EUROPEAN VISA DATABASE

DATABASE CODEBOOK

30.05.2013

Version 3.0

Table of contents

1. Introduction

This codebook describes in detail how the European Visa Database (EVD) was constructed.[1] The purpose is to provide full transparency of the coding process and thereby ensure a high reliability of the dataset. It is structured as follows. The next and second part sets out how the basic tables in the database on countries, cities, travel distances and population sizes were put together. The third, fourth and fifth describes the coding of the three dimensions of visa policies: requirements, issuing practice and consular services. The sixth part explains how the final combined data-table on visa requirements, visa issuing practices and consular representation, including the mobility barriers index, was devised. The seventh present a set of equations used and the final eighth part contains a technical diagram of the database structure.

All raw data, coding schemes as well as the database itself are available as a downloadable file from the EVD website. In this way, it is possible to inspect precisely how the dataset was compiled and organized in order to verify, repeat or modify the process. On a technical note, the downloadable file contains the ASP.NET computer scripts, SQL syntaxes, original and processed data as well as the database (MSSQL mdf file). This file also contains, where applicable, links to the raw data sources. To replicate the coding process a Microsoft SQL Server and Microsoft Internet Information Service running ASP.NET is required. The latter is available on most Windows computers whereas the former is specialized software which needs to be purchased separately. To repeat the conversion of pdf documents to excel it is also necessary to acquire software for doing so. For the project I used ABBYY's PDF Transformer 3.0.

In general, the time period covered by the database is 2005 to 2013. However, as data availability varies for the three dimensions (visa requirements, issuing practices and consular services) and receiving countries the years covered can in practice differ. The table below details the years included per receiving state and analytical dimension. For each year data is available for all relevant sending countries. The database is expanded with additional years and receiving countries as new data becomes available.

[1] The database can be accessed via www.mogenshobolth.dk/evd.

Receiving country	Visa requirements	Visa issuing practice	Consular representation
Austria	2001-2012	2005-2012 [1]	2004-2012
Belgium	2001-2012	2005-2012	2004-2012
Bulgaria	2007-2012	2006-2010	2006-2010 [2]
Cyprus	2004-2012	2005-2010	2005-2010 [2]
Czech Republic	2004-2012	2005-2012	2005-2012 [3]
Denmark	2001-2012	2005-2012	2004-2012
Estonia	2004-2012	2005-2012	2005-2012 [3]
Finland	2001-2012	2005-2012	2004-2012
France	2001-2012	2005-2012	2004-2012
Germany	2001-2012	2005-2012	2004-2012
Greece	2001-2012	2005-2012	2004-2012
Hungary	2004-2012	2005-2012	2005-2012 [3]
Iceland	2001-2012	2005-2012	2004-2012
Italy	2001-2012	2005-2012	2004-2012
Latvia	2004-2012	2005-2012	2005-2012 [3]
Lithuania	2004-2012	2005-2012	2005-2012 [3]
Luxembourg	2001-2012	2005-2012	2004-2012
Malta	2004-2012	2006-2012	2005-2012 [3]
Netherlands	2001-2012	2005-2012	2004-2012
Norway	2001-2012	2005-2012 [4]	2004-2012
Poland	2004-2012	2005-2012	2005-2012 [3]
Portugal	2001-2012	2005-2012	2004-2012
Romania	2007-2012	2006-2010	2006-2010 [2]
Slovakia	2004-2012	2005-2012	2005-2012 [3]
Slovenia	2004-2012	2005-2012	2005-2012 [3]
Spain	2001-2012	2005-2012	2004-2012
Sweden	2001-2012	2005-2012	2004-2012
Switzerland	2009-2012	2009-2012	2008-2012
United Kingdom	2003-2012	2001-2008	2001-2008 [2]
United States	2001-2012	2006-2012	No data

[1] Information on visas refused missing for 2007 and 2008

[2] Information on consular representation is available via the data on cities where visas were issued

[3] Information on consular representation in 2006 and 2007 is available via the data on cities where visas were issued

[4] Information on visas refused missing for 2009

As illustrated in the table, the database fully covers the period from 2005 to 2012 on all three dimensions in all receiving countries with a few exceptions. For some states, such as Austria, data on refused visas is missing for certain years. For the UK empirics on issuing practices is lacking for 2009 and 2010, but here information is available already from 2001. It is also important to note, that there is no data on consular services for the US. I expand in more detail below on the missing data where relevant.

