of international migration based on administrative data sources. The coronavirus (COVID-19) pandemic has led us on an innovative journey to estimate international migration in the absence of the International Passenger Survey (IPS).

This paper outlines the approaches and methods that have been developed by ONS Methodology and the ONS Centre for International Migration to estimate international migration in the absence of the International Passenger Survey (IPS). It is not the intention of ONS to use the IPS as the leading indicator of migration in the future so these approaches will form the basis for international migration estimates in the transformed population statistics system.

Two methods to estimate international migration were developed independently:

1. A tactical approach to model international migration flows during Quarter 2 2020 in the absence of the IPS, developed by Methodology. This method used Home Office (HO) border data and made assumptions around migrant behaviour during the coronavirus pandemic within a statistical time series methodology. The model estimated total immigration, emigration and net migration by broad nationality group (GB, EU and Non-EU). Panel members commented on an early iteration of these models in December 2020.

2. A strategic approach to provide high-level international migration estimates by broad nationality group (EU, Non-EU) using Department for Work and Pensions (DWP) and Home Office administrative based data, developed by the ONS Centre for International Migration. These were labelled as Admin-Based Migration Estimates (ABMEs). This method produced estimates up to March 2020.

Traditionally, data from the IPS are the main source for estimating international migration to and from the UK. The data allow estimates of immigration, emigration and net migration by citizenship and reason for migration. These estimates feed into the annual Mid-Year Population estimates for England and Wales that are published in June each year for the previous mid-year point.

ONS are delivering, through the Census and Data Collection Transformation Programme (CDCTP), a transformed population, migration and social statistics systems using administrative data. This work

will inform the recommendation in 2023 on the future of the census and is a key enabler for delivering the Statistics for the Public Good strategy.

Work to transform international migration statistics to put administrative data at their core has been ongoing for several years. Previous research identified a range of data sources held across government that can help estimate international migration including immigration, income, benefits, and education data. However, some of these data sources can only tell us about migration into the UK. Therefore, a strategic approach to develop Admin-Based Migration Estimates (ABMEs) has focussed on data sources that can tell us about both immigration and emigration. So far, the two sources of administrative data which have shown greatest potential for the estimation of long-term immigration and emigration are:

- Department for Work and Pensions (DWP) Registration and Population Interaction Database (RAPID)
- border data from the Home Office exit checks dataset

In the first iteration of our development of Admin-Based Migration Estimates (ABMEs), published in April 2021, we explored each data source individually, developing methods to estimate long-term migration within each source. We are still at a relatively early stage in this work and there remains further development, including work to further understand data quality, before we will be able to produce official estimates of international migration using these sources.

In 2019, ONS Methodology began exploring time series modelling as an approach to estimate international migration using administrative outcomes-based data. Time series modelling was selected because of the strong seasonal trends that are evident in international migration over time. This research was accelerated during 2020 with the suspension of the IPS on 16 March 2020.

Details of both the tactical and strategic approaches were published on 16 April 2021. A summary of the methods is provided in sections 5.2 and 6.2. respectively. The 16 April publications demonstrate our ability to use the data that we currently have and the insights that we have gained from our administrative data research to estimate international migration using a model-based method. These modelled outputs were endorsed and welcomed by expert users. The provisional 2020-based mid-year population estimates integrated the modelled estimates and underline the relevance and importance of producing timely population outputs during the pandemic.

Going forward, we need to consider the statistical design and methods for delivering international migration estimates for Quarters 3 and 4 during 2020 and beyond. This includes how we respond to user needs in the short- to medium term, but also the challenges of delivering sustainable, timely and high-quality estimates in the longer-term. This means that as well as continuing the methodological research and data evaluation and exploration, we must produce estimates along the way to inform our users. We are working closely with stakeholders so that they understand this, so it is important that we are transparent about the methods, data and assumptions which underpin any estimates. We are also aware that we need to be able to provide quality measures alongside the estimates to ensure their use is proportionate to their quality.

For the remainder of this paper we:

- describe user requirements for international migration statistics
- outline the data sources and methods for our two approaches to estimate international migration
- review the advantages and disadvantages of the two methods

- present our strategy for future iterations of international migration estimates beyond 2020
- introduce how our estimates will integrate with the future population, migration and social statistics system

4. User requirements for international migration estimates

One of the principal purposes of international migration statistics is to inform the overall measurement of the UK population. At the national level, the size of the population is determined by the number of births, deaths, and net migration, and then regionally, internal migration. As such, there is a clear need to provide estimates of long-term international immigration and emigration broken down for example by:

- UK country/Devolved Administrations
- local authority
- single year of age
- sex
- reason for migration
- citizenship/nationality

Alongside this requirement, there are a range of key stakeholders and users, which Migration Statistics Division regularly consult with to understand needs. Broadly, they require coherent statistics on the size (or stock) of the population and how it changes over time (flows, both nationally and locally, and by age and sex), as well as relevant migration analysis in a rapidly changing society.

More specifically, recent engagement with various user forums has identified:

- Coherence across population and migration estimates produced by the transformed population and migration statistics system.
- A need for evidence on the reasons why people choose to migrate (most likely through Visa data).
- Insight into subgroups of the population, such migrant workers, students, and families, while being alert to coherence with overall population measures.
- Frequent and timely statistics, however our users are sympathetic to the challenging environment that currently faces the production of migration statistics.
- A continuation of recent collaboration with GSS to research migration and the labour market (e.g. industry breakdowns).
- A need to be explicit about the difference in how we're proposing to measure population change when moving from survey to administrative data.
- Publication of an output strategy detailing the format and granularity of estimates that can be produced using administrative data, that also considers needs for international comparability.
- The current UN 12-month definition is not essential and other definitions should be considered, for example majority of time. Similarly, more consideration is required around definitions of migrants in terms of country of birth vs nationality.
- Consider how we will measure and communicate uncertainty under the new estimation approaches.
- How to identify and account for non-residents who may appear as active in UK on admin data (circular migration).

5. Modelling international migration (tactical) for Quarter 2 (April to June) 2020

5.1. Data sources

We used International Passenger Survey (IPS) data from January 2010 to February 2020 in a time series modelling approach to estimate seasonality and trends over time. Whilst the general pattern of migration has been relatively stable in the past, it can be impacted by a range of one-off events, such as the 2008 economic recession or the UK's exit from the EU, among others. IPS data for March 2020 were extrapolated as only partial data were available for that month[4]. These estimates, based on extrapolated data, were used to produce the published estimates of international migration for the year ending March 2020[5].

We explored the best available data sources:

- HO border crossing data, visa, and Advanced Passenger Information (API) data
- DWP National Insurance Number allocations data for overseas adult nationals
- Civil Aviation Authority (CAA) train and ferry passenger arrival and departure data
- Personal Demographic Service (PDS) data from the NHS on new registrations from overseas and embarkations

Other sources were explored but were not considered as they were less timely and did not cover the period of interest, i.e. Quarter 2 2020. However, they may be useful for estimation over longer time periods or for sub-groups of migrants. These included Higher Education Statistics Agency (HESA) data, and English and Welsh School Census data. Annual Population Survey (APS) data were also considered but were not used as there would be circularity introduced into the model since APS data are weighted to population estimates that include an international migration component.

We investigated Her Majesty's Revenue and Customs (HMRC) Real Time Information (RTI) data but concluded that while these data have potential, they are only representative of movements into and out of the labour market, not migration. Additionally, we believed that the link between where people live and work during the pandemic had changed and this may not be reflected in RTI data. For example, some overseas nationals who are furloughed may choose to return to their home country to be with family. Equally, many people are working for a UK business but are resident with family overseas.

We continue to explore other data sources as they become available and existing sources as they are updated. In the coming year, we anticipate data supplies that will shed light on migration behaviours during 2020, for example, DWP RAPID data. The 2021 Census will provide the most robust and comprehensive picture of the England and Wales population possible and an opportunity to validate our estimates. The IPS was reinstated early 2021 but with no migrant focussed shifts, so the sample of migrants is much reduced. We will continue to assess these data and their inclusion in our models.

[4] The International Passenger Survey (IPS), which underpins our existing UK international migration statistics, was suspended because of the impact of the coronavirus pandemic on 16 March 2020. While the IPS resumed operations in January 2021, the decision was taken and announced in the August 2020 Migration Statistics Quarterly Report (MSQR) that IPS data would no longer be used to calculate international migration

[5] The latest available IPS data cover the vast majority of the year ending March 2020, which have been published in the August 2020 Migration Statistics Quarterly Report (MSQR).

5.2. Methods

We used multivariate State Space Models (SSMs) to estimate international migration flows. Time series approaches are a good tool for modelling time-varying processes and their different features, such as trends, seasonality, and random components. SSMs offer the added flexibility of the time series approach by explicitly modelling the latent (unobserved) process and its (observed) measurements. These models were fitted to the historical International Passenger Survey (IPS), and administrative data including Home Office border crossing data and air, train, and ferry passenger data.