2. Basic tables

The database stores information on regions, countries and cities in the world as well as data on travel distances and population sizes. This data is referenced in the other tables containing the information on visa requirements, issuing-practices and consular representation.

2.1 World regions

The table structure is as follows:

Column name	Data type	Comments
regionID	Int	-
regionName	nvarchar(250)	-

The list of world regions is based on the CIA World Factbook 2012. I downloaded a background data file containing an XML file listing all countries in the world grouped into regions. I copied this information into an excel file, and subsequently imported the list of regions into the database using a country import script. The database thus contains a list of 11 regions: Africa, Central Asia, East Asia, South Asia, Europe, Middle East, North America, Central America, South America, Oceania and Antarctica. This classification of regions can of course be contested. For example, is Egypt part of Africa or the Middle East? Does Mexico belong in Central or North America? For some research purposes it could also be relevant to consider for example Sub-Saharan or Northern Europe as separate regions, and indeed in the second paper of the thesis I pursue such a sub-regional approach within a European context. The database does not preclude a re-coding along these lines.

2.2 Countries

The country table is structured as follows:

Column name	Data type	Comments
countryID	Int	-
countryName	nvarchar(250)	-
countryCode	varchar(2)	-
regionID	Int	Foreign key (FK) reference to the regions table

The list of countries in the world is based on the European visa list (Council Regulation 539/2001). I copied all the entries from annex 1 (nationals requiring a visa) and from annex 2 (nationals not requiring a visa) to an excel file. For abbreviations of the country names I downloaded a table with standard country codes from the International Organisation for Standardization (ISO) and added it to the import excel file.

The visa regulation makes a distinction between "States", "Special administrative regions of the People's Republic of China" and "Entities and territorial authorities that are not recognised as states by at least one member state". I included all entries as countries.

Doing so yields a list of 167 countries. To this I then added Iceland, Liechtenstein, Norway and Switzerland as countries associated with the European Union. Finally, I added the current 27 states member states of the European Union. The final list of countries thus contains 198 data points covering almost all political entities in the world. After the excel file was put together I copied the data from excel to the database using the country import script.

In the main time period under consideration in this project the political landscape in the Balkans changed. Serbia and Montenegro split up in 2006, and in 2008 Kosovo declared itself an independent state separate from Serbia. The database contains separate entries for "Serbia", "Montenegro" and "Kosovo". There are different ways of handling such secessions. One approach is to backdate the data so that earlier information as much as possible reflects the current political landscape. For example, because Kosovo has now become independent the visa practice in Pristina should be retrospectively removed from the data on Serbia and grouped under Kosovo. This might seem a reasonable strategy when shorter time periods are investigated as is the case in the database currently. However, in a longer time-scale the approach becomes problematic as it creates pseudo political entities not reflecting the legal-political landscape in which decision-makers acted at the time. Consequently, the validity of the inferences drawn is reduced.

In order to ensure that the database can at a later stage be expanded with additional years I have thus chosen a different approach which do not involve back-dating. Earlier entries in the database for Serbia simply include the visa-issuing practices occurring at consulates in what are now the states of Montenegro and Kosovo up until 2005 and 2007 respectively. Consequently, no data is available for these two new states prior to their establishment. If, however, for particular research purposes it is relevant to backdate the data can easily be re-coded.