Estimates are based on an extrapolation of the trend and seasonal components from the IPS, together with the Home Office border crossing and travel data for the period March to June 2020. The IPS and administrative data are assumed to be correlated. The error term in the model captures the change in the trend for the administrative based time series (caused by COVID-19). It is this error term that is used to predict the change in migration. Estimates were generated from the models for British, EU and non-EU nationals' immigration and emigration and are summed to estimate the corresponding totals. Net migration is calculated by taking the difference between immigration and emigration.

We developed these models with time series experts in ONS Methodology and with modelling and migration experts at the Universities of Southampton and Warwick. We used a Delphi approach to gather expert judgement on model assumptions and early modelled estimates based on different scenarios. From this process we produced a range of provisional immigration, emigration and net migration estimates with uncertainty measures for Quarter 2 (April to June) 2020. This process was guided by expert input. We provided uncertainty intervals around the migration estimates to reflect uncertainty in the predicted values from March onwards, where we do not have observations from the IPS. Please refer to Appendix 1 for further information on estimating uncertainty.

The uncertainty intervals for net migration were calculated using the sum of the prediction error variances derived from the immigration and emigration confidence intervals. They should be treated as indicative; in future iterations we will develop a more precise measure which will not necessarily be symmetrical around the estimated net migration values.

We evaluated the fitted models using a set of diagnostic statistics (reported in Appendix 1). Appendix 1 also gives further detail on the mathematical specification of the multivariate SSMs.

The Delphi-approach

The modelling used several assumptions, regarding:

1) the interpretation of International Passenger Survey (IPS) estimates for early 2020
2) whether the changes in migration behaviours observed for non-EU migrants in Home Office border crossing data also applied to EU and British nationals who are migrating
3) whether EU migrants had more opportunity to come to the UK or return home than non-EU migrants, given the continued operation of cross-channel travel services when air travel was virtually halted from April 2020

We invited migration experts and stakeholders to give us their view of the assumptions and invited them to provide further evidence that we should consider. The discussion was guided by a

questionnaire and empirical evidence on each assumption. The assumptions considered by experts are presented in Table 1.

Table 1: Assumptions considered by experts

	Title	Assumption
1	International Passenger Survey (IPS) data for Q4 (Oct-Dec) 2019	We are content that the IPS migration estimates for Q4 2019 are broadly in line with other reporting on international migration for that period
2	IPS immigration estimates for January 2020	The increased immigration of students from Southern Asia and China in January 2020 reported by the International Passenger Survey was real
3	IPS estimates for March 2020	We should use modelled estimates for March 2020, not IPS. IPS results from 1-15 March did not capture changed migration patterns after March 16 2020
4	Non-EU migration change in Q2	Flows for non-EU visa nationals in Home Office border crossing data are a true representation of changed patterns of arrivals and departures for this group in Q2. Time will reveal if these departures are long-term migrations. For now, we can use deferred arrivals, and departures up to 9 months ahead of a visa end date, as an indicator of behavioural change.
5	EU migration change in Q2	We can use non-EU border crossing evidence of changed migration behaviour to model EU migration in Q2. We can modify this to reflect increased travel options for EU migrants.
6	GB migration change in Q2	We can use non-EU border crossing evidence of changed emigration behaviour to model British *immigration* in Q2. We can use non-EU border crossing evidence of changed *immigration* behaviour to model British emigration in Q2.

We ran a second round with experts where we presented a report that included their collated comments and advice, together with outputs from six models (two for immigration and six for emigration. Further detail on these six models is available in Section 8 of our published report). Round 2 resulted in experts reaching a majority view on the assumptions underpinning our immigration and emigration models. These are summarised in Table 2.

Table 2: Implemented assumptions following expert advice

	Non-EU	**EU**	**GB**
Immigration	Model from March 2020	Model from March 2020 but apply travel options adjustment from April 2020	Model from March 2020 using non-EU Home Office border crossing data on departures
Emigration	Model from March 2020	Model from March 2020 but apply travel options adjustment from April 2020	Model from March 2020 using non-EU Home Office border crossing data on arrivals

6. Admin-Based Migration Estimates (Strategic)

6.1. Data sources

The ONS Centre for International Migration is engaged in a programme of work to develop estimates of international migration based on administrative data sources. In April we published a report on progress in the development of these admin-based migration estimates (ABMEs).

The main focus of our work developing ABMEs has so far centred on two sources of data, which have shown greatest potential for the measurement of long-term migration:

- Department for Work and Pensions (DWP) Registration and Population Interaction Database (RAPID)
- Home Office visa and border data

RAPID provides a single coherent view of citizens' interactions across the breadth of systems in both DWP and HM Revenue and Customs (HMRC) including benefits, employment, self-employment, pensions and in-work benefit. We have been developing the use of these data to determine signs of "activity" among the population in the UK, and how we can use that information to infer migrant flows based on when that activity commenced (indicating arrival in the UK) and when it ends (indicating departure). So far, our work has used aggregate data which provide counts of the number of weeks of activity for everyone in the dataset. DWP are now working on adding monthly indicators of activity to the dataset which would allow for more sophisticated analysis of individuals' activity on administrative systems. Crucially, the coronavirus pandemic has highlighted instances where overseas nationals, who usually live and work in the UK, may have chosen to move back to their home country during the pandemic to be with family. Therefore 'activity' in the RAPID data may not be an adequate measure of presence in the UK over this period.

As RAPID covers everyone with a National Insurance number (NINo) it covers both migrants from EU countries and those from countries outside of the EU. While the coverage is extensive, and most migrants will appear in RAPID, there are some groups less well covered in the data. For example, migrant children under 16 years of age are not separately identifiable in RAPID and visiting students who do not hold a NINo will not be included in the dataset.

Home Office visa and border data provide a more direct measure of movement and can be used to identify when non-European Economic Area (EEA) nationals have entered or exited the UK. Previous analysis published in February 2020 used Home Office visa and border data to explore patterns and definitions of long-term international immigration. Our latest ABME research has built on this method, which looks at first arrival and last departure within a visa period as an approximation for length of stay in the UK. Visa periods are constructed by linking together any consecutive or concurrent visas held. If there is a gap between visas, then a new visa period is started.

Home Office border data currently only include nationals of non-European Economic Area (EEA) countries, and have only been collected from April 2015, allowing for migration estimates from the year ending 7 April 2017 onwards. Future work will aim to build further understanding of the movement of EEA nationals in these data, as this information is collected from 1 January 2021 following the UK's exit from the EU.

Going forward, it is clear we need to consider, through data linkage, how we can identify EU and Non-EU nationals who have since been granted citizenship in the RAPID data, as well as identifying

EU nationals who have been previously resident and have settled status through the EU Settlement Scheme as well as new EU national visa holders arriving in the UK post 1 January 2021.

6.2. Design and Methods

Delivering new measures of international migration using administrative data sources presents a substantial change in the measurement of migration. Until now, estimates of international migration have been based on the International Passenger Survey (IPS) which interviewed migrants to record how long they were intending to remain in or out of the UK in the next 12 months. Administrative data on the other hand are retrospective and tell us about actual activity that has already happened. It is important to understand this shift in the underlying data and the methodology when comparing estimates based on administrative data to the previously published IPS estimates.

To estimate international migration to and from the UK using RAPID data, the DWP and the ONS worked together to develop a methodology using the annual tax year summary datasets to create a version of RAPID including only Non-UK nationals (RAPID Migration Dataset). There are two main steps in the creation of the RAPID migration dataset which are important in understanding the methodology for identifying patterns of migration (both long-term and short term).

Firstly, use of information from the Migrant Worker Scan (MWS) - RAPID includes data from the MWS which identifies all non-UK nationals registering for a National Insurance number (NINo) from 1975 onwards. This gives us further information including, self-reported date of first arrival, NINo registration date, nationality at registration and previous country of residence. To ensure the most accurate information from the MWS is included DWP processed all MWS files from 2011 onwards to extract the earliest self-reported date of arrival.

Secondly, creating rules of residency in the UK - RAPID contains a combination of information on geographical location and activity with the underlying datasets as a proxy for inferring if someone was resident in the UK at any point during each tax year. For someone to be classed as resident in the tax year they must have at least one week of interactions with tax or benefit systems. In addition, they must also have at least one indicator showing they have a UK address. For those on State Pension, Bereavement Benefit or Widows Benefit this must not be paid abroad in a frozen rate country. Non-UK nationals who have registered for a NINo will be counted as resident from the tax year of their self-reported date of arrival.

Using all this information we are able to create the RAPID migration dataset by removing UK nationals and determining whether non-UK national records within the dataset are either short-term or long-term. For each non-UK national record, the RAPID migration dataset contains one row for each tax year since first arrival in the UK, containing a summary of the activities for employment (including self-employment), DWP and HM Revenue and Customs (HMRC) benefits, housing benefit and pensions as well as the self-reported date of first arrival in the UK. This longitudinal data allows us to assess patterns of interactions over time. Records are then categorised as either long-term or short-term by looking for patterns of interactions with the tax and benefits system (see Figure 1).