After importing the information into the database I made a set of changes to the country list. The country codes for Bolivia, Myanmar, Congo (Democratic Republic of), Djibouti, Iran, Laos, Micronesia, Korea (North), Russia, Sao Tome and Principe, Suriname, Syria, Tanzania, Comoros, Vietnam, Bahamas, Macedonia, Holy See, Korea (South), United States, Venezuela, Palestinian Authority, Hong Kong, Macao and Taiwan were added manually. This was necessary because the automated script failed due to differences in the naming of the countries between the EU and the ISO organization. Kosovo does not figure on the ISO list. Here I used the country code KV.

The regional coding of Burma, Congo, Côte d'Ivoire, Congo (Democratic Republic of), Djibouti, Gambia, Micronesia, Korea (North), São Tomé and Príncipe, Suriname, Comoros, Vanuatu, Palestinian Authority, Bahamas, Brunei Darussalam, Macedonia, Holy See, El Salvador, Korea (South), United States of America, Hong Kong SAR and Macao SAR were done manually as again the automated process failed due to differences in country names. The information on Vanuatu was looked up manually on the CIA World Factbook website as the country was not listed in the downloadable dataset. Note also that Palestinian Authority is listed as the West Bank in the CIA raw data. Croatia was recoded as part of Europe since the CIA data erroneously grouped it under Africa.

To correct spelling errors and for presentation purposes I adjusted the country names "Burma/Myanmar" to "Burma", "Djijbouti" to Djibouti", "North Korea" to "Korea (North)", "Surinam" to "Suriname", "The Comoros" to "Comoros", "Bahamy" to "Bahamas", "Former Yugoslav Republic of Macedonia" to "Macedonia", "Salvador to "El Salvador", "South Korea" to "Korea (South)" and "The Democratic Republic of Congo" to "Congo (Democratic Republic of)".

Finally, after the main database was constructed South Sudan - as “Sudan (South)” - was added with the country code SS. As it declared independence in July 2011 and therefore was independent for less than half of this year it has been coded as a separate country from 2012 and onwards.

2.3 Cities

The city information is contained in two tables. The first stores the basic information on each location (name, geographical location). The second links the cities with the different countries in the database, and includes a coding of whether a city is the capital of the country. The structure of the tables is as follows:

Column name	Data type	Comments
cityID	Int	-
cityName	nvarchar(250)	-
Lat	decimal(18, 7)	Location of city (latitude)
Lng	decimal(18, 7)	Location of city (longitude)

Column name	Data type	Comments
countryCityID	Int	Primary key
cityID	Int	FK reference to the cities table
countryID	Int	FK reference to countries table
capital	Tinyint	-

The list of cities in the world is based on a 2012 list of embassies and other forms of European consular representation and an overview of EU-Schengen visa-issuing practices in 2011. The former contains lists of cities outside the Schengen area; the latter also includes data on cities inside the EU, EEA or Schengen area where visas are issued. Both documents were put together by the EU Commission based on information from the member states. The raw data on representation sets out countries and cities in the world and specifies what diplomatic services if any the different Schengen member states offer. The information on issuing-practices details the number of visas issued abroad in different cities. Both documents are described in more detail in the sections on issuing-practices and consular representation. Here I only detail how I used the data to code cities.

As a first step in the import process I converted the tables in the original pdf files to excel. After having done so I inspected the converted data for apparent errors. In the process I also made a few

adjustments to the information in order to ensure the consistency in structure necessary for the automated import to succeed. These minor alterations are documented in the excel files.

In the second step I put together a conversion table to take into account that the naming of countries and cities vary. For example, in the database there is an entry for "Russia" but in the raw data on consular representation the name "Russian Federation" is used. I put the table together by stopping the import whenever I could not look up a country in the database. I then identified what name it was coded under. I followed a similar strategy for putting together a list of alternative city names. For the import I used four sheets: primary consular representation, secondary representation (e.g. honorary consuls), visa-issuing data for full Schengen countries and visa-issuing data for partial Schengen countries. After coding the first sheet I double-checked that the entries from remaining data were indeed new cities and not duplicates with a slightly varying spelling of the city name.