Our analysis has so far focused on long-term migration of non-UK nationals. Looking ahead, in the next phase of the development of ABME's we will be exploring how similar methods can be applied to estimate long-term migration of UK nationals. Results are reported in an article 'Developing our approach for producing admin based migration estimates', April 2021.

Identifying long-term migrants in the RAPID data

Both long-term and short-term migrants can be issued with a NINo, therefore the process of being issued with a NINo is not enough to indicate long-term migration into the UK. To determine long-term immigration of non-UK nationals we use a combination of data from the MWS showing when a NINo was issued alongside the "activity" within DWP and HMRC datasets. There are a series of steps taken in processing the data which form the main components in determining if arrivals are long-term. The main components to this processing are:

1. identifying migrants - migrants are identified from the MWS where any person registering for a NINo with a non-British nationality at the point of registration will be included
2. identifying first arrivals and registrations - first arrival and NINo registration date are obtained from the MWS
3. amalgamate all tax year datasets into a single dataset - in order to estimate a person's activity over time, all separate tax year data files in RAPID are amalgamated into a single dataset holding all tax years from 2010, or for those arriving after 2010, the tax year of first arrival
4. activity - RAPID captures interactions with DWP, HMRC and local authority systems and calculates the length of each interaction in the tax year (the number of weeks); these interactions therefore show that that person is "active" within the source systems and we therefore use this to show "activity" within the administrative data

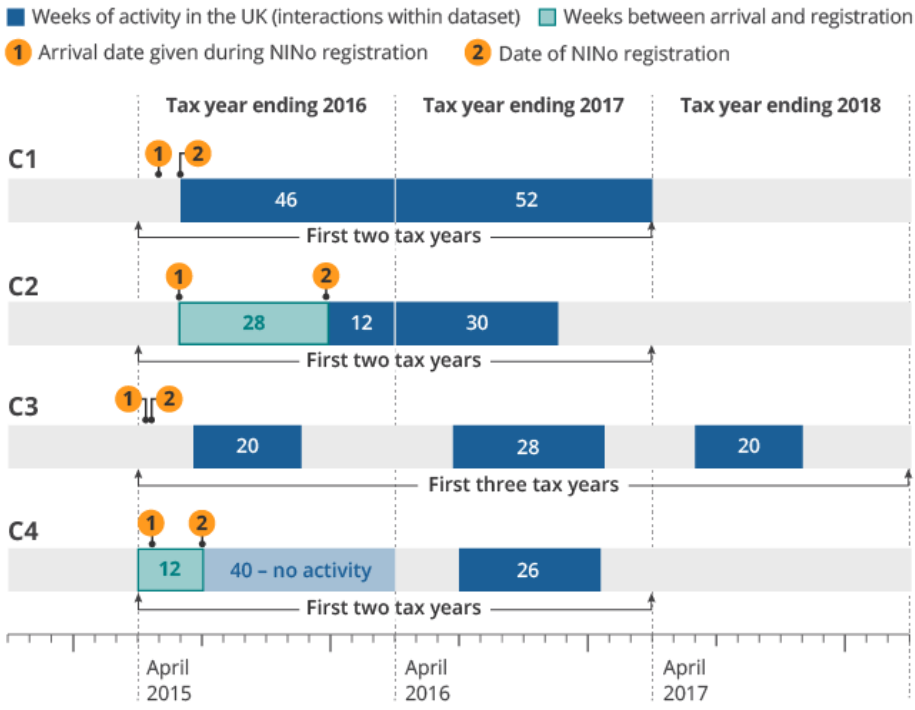
To estimate a person's total activity in the year, activities that cannot occur at the same time, such as employment and an out of work benefit, are added together. Other interactions may be allowed to overlap and therefore the maximum activity value is used. This gives a total number of weeks as an estimate of a person's total activity in the tax year. The number of weeks between the first arrival date and NINo registration date are also calculated. During this period the person did not have a NINo and therefore there are unlikely to be activities on the source systems covering this period.

Using all this information we can make an estimation of whether arrivals to the UK are long-term or short-term based on their activity profile in the RAPID data. All our research using administrative data so far has shown that people's lives are complex, therefore we have created multiple categories of long-term interactions to account for this.

We have created four categories defining patterns of activity of long-term arrivals. The first two categories most closely align with the UN definition of a long-term migrant whereby we are looking for sustained long-term interactions after arriving in the UK. It is important to note that the information contained within the RAPID data only specifies the total number of weeks of interactions within each tax year and does not specify that this activity is continuous, however, we have assumed the total activity measured to be sufficient to indicate long-term presence in the UK. These two categories make up the largest proportion of long-term arrivals in the RAPID data (over 90%). We have also included two further categories that expand on this definition of long-term activity, in order to reflect the complexity of people's lives, although it is important to note that each these groups only make up a small proportion of arrivals. These categories, which are outlined below, will be subject to further research and it's likely they will be refined as we learn more about how individuals interact with the administrative systems covered by RAPID. For example, Category 3 may capture circular migration or seasonal workers without the ability to distinguish between the two. The inclusion of monthly activity indicators in the RAPID data will allow for more nuanced interpretation of activity.

- **Category 1:** activities in the registration year and registration year plus one suggest they are resident for 52 weeks or more over that two-year period
- **Category 2:** the period between arrival and registration, plus the duration of activities in registration year and registration year plus one, suggest they are resident for 52 weeks or more
- **Category 3:** activity occurred in **three consecutive** years from registration (where registration is counted as an activity), and **where the 52-week activity criteria is not met** but where the activity profile suggests they are resident for 52 weeks or more over a two-year period
- **Category 4:** where the number of weeks between the registration date and the end of the tax year, plus the activity in the registration year plus one suggests they have been resident for 52 weeks or more; there must be at least one week of activity in the registration year plus one

Figure 1: Illustrative examples of identifying long-term international arrivals using RAPID data



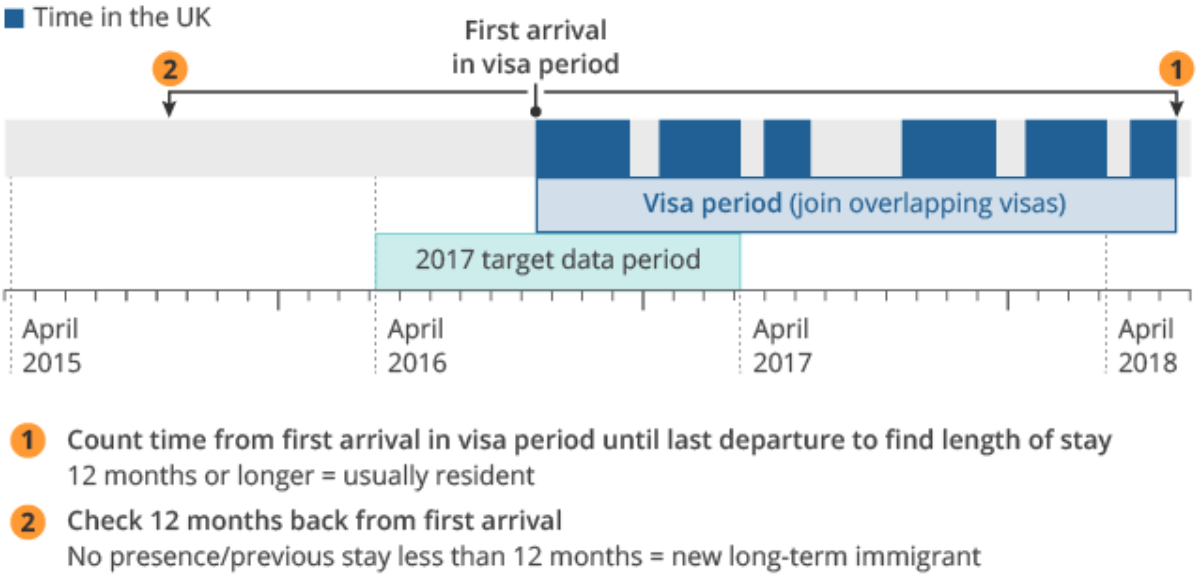
Identifying long-term immigrants in the Home Office visa and border data

The latest method looks at an individual's first arrival and last departure within a visa period as an approximation for length of stay in the UK. Visa periods are constructed by linking together any consecutive or concurrent visas held. If there is a gap between visas, then a new visa period is started. Visits from non-visa nationals and those on long term visit visas are excluded.

The process of estimating long-term international immigration for any given 12-month reference period is illustrated in Figure 2, and uses a three-step process:

1. Identifying those people who have a visa period with a first arrival date within the reference period.
2. Using the time between the first arrival and last departure within a visa period to identify whether they have been resident in the country for 12 months or more (i.e. whether they meet the usual residence threshold applied in the UN definition). This method means that short trips abroad over the course of an extended period of residence are excluded. If either the first arrival or last departure information is missing, then visa start or end dates are used as a proxy.
3. Looking at any previous visa period to determine if this is a new long-term immigrant or one who has previously been in the country. If no presence is identified in the country during the 12 months preceding first arrival on a given visa, or the previous visa period had a length of stay of less than 12 months, then this pattern of travel will be considered as identifying a new long-term immigrant.

Figure 2: Illustrative example of identifying long-term international immigration using border data



Further detail on methods used to derive international migration flows from RAPID data and Home Office data are published in the methodological paper '[Methods for measuring international migration using RAPID administrative data]'. Comparing estimates from RAPID to IPS estimates to assess the quality in measuring migration. Our report "developing our approach for producing admin-based migration estimates" presents high-level comparisons between estimates from the Registration and Population Interaction Database (RAPID) and previously published long-term international migration (LTIM) estimates. As part of building our understanding of RAPID and its measurement of long-term international migration, we completed further analysis comparing the findings to those from the International Passenger Survey (IPS). This includes demographic analysis of arrivals, departures and net migration for both EU and non-EU nationals. We also completed country level analysis of EU migration patterns to help understand what could be driving the trends seen in RAPID.

7. Review of current approaches

The modelling approach was developed as a tactical solution to address IPS data no longer being available to produce long-term international migration estimates after March 2020, and as a way of being able to incorporate a variety of sources including assumptions based on expert opinion. This was particularly important as the pandemic resulted in changing behaviours and affected some of the processes which underpin administrative sources. While we have accelerated the move to Admin-Based Migration Estimates (ABMEs), the data used in this work only cover the period up until the end of March 2020.

Estimating international migration using non-survey data sources is complex and challenging, even before the coronavirus pandemic. Definitionally, 12 months or more need to pass before we can assume if someone who has travelled to or from the UK at the start of that period is a migrant according to the long-established UN international definition[1]. Administrative data sources are not designed to measure international migration flows; often there are coverage issues, they are less timely and have inherent time lags due to (a) when an individual interacts with a service for example

and, (b) the definitional constraints for international migration. As described above, we are essentially repurposing and using these data as a proxy for international migration.

Table 3 summarises the advantages and disadvantages of the two approaches.

Table 3: Summary of advantages and disadvantages of the two approaches

	Advantages		Disadvantages	
	Model	**ABME**	**Model**	**ABME**
Data	Uses primarily aggregate level data, so can be timelier. Timely aggregate monthly data feeds using Home Office Border data on Non-EU nationals' arrivals and departures by visa route. Potential for EU going forward but possibly not until 2023. Air, ferry and train passenger data by month – travel patterns. Home Office Advanced Passenger Information on air arrivals and departures for GB, Non-UK and EU nationals Potential to include 'signal' data such as mobile phone, Facebook etc.	Uses mainly individual level data which potentially could be more accurate. Based on behaviours rather than intentions. Home Office Border data for Non-EU nationals' arrivals and departures. Potential for EU going forward but possibly not until 2023. RAPID potential rich source EU nationals working or claiming benefits. Granularity by geography, nationality, age/sex in RAPID data. Visa data on routes applied for/time in UK/out of UK. RAPID has potential to provide signal/patterns to feed into models for EU.	May lack sub-population levels detail if not able to be included in the aggregated data. Lacks equivalent data on EU and GB nationals' migration behaviours. Home Office data will exclude Irish Nationals. Known coverage issues for Common Travel Area (CTA).	Inherent lags in administrative data (definitionally and in timeliness of collection). If linking data sources, errors may begin to accumulate. Under-coverage in RAPID data: Under 16s, students, economically inactive/not claiming benefits or have recourse to public funds. Possibly includes those employed by UK organisations but residing overseas during pandemic. Need to consider remote working as an increasing number of people spend most or all their time in one country but working full time in another. Activity and presence may not be the same thing. Use of no activity as an indicator for emigration can be problematic for GB nationals. Home Office data will exclude Irish Nationals. Known coverage issues for Common Travel Area (CTA).
Methods	Models time-varying processes and their different features, such as trends, seasonality, and random components. Explicitly model the latent (unobserved) process and its (observed) measurements.	Aggregation/counts based on a set of rules to classify migrants that can be fairly easily explained to users. Flexibility to provide a range of breakdowns as available in the RAPID and Home Office datasets (e.g. for migrant workers – industry	Heavily assumption based for EU and GB nationals for immigration and emigration. Assumptions will need testing going forward. Evidence gathering and assumption setting process needs to be built into modelling timetable. More uncertainty in estimate as move further away from historic IPS timeseries. However, there is an opportunity to review how we can use IPS data as	Reliant on assumption that the hierarchical deterministic rules to classify long-term migrants in RAPID are accurate. No possibility of moving from long-term migrant to short-term. 'Activity' is used to establish residency (or not) in the UK. Adjustments are made for under-coverage, changing citizenship and right censoring in the data (not enough data to know outcome in latest year).

	Flexibility to include other data as they become available. Allows multiple usage of data sources, taking advantage of the statistical properties at different levels of granularity required for Mid-Year Population estimates	sector, earnings, nationality, age and sex) If the underpinning sources are stable, then the estimates should be coherent over time.	the survey was reinstated in January 2021. Going forward, assumption that the relationship between the covariates used in the model and the outcome variable continues to hold (unless exploratory analysis tells us otherwise). Requires upskilling and support for business area to operationalise. Will be difficulty to explain models to users with different levels of technical expertise and modelling experience	Home Office border data and RAPID data used independently of each other to produce estimates. Methods for uncertainty measurement for this type of approach have not yet been developed.
Outputs	Provisional timelier monthly estimates for UK immigration, emigration and derived net migration by nationality group. Uncertainty intervals provided for estimates derived from the model.	Granular annual estimates for UK immigration, emigration and derived net migration by nationality group. Possible to breakdown further by geography, age/sex Estimates for work migration by nationality group etc.	Estimates will be subject to later confirmation – User acceptability? Method needed for more granular outputs by geography (E&W, LA), age and sex	Not able to produce estimates by some migration streams e.g. study, family Timeliness of outputs, currently a 12+ month lag i.e. reporting 2019/20 in April 2021 – User acceptability (not timely enough for Mid-Year population estimates) Uncertainty needs careful explanation, as may not have a 'confidence interval'. A method is needed to produce some type of uncertainty measure for estimates.

8. Future iterations of international migration estimates

We refer to the two approaches we developed as a 'tactical' model-based approach to deliver timely ***provisional*** estimates with more uncertainty and a 'strategic' approach that uses administrative data to produce a set of ***final*** estimates with less uncertainty, but with a significant time lag. We will continue to develop both methods and explore data sources and incorporate them into both approaches if appropriate.

Our ambition is to bring together and therefore narrow the time between the provisional and final estimates if possible. To do this we aim to use modelling as the statistical framework around which we will develop international migration estimates. In section 9 we introduce the admin-based dynamic population model that aims to produce timely provisional England and Wales population estimates that will be at the core of the transformed population and migration system. Timely international migration estimates are critical to the dynamic admin-based population models and will feed into this system.

For future iterations of international migration estimates, we plan to produce UK totals for immigration, emigration and net for Quarters 3 and 4 2020 using the multivariate State Space Models (SSM) developed for Quarter 2. We will extend the covariate data to include up to the end of December 2020 and assume that the relationship between migration flows (the outcome variables) and the covariates continues to hold.

We will revisit Quarter 2 estimates and revise if needed in the light of additional data on migration behaviours. We acknowledge that the estimates for this period are provisional as our models develop and will be subject to retrospective confirmation and adjustment as data become available or mature. We anticipate that a tactical approach will be required until a steady state is reached when migration behaviours stabilise into a new normal and, administrative data systems reflect this.

Alongside this we will continue to explore how to make best use of the administrative data sources that we already have, e.g. RAPID and Home Office border data, as well as new sources as they become available.

We are considering how best to incorporate a RAPID data time series into our models given the current constraints and coverage issues. We are applying our error framework for longitudinally linked administrative data sources to optimise the use of RAPID data in our models, and to inform the quality of the strategic approach. We are also planning to explore the use of predictive machine learning models to classify non-UK and UK long-term migrants in the RAPID data to meet the need for timely initial estimates. Predictive models should mitigate the impact of the current time lag when using the outcomes-based approach to producing provisional international migration estimates. A similar approach has been used successfully by Statistics New Zealand to classify international migrants. This approach has used a three-step process to classify migrants which we plan to apply to the DWP RAPID data:

- deterministic classification of international migrants where we have complete certainty of the outcome.
- predictive model-based classification of 'uncertain' cases. Although these outcomes cannot be classified with complete certainty, information from RAPID data on activity could suggest what the eventual classification might be. We can exploit this information by training a statistical model on historical data, and then applying it on recent uncertain cases.
- aggregation of results and estimation of migration to produce provisional estimates. We will use the output from the model to generate distributions of possible outcomes for uncertain