After this step was completed city information for 16 countries were still missing. For these states there was no representation or visa-issuing data from the EU-Schengen area. The specific countries are: Bahamas, Bhutan, Dominica, Kiribati, Liberia, Liechtenstein, Marshall Islands, Micronesia, Nauru, Palau, Saint Vincent and the Grenadines, Samoa, Solomon Islands, Somalia, Tonga and Tuvalu. Here I manually added the city listed as capital of the country in the CIA World Factbook.[2] Furthermore, I deleted the entries for Hong Kong and Macao created under the entry of China as they also figure under their separate country entries. I also renamed 'Antigua' to Saint John's to correctly register the capital of Antigua and Barbuda.

Having put together the list of cities I then turned to the coding of capitals. In the Commission representation data the capital is the first city listed under a given country. I thus made use of this ordering in the coding and flagged the first imported city as the capital. As this data does not cover all countries I manually had to code the capital in some cases, in particular for the EU states themselves. Using the Factbook I coded the following cities as capitals: Vienna, Brussels, Amsterdam, Athens, Barcelona, Bern, Berlin, Budapest, Copenhagen, Helsinki, Hong Kong, Ljubljana, Luxemburg, Macao, Madrid, Oslo, Paris, Prague, Reykjavik, Riga, Rome, Stockholm,

[2] The Factbook defines the capital as the seat of the government.

Tallinn, Vilnius, Warsaw, Lisbon and Valetta.[3] After having done so I recoded all remaining cases as not capitals.

I then inspected the dataset to check whether or not all the imported city names were unique. In doing so, I discovered an entry for Valencia in Spain and Valencia in Venezuela. I therefore renamed the former to 'Valencia (Spain)' to minimize further the risk of later coding errors.

The information on latitude and longitude of the cities was inputted using Google Maps accessed programmatically via a geo-coding web-service. I set up a computer script querying the Google servers for location data based on the country and city names stored in the database. A number of records could not be geo-coded in this way. For some of these the problem was variation in the naming of countries and cities. The code-file details the instances where I temporarily renamed a city or a country in order to make the correct data request. In three cases – Kuwait City, Funafuti and Majuro – I could not fetch the coordinates automatically. I thus inputted information on these through a manual look-up via the Google Maps webpage. Location data can be revised as errors and other issues are identified, and it should therefore be noted that future lookups could return somewhat different coordinates if Google has identified and corrected data errors. I checked the accuracy of the coding by randomly selecting 10 entries (capital cities) from the database and comparing the coding with the coordinates found via another data source, Wikipedia's mapping service.

Having set up the basic table, containing 368 cities, I then added new entries manually as need arose during the import process. All cities manually added are listed in the excel import files.

2.4 Distances

The distance table structure is as follows:

Column name	Data type	Comments
distanceID	Int	-
rcID	Int	Receiving country, reference countries table (FK)
scID	Int	Sending country, reference countries table (FK)

[3] South Sudan is coded as an independent country as of 2012 with Juba as the capital.

DistanceKm	Int	

Travel distance is a key variable when understanding global travel flows and variation in visa application numbers. I use it, specifically, in the calculation of the mobility barriers index as detailed below. I calculated the travel distance between sending and receiving countries as the distance (in kilometres) between capital cities using the Haversine formula. This equation can be used to approximate the distance between a given set of coordinates. I relied on a pre-coded implementation of it as a database script, adjusting the radius of the earth to 6.378 to fetch the results in kilometres. I thus set up an automated script parsing the coordinates of all potential sending and receiving countries for getting the distance, and inserted the result into the distance table. The result yielded a total of 39.006 rows, of which only a subset is of immediate relevance as the database does not contain information on the mobility barriers of all potential receiving countries in the world. I checked the accuracy of the coding by cross-checking the calculation in the database for ten sending countries listed for France with the distance found via the Google Earth application.

2.5 Population size

The population table structure is as follows:

Column name	Data type	Comments
populationSizeID	Int	-
countryID	Int	Reference to countries table (foreign key)
dYear	Int	
populationSize	Int	Measured in 1000s

As with travel distance, population size is a key indicator for understanding trends and variations in visa figures. The measure is used in the construction of the mobility barriers index. To code population size I draw on the United Nation's 2010 revision of world population figures. This dataset contains information on almost all countries included in the dataset. Data is only missing for Taiwan and Kosovo. For Taiwan I instead used the size and annual growth estimate of the population in the CIA Factbook for 2010, and backdated this to earlier years assuming that growth-rates were constant in the time-period. For Kosovo I made use of the alternative population dataset provided by the World Bank. In coding the figures for Serbia, Montenegro and Kosovo I took into

account that the latter two were not sovereign states in the entire time-period. The country section above explains the strategy followed. All transformations are detailed in the data files. Figures for 2011 and 2012 are the 'medium' estimates provided by the UN.

In total, I imported population data for countries for the period from 2000 to 2012 equal to 2561 observations (13 per country, in general). I checked the accuracy of the import by randomly comparing five imported values with the information in the original dataset.

3. Visa requirements

Visa requirements are stored in a table with the following structure:

Column name	Data type	Comments
visaReqID	Int	-
rcID	Int	Receiving country. FK reference to countries table
scID	Int	Sending country. FK reference to countries table
dYear	Int	
shortStayVisaRequired	Tinyint	

The table contains information on whether a receiving country in a given year required the nationals of a sending country to obtain a visa before embarking on a short trip. The database does not yet contain information on transit visa requirements but these can be added later on.

I view nationals of a sending country as ordinary citizens without a special or diplomatic passport. This is a reasonable assumption where most travellers are concerned. There are, however, exceptions. Particular visa rules for diplomats are not covered by the database. For a few countries there is visa free access for all but holders of specific types of identity documents. Here I code no visa obligation for the state as such even though some categories of travellers still require a permit to travel. In recent years, sending states have in some cases lifted the permit requirement provided that foreign nationals hold a biometric passport. In these instances, I code the visa requirement as having been lifted for the country. Furthermore, if a visa obligation was instituted or lifted during a year, the classification follows the status the country had for the main part of the year. These coding choices slightly reduce the validity of the visa requirement measure. Ideally, the different

exceptions should be explored in-depth to ascertain their precise impact and then re-coded accordingly. This, however, falls outside the scope of the research project.

The following sections describe the construction of the data for the different receiving countries.

3.1 European Union (Schengen)

The information on the permit requirements of the Schengen states is based on the 2001 common EU visa list and subsequent revisions. In 2006 Bahamas, Saint Kitts and Nevis, Antigua and Barbuda, Barbados, Mauritius and Seychelles were removed from the list pending the conclusion of a visa waiver agreement with these states. The agreements eventually reached were thus also found and consulted. All the documents were located by searches on the EUR-Lex website. I looked up the law from 2001, and used the links from there to find subsequent acts altering the visa list. The references and source documents are stored together with the database files.

To ensure consistency, I coded the visa requirement per individual Schengen receiving state for all potential sending countries – including members of the EU. As there are no travel permit obligations in force amongst EU states (including the European Economic Area), I also had to compile a list of membership status for each of the sending states. This was done using a European Commission overview of the EU and a summary of the EEA agreement by the European Free Trade Association. For example, from the perspective of Germany up until Bulgaria and Romania joined the EU their visa free status originated from the common visa list. Afterwards, it flowed from their status as EU member countries.