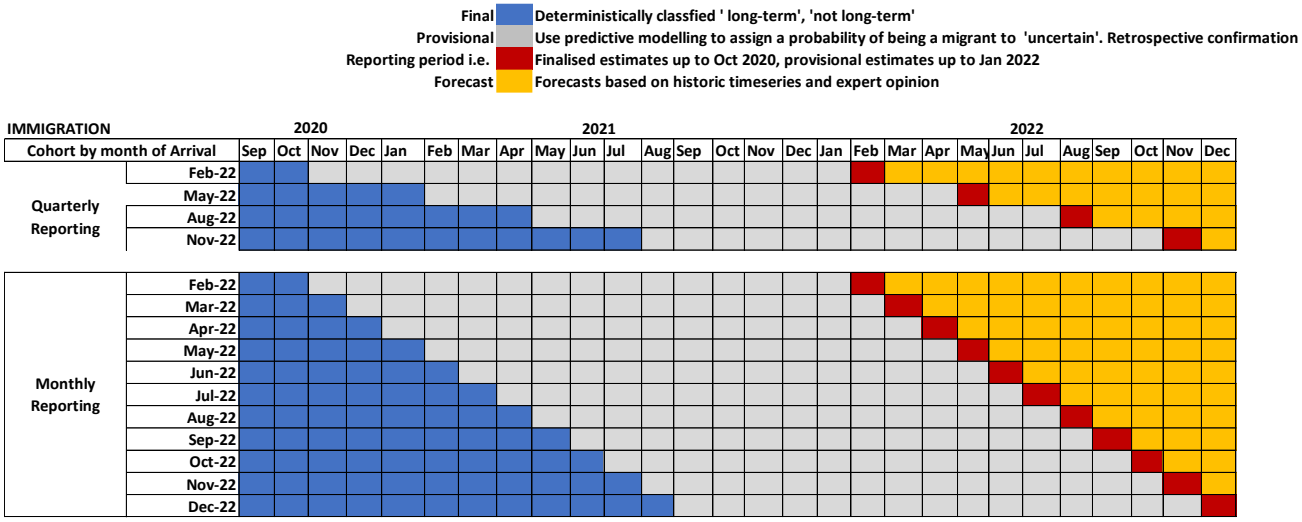
cases. By combining these distributions with the information on known outcomes, we can potentially generate estimates for different subsets or breakdowns (i.e. nationality, age/sex, geography) with measures of uncertainty.

Any provisional modelled estimates will have inherent uncertainty; data in more recent provisional periods will have higher uncertainty while older provisional periods will become more certain. Therefore, provisional estimates will need to be regularly revised until finalised after the reference period. Figure 1 below illustrates how this could look for reporting of quarterly or monthly modelled long-term immigration estimates from administrative data. We will be relying on predicting migration outcomes, given the definitional requirement for a migrant to be resident a year after arrival. Early outputs from the modelling will be provisional, subject to confirmation or correction as we follow up migrant outcomes and as data sources mature and become available to us.

For example, in February 2022 we would produce provisional estimates for the previous 15 months[6] and confirm estimates for October 2020 and earlier. If required, it would be possible to use our SSM methods to predict long-term international migration beyond the reporting period. These forecasts would be based on historic time series and expert opinion.

Clearly this would represent a significant change in the way that international migration statistics are produced and reported. We are working closely with stakeholders to ensure this is meeting their requirements and they understand the direction of travel.

Figure 3: Illustration of estimation and reporting window for long-term immigration estimates using RAPID or Home Office border data



Notes:

1. Assumption: 15-month lag allowed; key predictors available monthly.
2. Use decision rules to confirm travel and short-term visits based on time left in 16-month observation window.

[6] Assumes 15 months need to pass before a long-term immigration outcome is known. Longer may be necessary for emigration based on the current UN definition.

9. Integration with the future population, migration, and social statistics system

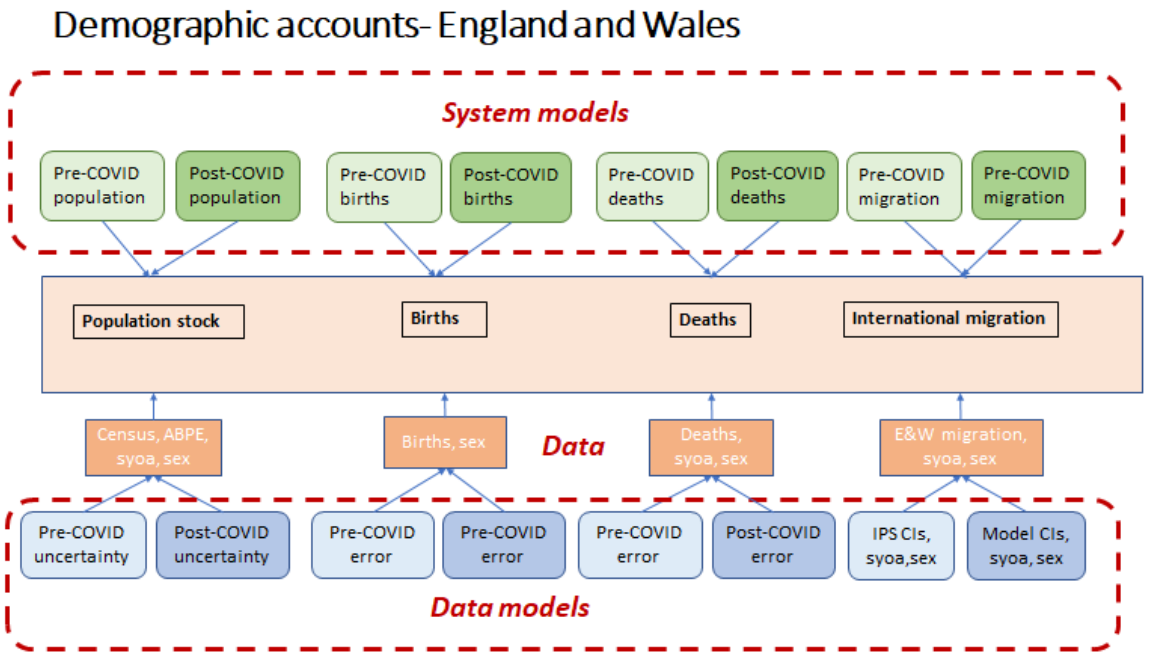
The pandemic and Brexit have put more focus on the importance of delivering the benefits of new data and transformed methods quickly and iteratively. We need to provide the best possible estimates to users as quickly as possible. We are seeking to harness the current opportunities to be ambitious in providing new responsive statistics, analysis and early insights on the size and characteristics of the population now and in future.

In addition, the pandemic has highlighted the importance of coherence and ensuring a 'one ONS' approach in how we are developing statistical methods. We need timely access to data to deliver the best possible statistics. The transformation strategy aims to bring coherence in how we address the immediate impact of the pandemic, our recovery from it, and our longer-term, sustainable system.

As a priority, we are pushing forward with developing an admin-based dynamic population model to produce timely provisional population estimates that are the critical core of the transformed population and migration system. This system will provide coherence across the Transformation Programme, delivering demographic accounts, initially provisional ones and then final ones, across time. These will form the benchmark and calibration tool for social surveys, to enable the production of coherent transformed social statistics, and for benchmarking the linked administrative and survey data being provided to support research through the Integrated Data Platform.

This dynamic population model is in the early stages of development and we anticipate this will be presented to the panel in due course. Figure 4 is a high-level representation of what the model would look like at the national (England & Wales) level.

Figure 4. Overview of the England and Wales Dynamic Population Model



10. Conclusion

This paper provides an update on our approach to estimate international migration using administrative data, summarises our current tactical and strategic approaches, and outlines our ambition to bring these two independent approaches together using modelling as a statistical framework. We also set out the challenges of using non-survey data to estimate international migration and assess the advantages and disadvantages of the two approaches we have taken so far.

We welcome discussion with the panel on our approach and seek feedback on:

- the tactical and strategic approaches to estimating international migration set out in Sections 5 and 6
- our proposals to bring together the two approaches for future iterations of international migration estimates (ABMEs) outlined in Section 8 to inform our submission to the Methodological Assurance Review Panel this month, of which this paper is the basis.

And, for the panel to consider the following questions specific to our approach:

  - Should we pursue this approach for estimating international migration?
  - Are there any other approaches we should be considering? For example, co-operation with other National Statistical agencies who hold good data.
  - Do you have any specific comments on the strategic and tactical approaches?
  - Do you have any suggestions for bringing together the tactical and strategic approaches?
  - Are you aware of similar work?

As our work develops, we plan to bring future iterations to this panel for discussion and feedback on the statistical design and methodological considerations.

11. References

Office for National Statistics (2021) Methods for measuring international migration using RAPID administrative data, 16 April 2021

Office for National Statistics (2021) International migration: developing our approach for producing admin-based migration estimates, 16 April 2021

Office for National Statistics (2021) Using statistical modelling to estimate UK international migration, 16 April 2021

Office for National Statistics (2020) Migration Statistics Quarterly Report: August 2020, 27 August 2020

Office for National Statistics (2020) Population and migration statistics system transformation – recent updates: 3. *Update on analysis to explore how the DWP's RAPID could be used to measure migration*, 21 May 2020

Office for National Statistics (2020) Defining and measuring international migration, 14 February 2020

Office for National Statistics (2020) ONS Working Paper Series No 19 – An error framework for longitudinal administrative sources; its use for understanding the statistical properties of data for international migration, 14 February 2020

Office for National Statistics (2019) Update on our population and migration statistics transformation journey: a research engagement report *8. Data sources: what can we use to measure population and migration?*

Statistics New Zealand (2018) Update on the development of provisional external migration estimates. Retrieved from www.stats.govt.nz

Further references used in the development of time series models:

Bijak J (2016) Migration forecasting: Beyond the limits of uncertainty. IOM Global Migration Data Analysis Centre, Data Briefing 6, 29 November 2016.