I coded the time-period from the establishment of the EU's common visa list in 2001 to 2012. The new member states (from the 2004 and 2007 enlargement) are coded as receiving countries from the year of their EU-membership. As above, information for Montenegro is coded from 2006 and onwards; for Kosovo from 2008 and onwards.

All in all, for this group of receiving states the database contains 49.993 entries. It is important to note, however, that the visa lists are identical for these states. Hence, the only variation is over time as the common list is expanded and contracted. Future revisions could, however, introduce variation also between the receiving states by including earlier years. Additionally, rules on short-stay visa

requirements for diplomats are not fully harmonized nor are the regulations on transit visas. Hence, introducing these would also introduce further variation in the dataset.

3.2 United Kingdom

The visa requirements of the United Kingdom are not defined in legislative acts. They are set administratively by the Home Office with a notification of Parliament. Currently, the categories of persons requiring a visa to visit the UK for a short stay (defined as a period of up to 6 months), are set out in 'Appendix 1' to the Immigration Rules. Nationalities not listed in this annex can in general visit the UK without having to obtain a visa beforehand.

It is somewhat more difficult to track changes in the UK visa list as these cannot be looked up via databases over acts of Parliament. I constructed the data entries using the following procedure. As a first step I coded the country list in the current (October 2012) appendix 1. The list was inspected for different exceptions and qualifications following the rules set out above. The entry for the "The territories formerly comprising the socialist Federal Republic of Yugoslavia" was split into Kosovo and Montenegro (Serbia was already on the list). Finally, I added Palestinian Authority to the list as it falls under the general category of "Persons who hold non-national documents". This I double-checked with the Home Office webpage.

Having put together this basic list I then tracked the changes to the appendix using the 'Statements of changes in Immigration rules' published from 2003 and onwards on the UKBA website. I went through each of the documents from 2003 to 2011 searching for visa, and then downloaded and inspected those that made a change to the list of visa nationals. Using this procedure I identified changes with regards to Lesotho, Swaziland, Bolivia, Taiwan, South Africa, Bulgaria, Romania, Croatia and Malawi in the time period. These were then entered into the excel sheet. In the final step I imported the information to the database, assuming that for all other countries than those identified on the list no visa requirement was in force.

3.3 United States

As a main rule, the United States requires that all persons who seek to visit obtain a 'non-immigrant' visa before embarking on their trip. This requirement, however, has since the late 1980s been lifted for a limited group of countries through the 'Visa Waiver Program'. Whether or not a

country is included in the program is an administrative decision based on guidelines set by Congress. After the 9/11 terrorist attacks the visa exemptions was criticised for endangering the security of the United States. As a response, the rules were tightened and visa-waiver nationals now need to be authorised prior to the travel through the so-called ESTA system. To obtain the permit it is necessary to pay a fee, fill out a form, submit passport information and consent to the US authorities using the information. ESTA raises the question of whether the US has de facto re-introduced a visa obligation for all. I have not coded this to be the case because of the short time and ease with which it is possible to go through the ESTA screening. However, it is debatable to what extent the ESTA barrier is much different from – for example – the relatively lenient visa procedures encountered by many in for example Taiwan.

I constructed the visa list for the United States by starting out with the current list of countries participating in the waiver program. After having done so, I searched through the US Federal Register for departmental notifications on changes in the program. The precise documents found and the search criteria used are listed in the source files. I identified the following changes to the program in the time-period investigated: Argentina was removed from the list in 2002, Uruguay removed 2003, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Republic of Korea, Slovak Republic and Malta added in 2008, and finally, Greece added in 2010. Taiwan was added in 2012.

Additionally, the visa requirement was also lifted for a limited set of countries through other legislative means in the time-period. Canadian citizens are generally able to travel freely to the US. A similar option is open for nationals of Micronesia and the Marshall Islands. These countries are hence also coded as having visa-free access to the United States.