Bijak J, Disney G, Findlay AM, Forster JJ, Smith PWF and Wiśniowski A (2019) Assessing time series models for forecasting international migration: Lessons from the United Kingdom. Journal of Forecasting, 38(5), pages 470 to 487.

Disney G (2015) Model-based estimates of UK immigration. PhD Thesis, University of Southampton.

Durbin, J., and Koopman, S. (2001). Time series analysis by state space methods. Oxford: Clarendon Press.

Eurostat (2020) Guidance on time series treatment

Helske, Jouni (2017). KFAS: Exponential Family State Space Models in R. Journal of Statistical Software, 78(10), pages 1 to 39.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raymer J, Wiśniowski A, Forster J J, Smith PWF and Bijak J (2013) Integrated Modeling of European Migration. Journal of the American Statistical Association, 108(503), pages 801 to 819.

Wiśniowski A, Bijak J, Christiansen S, Forster JJ, Keilman N, Raymer J and Smith PWF (2013) Utilising expert opinion to improve the measurement of international migration in Europe. Journal of Official Statistics, 29(4), pages 583 to 607.

Wiśniowski A, Forster JJ, Smith PWF, Bijak J and Raymer J (2016) Integrated modelling of age and sex patterns of European migration. Journal of the Royal Statistical Society, Series A, 179(4), pages 1,007 to 1,024.

Further references used in the development of the admin-based dynamic population model

Bryant, John & Graham, Patrick. (2013). Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources. Bayesian Analysis. 8. 10.1214/13-BA820.

Bryant, John & Zhang, Junni L. Zhang (2018) Bayesian Demographic Estimation and Forecasting. Chapman and Hall/CRC

Appendix 1: Mathematical specification of the multivariate State Space Models and model diagnostics

This appendix contains the details of the mathematical specification of the multivariate State Space Model (SSM).

Purpose of the model

The multivariate SSM is fitted to the historical International Passenger Survey (IPS), and Home Office border crossing and travel data. Predictions are based on an extrapolation of the trend and seasonal components, together with the Home Office border crossing and travel data for the period March to June 2020. Predictions are generated from the model for British (GB), EU and non-EU immigration and emigration, and are summed to estimate the corresponding totals.

Data

IPS estimates for GB, EU and non-EU immigration and emigration, for the period January 2010 to February 2020. Note that we set January 2020 as an outlier by treating it as missing in the implementation.

Home Office border crossing (Exit Checks) data provide the number of incoming and outgoing non-EU migrants with a visa (of all types, be it for work, study, or family and other purposes). They are used in the models from March 2020 until June 2020.

The proportion of incoming and outgoing travellers using ferries or trains, from February 2020 until September 2020. This is calculated as the number of travellers using the ferry and train divided by the number of travellers using ferries, trains and flights (as identified in Civil Aviation Authority (CAA) data).

SSM equations

The multivariate SSM is a combination of a trend, seasonal, and an error component, for each IPS time series and for the administrative time series. SSMs consist of two equations:

1. The observation equation tells you the relationship between the observed vector of time series (IPS estimates and administrative time series) and the unobserved components of these series (trend, seasonal, and other components) contained within the state vector.
2. The transition equation tells you how the values of the components in the state vector change from one time-period to the next; in this case, the periodicity of the data is monthly.

Observation equations

Each observed time series y_t^i is the observed value of the outcome variable i, for $i \in \{GB, EU, nonEU, EXa, EXp, EX\}$, at time t, where EX refers to Exit Checks data , EXp to Exit Checks data adjusted by the proportion of travel by train and sea, and EXa refers to Exit Checks data for the opposite direction of movement for the IPS data. For example, if the IPS outcome variables are immigration (emigration) then EXa refers to Exit Checks data for departures (arrivals), EXp refers to

Exit Check data for arrivals (departures) adjusted by the proportion of arrivals (departures) by air and sea, and EX refers to Exit Checks for arrivals (departures). The trend component is denoted μ_t^i, and γ_t^i is the seasonal component.
The observation equation for the log of the IPS estimate for immigration (emigration), for stream $s \in \{GB, EU, nonEU\}$, in month t, denoted y_t^s, is

$$y_t^s = \mu_t^s + \gamma_t^s + \varepsilon_t^s \tag{1}$$

where $\varepsilon_t^s \sim N(0,\ k_{s,t}^2 \sigma_{\varepsilon_s}^2)$ is an error component that accounts for the IPS sampling error ($k_{s,t}^2$ is the estimated variance of the IPS estimate in month t for stream s), and $\sigma_{\varepsilon_s}^2$ is to be estimated (but expected to be 1). It is also assumed that the sampling errors of the three IPS series are contemporaneously correlated. There are three of these equations for each direction of migration.

The remaining administrative based time series in the observation equation, $a \in \{EXa, EXp, EX\}$ are also log transformed and modelled as:

$$y_t^a = \mu_t^a + \gamma_t^a + \epsilon_t^a \tag{2}$$

where $\epsilon_t^a \sim N(0,\ \sigma_{\epsilon^a}^2)$ is an error component. These error terms are assumed to be independent of one another.

Transition equations

The unobserved components in the state vector evolve over time, with the following equations that are then incorporated into the transition equation. The trend components are as follows:

$$\mu_{t+1}^{GB} = \mu_t^{GB} + \nu_t^{GB} + r_t \eta_t^{\mu_{EXa}} \tag{3}$$

$$\nu_{t+1}^{GB} = \nu_t^{GB} + \eta_t^{\nu_{GB}} \tag{4}$$

$$\mu_{t+1}^{EU} = \mu_t^{EU} + \nu_t^{EU} + r_t' \eta_t^{\mu_{EXp}} + (r_t - r_t') \eta_t^{\mu_{EX}} \tag{5}$$

$$\nu_{t+1}^{EU} = \nu_t^{EU} + \eta_t^{\nu_{EU}} \tag{6}$$

$$\mu_{t+1}^{nonEU} = \mu_t^{nonEU} + \nu_t^{nonEU} + r_t \eta_t^{\mu_{EX}} \tag{7}$$

$$\nu_{t+1}^{nonEU} = \nu_t^{nonEU} + \eta_t^{\nu_{nonEU}} \tag{8}$$

$$\mu_{t+1}^{EXa} = \mu_t^{EXa} + \nu_t^{EXa} + r_t \eta_t^{\mu_{EXa}} \tag{9}$$

$$\nu_{t+1}^{EXa} = \nu_t^{EXa} + \eta_t^{\nu_{EXa}} \tag{10}$$

$$\mu_{t+1}^{EXp} = \mu_t^{EXp} + \nu_t^{EXp} + r_t \eta_t^{\mu_{EXp}} \tag{11}$$

$$\nu_{t+1}^{EXp} = \nu_t^{EXp} + \eta_t^{\nu_{EXp}} \tag{12}$$

$$\mu_{t+1}^{EX} = \mu_t^{EX} + \nu_t^{EX} + r_t \eta_t^{\mu_{EX}} \tag{13}$$

$$\nu_{t+1}^{EXp} = \nu_t^{EX} + \eta_t^{\nu_{EX}} \tag{14}$$

where $\eta_t^{\mu_a} \sim \mathrm{N}\left(0, \sigma_{\mu_a}^2\right)$, for $a \in \{EXa, EXp, EX\}$, and $\eta_t^{\nu_i} \sim \mathrm{N}(0, \sigma_{\nu_i}^2)$ for $i \in \{GB, EU, nonEU, EXa, EXp, EX\}$. For immigration

$$r_t = \begin{cases} 0, & t < \text{March 2020} \\ 1, & \text{otherwise} \end{cases} \text{ and } r_t' = \begin{cases} 0, & t < \text{April 2020} \\ 1, & \text{otherwise} \end{cases},$$

while for emigration

$$r_{\mathrm{t}} = \begin{cases} 0, & t < \text{February 2020} \\ 1, & \text{otherwise} \end{cases} \text{ and } r_t' = \begin{cases} 0, & t < \text{April 2020} \\ 1, & \text{otherwise} \end{cases}.$$

These equations say that the trend component in month $t + 1$ is equal to that in month t plus a variable ν_t with error, plus the addition of an error term during the coronavirus (COVID-19) period, (determined by r_{t} and r_t'), $\eta_t^{\mu_a}$ for $a \in \{EXa, EXp, EX\}$. The latter term allows us to capture potentially large step changes in migration because of the coronavirus. On the original scale of the data, they are proportional modifications to what might have been expected had COVID-19 (and other effects during this period such as effects from the EU exit - it is not possible to distinguish between them) not happened. The proportional modifications are estimated from the large movements in the Exit Checks data time series.