4. Visa-issuing practices

The database contains three tables on short-stay visa-issuing practices for respectively the Schengen group, the UK and the US. A short-stay is here usually defined as a trip for no more than three (Schengen, US) or six months (UK). A permit may be issued allowing only for a shorter stay, for example a week's conference attendance. It can also be valid for multiple entries and thereby enable the holder to conduct several small visits over a longer time period.

In the following sections I describe how I coded the data for the individual receiving countries.

4.1 European Union (Schengen)

The table structure is as follows:

Column name	Data type	Comments
visaPracticeEuID	Int	-
rcID	Int	Receiving country. FK reference to countries table
scCityID	Int	Location of consulate. FK reference to cities table
dYear	Int	
shortStayAppliedFor	Int	Calculated
shortStayIssued	Int	Calculated
shortStayRefused	Int	Calculated
shortStayRefusalRate	Decimal(5,2)	Calculated
issuedA_All	Int	A = Airport transit visa
issuedA_Mev	Int	Mev = valid for multiple entries. 2010 and onwards
issuedB	Int	B = Transit visa
issuedC_All	Int	C = Short-stay visa
issuedC_Mev	Int	Data only available for 2010 and onwards
issuedD	Int	D = National long-stay visa
issuedDC	Int	D + C = National long-stay also valid as C visa
issuedVTL	Int	VTL = National short-stay visa
issuedADS	Int	ADS = Chinese tourist group visa
issuedABC	Int	
issuedABCDDCVTL	Int	
appliedC	Int	
appliedABC	Int	
notIssuedA	Int	
notIssuedB	Int	
notIssuedC	Int	
notIssuedABC	Int	

The information on EU visa issuing practices is based on detailed tables setting out the number of visas applied for, issued, and not issued at the member states' consulates. These overviews were put together by the General Secretariat of the Council of the European Union up to and including 2009.

In subsequent years, due to the entry into force of new visa legislation, the collection and publication of the data has been taken over by the European Commission. The visa-issuing statistics are supplied per embassy.

The information was compiled in four steps. In step *one* I converted the raw pdf files to excel. This was only necessary for the Council data. The Commission made the figures available from the outset in excel. I then went through the processed data fixing errors in the conversion and standardising the layout. The precise changes I made to the files are set out in the source files. Step *two* was the actual import of the information in the excel sheets to the database. This was done using a computer script. During this process I added new cities to the database as necessary. In step *three* I compared the sum-totals for each year and visa column with the similar totals in the raw data, and tracked down and corrected any errors there might have been in the import. This check was primarily done per receiving country and for the Schengen area as such. I checked whether the data had been coded correctly under the different sending countries and cities by randomly looking up a limited set of data entries. As part of this step I also inspected the data from the member states for apparent major errors and problems. This revealed that 2009 figures for Norway deviated greatly with earlier and later years. I hence excluded these.

In step *four* I estimated the short-stay visas applied for, issued and refused by the EU states. Here I made use of the columns detailing the total number of 'ABC' visas applied for, issued and not issued. 'A' and 'B' are visas for transit and 'C' for short stays. In general, the member states have supplied data on these variables consistently across the years. Austria, however, did not report information in 2007 and 2008. In a set of cases data on the visas not issued were left blank. In general, I interpreted this as missing data. However, when the number of visas applied for and issued was the same I assumed that the column was left blank because no visas were refused that year. In most cases a zero would have been entered in the field but in some cases this was apparently not done.

The precise equations used to calculate the visas applied for, issued, refused and the refusal rate are listed in section 7. The main ideas behind them are as follows:

Firstly, by refused I understand both formal rejections and informal advice or delays resulting in the withdrawal of an application. Often a visa application results in a formal refusal. In other instances an applicant is encouraged to withdraw his or her request before this stage is reached. For example, some consular officials might encourage persons to withdraw 'for their own sake' so that a formal refusal is not entered into government databases. Delaying tactics might also mean that the process is prolonged and that the applicant therefore gives up. In some cases, of course, a withdrawal might be entirely voluntary and not related to state practice. For example, a conference is cancelled well in advance and a visa is therefore no longer necessary. Yet it is fair to assume that the latter form of withdrawal is comparatively rare especially as applicants have already paid a not insubstantial visa handling fee.