The proportional modifications estimated by the unadjusted Exit Checks time series (EX) are used to modify the non-EU predictions from March (February) onwards for immigration (emigration) and also to modify the EU prediction for March (February and March) only for immigration (emigration). The proportional modifications estimated by the adjusted Exit Checks time series (EXp) are used to modify the EU predictions from April onwards. Proportional modifications estimated by Exit Checks data for departures (EXa) are used to model immigration by British nationals (GB) while those estimated by unadjusted Exit Checks arrivals are used to modify GB emigration. This is because we believe that the behaviour of GB nationals migrating back to the UK may be similar to visa holders' early departures from the UK for home. Conversely, emigration by GB nationals is likely to be constrained, as evidenced in the deferred or cancelled arrivals seen for visa holders in the Exit Checks data.

A summary of proportional modifications and their corresponding equations.

Proportional modifications that multiply the predictions for IPS migration in the coronavirus period, and reference to the corresponding equations:

Immigration

- British (GB): unadjusted Exit Checks departures; see equations (3) and (9)
- EU: Unadjusted Exit Checks arrivals for March 2020 and adjusted Exit Checks arrivals from April onwards; see equations (5), (11), (13)
- Non-EU: Unadjusted Exit Checks arrivals; see equations (7), (13)

Emigration

- British (GB): unadjusted Exit Checks arrivals; see equations (3) and (9)
- EU: Unadjusted Exit Checks departures for March 2020 and adjusted Exit Checks arrivals from April onwards; see equations (5), (11), (13)
- Non-EU: Unadjusted Exit Checks departures; see equations (7), (13)

The seasonal components are modelled as:

$$\gamma_t^i = \sum_{j=1}^{6} \gamma_{j,t}^i \tag{15}$$

$$\gamma_{j,t+1}^i = \gamma_{j,t}^i \cos\left(\frac{2\pi j}{12}\right) + \gamma_{j,t}^{i*} \sin\left(\frac{2\pi j}{12}\right) + \omega_{j,t}^i, \qquad j = 1,2,\dots 6 \tag{16}$$

$$\gamma_{j,t+1}^{i*} = -\gamma_{j,t}^i \sin\left(\frac{2\pi j}{12}\right) + \gamma_{j,t}^{i*} \cos\left(\frac{2\pi j}{12}\right) + \omega_{j,t}^{i*}, \qquad j = 1,2,\dots 6 \tag{17}$$

for $i \in \{GB, EU, nonEU, EXa, EXp, EX\}$, where $\gamma_{j,t}^i$ is the effect of season j at time t for series i, $\omega_{j,t}^i, \omega_{j,t}^{i*} \sim N\left(0, \sigma_{\omega_i}^2\right)$. Note that both equations (16) and (17) are sets of 6 equations.

Matrix formulation

Model equations (1) to (17) can be written in matrix format – the required formulation for implementation in R. In the following notation, bold uppercase letters denote a matrix, bold lowercase letters denote a vector, and regular fonts denote a scalar (single value). Equation (18) is the matrix formulation of the observation equations, (1) and (2), and equation (19) is that of the transition equations (3) to (17). Note that the transition matrix $\boldsymbol{T}$ does not have a subscript t like the other terms because it is not time-varying.

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t \tag{18}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{R}_t\, \boldsymbol{\eta_t} \tag{19}$$

where

$$\boldsymbol{y}_t = \begin{pmatrix} y_t^{GB} \\ y_t^{EU} \\ y_t^{nonEU} \\ y_t^{EXa} \\ y_t^{EXp} \\ y_t^{EX} \end{pmatrix}, \boldsymbol{Z}_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} & \boldsymbol{z^\gamma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & k_t^{GB} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \boldsymbol{z^\gamma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & k_t^{EU} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z^\gamma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & k_t^{nonEU} \\ 0 & 0 & 0 & 1 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z^\gamma} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z^\gamma} & \mathbf{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z^\gamma} & 0 & 0 & 0 \end{pmatrix}$$

where $\boldsymbol{z^\gamma} = (1,0,1,0,1,0,1,0,1,0,1)$, each $\mathbf{0}$ (bold font) in the seventh column represents a row vector of length 6, and those in columns eight to thirteen represent a row vector of length 11, and

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \mu_t^{GB} \\ \mu_t^{EU} \\ \mu_t^{nonEU} \\ \mu_t^{EXa} \\ \mu_t^{EXp} \\ \mu_t^{EX} \\ v_t^{GB} \\ v_t^{EU} \\ v_t^{nonEU} \\ v_t^{EXa} \\ v_t^{EXp} \\ v_t^{EX} \\ \boldsymbol{\gamma_t^{GB}} \\ \boldsymbol{\gamma_t^{EU}} \\ \boldsymbol{\gamma_t^{nonEU}} \\ \boldsymbol{\gamma_t^{EXa}} \\ \boldsymbol{\gamma_t^{EXp}} \\ \boldsymbol{\gamma_t^{EX}} \\ \varepsilon_t^{GB} \\ \varepsilon_t^{EU} \\ \varepsilon_t^{nonEU} \end{pmatrix}, \text{ where } \boldsymbol{\gamma_t^i} = \begin{pmatrix} \gamma_{1,t}^{i} \\ \gamma_{1,t}^{*i} \\ \gamma_{2,t}^{i} \\ \gamma_{2,t}^{*i} \\ \vdots \\ \gamma_{5,t}^{i} \\ \gamma_{5,t}^{*i} \\ \gamma_{6,t}^{i} \end{pmatrix}, \text{ for } i \in \{GB, EU, nonEU, EXa, ECp, EX\},$$

$$\boldsymbol{\epsilon}_t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \epsilon_t^{EXa} \\ \epsilon_t^{EXp} \\ \epsilon_t^{EX} \end{pmatrix}, \text{ where } \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{H_t}) \text{ and covariance matrix } \boldsymbol{H_t} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\epsilon^{EXa}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\epsilon^{EXp}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\epsilon^{EX}} \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{I_6} & \boldsymbol{I_6} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I_6} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T_\lambda} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T_\lambda} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T_\lambda} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T_\lambda} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T_\lambda} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T_\lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0_3} \end{pmatrix}$$

where $\boldsymbol{T}_\lambda = \begin{pmatrix} \cos\lambda_1 & \sin\lambda_1 & 0 & 0 & \boldsymbol{0} & 0 & 0 & 0 \\ -\sin\lambda_1 & \cos\lambda_1 & 0 & 0 & \boldsymbol{0} & 0 & 0 & 0 \\ 0 & 0 & \cos\lambda_2 & \sin\lambda_2 & \boldsymbol{0} & 0 & 0 & 0 \\ 0 & 0 & -\sin\lambda_2 & \cos\lambda_2 & \boldsymbol{0} & 0 & 0 & 0 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \ddots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ 0 & 0 & 0 & 0 & \boldsymbol{0} & \cos\lambda_5 & \sin\lambda_5 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{0} & -\sin\lambda_5 & \cos\lambda_5 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{0} & 0 & 0 & \cos\lambda_6 \end{pmatrix}$ and $\lambda_j = \frac{2\pi j}{12}$

$$\boldsymbol{R}_t = \begin{pmatrix} 0 & 0 & 0 & r_t & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & r_t^{'} & (r_t - r_t^{'}) & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & r_t & \mathbf{0} \\ 0 & 0 & 0 & r_t & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & r_t & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & r_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{75} \end{pmatrix}, \boldsymbol{\eta_t} = \begin{pmatrix} \eta_t^{\mu_{GB}} \\ \eta_t^{\mu_{EU}} \\ \eta_t^{\mu_{nonEU}} \\ \eta_t^{\mu_{EXa}} \\ \eta_t^{\mu_{EXp}} \\ \eta_t^{\mu_{EX}} \\ \eta_t^{v_{GB}} \\ \eta_t^{v_{EU}} \\ \eta_t^{v_{nonEU}} \\ \eta_t^{v_{EXa}} \\ \eta_t^{v_{EXp}} \\ \eta_t^{v_{EX}} \\ \omega_{1,t}^{GB} \\ \omega_{1,t}^{*GB} \\ \omega_{2,t}^{GB} \\ \omega_{2,t}^{*GB} \\ \vdots \\ \omega_{5,t}^{GB} \\ \omega_{5,t}^{*GB} \\ \omega_{6,t}^{GB} \\ \boldsymbol{\omega_t^{EU}} \\ \boldsymbol{\omega_t^{nonEU}} \\ \boldsymbol{\omega_t^{EXa}} \\ \boldsymbol{\omega_t^{EXp}} \\ \boldsymbol{\omega_t^{EX}} \\ \varepsilon_t^{GB} \\ \varepsilon_t^{EU} \\ \varepsilon_t^{nonEU} \end{pmatrix}$$

where $\boldsymbol{\eta_t} \sim N(0, \boldsymbol{Q_t})$ and the 81 x 81 covariance matrix

$$\boldsymbol{Q_t} = \begin{pmatrix} \boldsymbol{\sigma_\mu^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma_v^2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_\omega^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_\varepsilon^2} \end{pmatrix}$$

where

$$\boldsymbol{\sigma_{\mu}^2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\mu_{EXa}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\mu_{EXp}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\mu_{EX}} \end{pmatrix},$$

$$\boldsymbol{\sigma_{\nu}^2} = \begin{pmatrix} \sigma^2_{\nu_{GB}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\nu_{EU}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{nonEU} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\nu_{EXa}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\nu_{EXp}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\nu_{EX}} \end{pmatrix}$$

$$\boldsymbol{\sigma_{\omega}^2} = \begin{pmatrix} \boldsymbol{\sigma^2_{\omega_{GB}}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma^2_{\omega_{EU}}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma^2_{\omega_{nonEU}}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma^2_{\omega_{EXa}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma^2_{\omega_{EXp}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma^2_{\omega_{EX}}} \end{pmatrix} \text{ where } \boldsymbol{\sigma^2_{\omega_i}} = \sigma^2_{\omega_i}\boldsymbol{I_{11}},$$

$$\boldsymbol{\sigma_{\varepsilon}^2} = \begin{pmatrix} \sigma^2_{\varepsilon_{GB}} & \sigma_{\varepsilon_{GB}}\sigma_{\varepsilon_{EU}} & \sigma_{\varepsilon_{GB}}\sigma_{\varepsilon_{nonEU}} \\ \sigma_{\varepsilon_{GB}}\sigma_{\varepsilon_{EU}} & \sigma^2_{\varepsilon_{EU}} & \sigma_{\varepsilon_{EU}}\sigma_{\varepsilon_{nonEU}} \\ \sigma_{\varepsilon_{GB}}\sigma_{\varepsilon_{nonEU}} & \sigma_{\varepsilon_{EU}}\sigma_{\varepsilon_{nonEU}} & \sigma^2_{\varepsilon_{nonEU}} \end{pmatrix}.$$

Note that the matrix $\boldsymbol{\sigma_{\varepsilon}^2}$ has off-diagonal terms because it is assumed that the sampling errors of the three IPS series are contemporaneously correlated.

Model implementation

The model is implemented and fitted in R (R Core Team, 2020). First, the variances and covariances of the error terms in the observation and transition equation are estimated using maximum likelihood estimation with the Broyden–Fletcher–Goldfarb–Shanno algorithm (an iterative method for solving unconstrained nonlinear optimization problems) and the R optim function (does general-purpose optimization), through the Kalman Filter and Smoother (KFAS) package for state space modelling (Helske, 2017). Then the state vector, $\boldsymbol{\alpha}_t$, and the residuals, $\boldsymbol{\eta_t}$ and $\boldsymbol{\epsilon}_t$ are estimated using the KFAS. Kalman filtering is an algorithm that estimates state variables that cannot be measured or observed with accuracy, and their uncertainties using a weighted average, with more weight being given to estimates with higher certainty. The initial values are set to zero with an infinite variance, which in terms of the Kalman Filter algorithm is referred to as exact diffusion initialisation. The exception is for components $\varepsilon_t^{GB}, \varepsilon_t^{EU}, \varepsilon_t^{nonEU}$ which have initial mean states of zero and variance 1, since these are stationary and we expect the variance to be 1.

Uncertainty intervals

Prediction intervals on the original scale of the data are based on the back transformation of prediction intervals from the model, where those predictions have been made on the logarithmic-scale (for log-transformed variables). The uncertainty of the predictions on the logarithmic scale accounts for the uncertainty in the estimation of the unobserved components in the state vector, given the prior distribution of the initial estimates. Uncertainty associated with the estimation of the covariances of the observation and transition error terms are not accounted for, which means that these intervals are likely to be underestimated.

Model diagnostics

The following set of diagnostic statistics and their corresponding p-values is provided for the standardised residuals for the fitted models: skewness (S), kurtosis (K), test for normality (N), test for Heteroscedasticity (H), and lags k for which Ljung-Box Q(k) is significant at the 5% level. See Durbin and Koopman (2001, Section 2.12.1) for details.

Table 1 gives the diagnostic statistics for immigration. There is some evidence of skewness in the distribution of the standardised residuals for IPS GB and EU, and some kurtosis in IPS all. The normality tests also suggest a lack of normality in IPS all, GB and EU. There is some evidence of residual autocorrelation for lags 2-24 for Exit Checks departures and lags 2-3 for Exit Checks arrivals. Some heteroscedasticity is indicated in IPS All. The Doornik-Hansen test for multivariate normality (test statistic 22.2, p = 0.07), provides no evidence to reject the null hypothesis that the data are multivariate normal.

Table 1: Diagnostic statistics for the standardised residuals from the fitted models for immigration

title	Series	S	K	N	H(11)	lags Q(k) at 5%
Test statistic	IPS All	0.679	5.5273	12.0051	3.196	
p-value	IPS All	0.101	0.0023	0.0025	0.033	
Test statistic1	IPS GB	-0.889	4.5412	8.0767	1.487	
p-value1	IPS GB	0.032	0.0627	0.0176	0.261	
Test statistic2	IPS EU	0.979	3.7583	6.4309	2.298	
p-value2	IPS EU	0.018	0.3598	0.0401	0.092	
Test statistic3	IPS Non-EU	0.730	3.5220	3.5027	1.343	
p-value3	IPS Non-EU	0.078	0.5284	0.1735	0.317	
Test statistic4	EC Departures	0.073	1.9918	1.5136	1.227	2 to 24
p-value4	EC Departures	0.859	0.2234	0.4692	0.370	
Test statistic5	EC Arrivals Ferry, Train adjusted	0.231	2.1883	1.2710	0.746	

p-value5	EC Arrivals Ferry, Train adjusted	0.578	0.3270	0.5297	0.317	
Test statistic6	EC Arrivals	-0.377	2.3180	1.5054	0.909	2 to 3
p-value6	EC Arrivals	0.363	0.4102	0.4711	0.438	

Table 2 gives the diagnostic statistics for emigration. There is some skewness in the distribution of the standardised residuals for IPS all and GB, but no evidence of kurtosis, non-normality or heteroscedasticity. There is some evidence of residual autocorrelation for lag 22 for IPS non-EU. The Doornik-Hansen test for multivariate normality (test statistic 19.0, p = 0.17), provides no evidence to reject the null hypothesis that the data are multivariate normal.

Table 2: Diagnostic statistics for the standardised residuals from the fitted models for emigration

title	Series	S	K	N	H(11)	lags Q(k) at 5%
Test statistic	IPS All	0.046	2.53	0.34	1.321	
p-value	IPS All	0.912	0.57	0.84	0.326	
Test statistic1	IPS GB	0.036	2.20	0.94	1.412	
p-value1	IPS GB	0.930	0.33	0.62	0.289	
Test statistic2	IPS EU	0.185	2.13	1.29	2.754	
p-value2	IPS EU	0.654	0.30	0.52	0.054	
Test statistic3	IPS Non-EU	0.372	2.56	1.09	0.641	22 to 22
p-value3	IPS Non-EU	0.369	0.60	0.58	0.236	
Test statistic4	EC Arrivals	-0.310	2.97	0.56	1.065	
p-value4	EC Arrivals	0.454	0.97	0.75	0.459	
Test statistic5	EC Departures Ferry, Train adjusted	-0.312	3.10	0.58	0.765	
p-value5	EC Departures Ferry, Train adjusted	0.451	0.90	0.75	0.332	
Test statistic6	EC Departures	0.055	1.94	1.65	1.358	
p-value6	EC Departures	0.895	0.20	0.44	0.310	

Out of sample forecasting results for the benchmark model

The Mean Absolute Percentage Error (MAPE), in Tables 3 and 4, provides an indication of the performance of the benchmark model, which consists solely of seasonal and trend components. For each year, 2017 to 2019, we forecast four months ahead from January, from May and then from September, and then we calculate the MAPEs for the whole year (last three columns) based on these forecasts. The second column (2017 to 2019) is the MAPEs for all three years.

Table 3 shows that the forecast for total immigration is better than for any of the sub-series alone, which suggests that some of the errors in the latter are cancelled out in the sum. The forecasts for non-EU were more accurate than for GB and EU, with GB being the worst overall. The model is the most accurate for 2017, and the least accurate for 2018. Table 4 shows that for emigration the errors appear to cancel out again in the sum. The forecasts for GB were more accurate than for EU, but like those for non-EU.

Table 3: Mean Absolute Percentage Errors (MAPEs) for immigration

Series	2017- 2019	2017	2018	2019
	%	%	%	%
All	19	12	25	22
GB	54	24	93	45
EU	51	34	76	44
Non-EU	26	23	25	28

Table 4: Mean Absolute Percentage Errors (MAPEs) for emigration

Series	2017- 2019	2017	2018	2019
	%	%	%	%
All	21	11	30	22
GB	28	21	25	37
EU	50	22	78	51
Non-EU	28	28	26	30