To what extent does it matter whether or not only formal rejections are included in the measure? Limited data is available on this. German statistics for 2003 suggests that formal rejections frequently are not made, whereas UK figures for 2006 to 2008 points in a different direction. It is likely that there is considerable variation in practices. Including informal withdrawals in the refusal rate is important to avoid significantly underestimating restrictiveness in certain cases. Reversely, with this strategy there is some risk of overestimating the refusal rate if it should be common that applicants withdraw freely. Yet of the two scenarios the latter is the least probable.

Secondly, in the calculations I include data on national 'VTL' visas. A VTL visa is a permit valid for transit or a short stay to one or more of the member states but not the entire Schengen area. It can be issued, for example, when a receiving state deems that an applicant does not meet all entry criteria but still wishes to issue a visa for humanitarian or political reasons. It is also issued when there is disagreement or insufficient clarity between the member states on what travel documents they recognize as valid.

VTL visas are not issued very often. In most cases, whether or not they are taken into account in the calculation of the refusal rate makes little substantial difference. Since the enlargement of the Schengen area in December 2007 the use of VTL visas has, however, increased. This is mainly due to a change in practice by France. In a French government report the new approach is justified with reference to the lack of information provided by Central and Eastern European member states on the travel documents they recognize. Apart from this, VTLs visas primarily matter in Macedonia

and Iraq where several EU countries issue them frequently. It is to avoid overestimating the refusal rate in these cases that I include VTL visas in the calculations.

Thirdly, transit visas (A, B) figure in the estimate. In principle they should be excluded since they do not allow for a short stay. This is not possible however. The statistical material for 2005 to 2010 contains information about the number of transit visas issued but does not state how many were applied for and refused. Recent 2011 and 2012 statistics include additional information making it possible to calculate transit and short stay refusal rates separately for this year. Doing so reveals that only in a very limited set of cases is the refusal rate substantially influenced by the transit visas. When these permits are excluded the refusal rate tends to increase. Hence having them in the calculation is thus not generally a problem but it does mean that in a few a cases the refusal rate is underestimated.

4.2 United Kingdom

The data from the United Kingdom is stored in a table structured as follows:

Column name	Data type	Comments
visaPracticeID	Int	-
rcID	Int	Receiving country. FK reference to countries table
scCityID	Int	Location of consulate. FK reference to cities table
dYear	Int	
shortStayAppliedFor	Int	Calculated
shortStayIssued	Int	Calculated
shortStayRefused	Int	Calculated
shortStayRefusalRate	Decimal(5,2)	Calculated
visitReceived	Int	Only for 2001 to 2004
visitIssued	Int	Only for 2001 to 2004
visitRefused	Int	Only for 2001 to 2004
visitFamilyReceived	Int	Only for 2004 to 2008
visitFamilyIssued	Int	Only for 2004 to 2008
visitFamilyRefused	Int	Only for 2004 to 2008
visitFamilyWithdrawn	Int	Only for 2006 to 2008
visitFamilyLapsed	Int	Only for 2006 to 2008
visitFamilyDecided	Int	Only for 2006 to 2008
visitOtherReceived	Int	Only for 2005 to 2008

visitOtherIssued	Int	Only for 2005 to 2008
visitOtherRefused	Int	Only for 2005 to 2008
visitOtherWithdrawn	Int	Only for 2006 to 2008
visitOtherLapsed	Int	Only for 2006 to 2008
visitOtherDecided	Int	Only for 2006 to 2008
transitReceived	Int	Only for 2005 to 2008
transitIssued	Int	Only for 2005 to 2008
transitRefused	Int	Only for 2005 to 2008
transitWithdrawn	Int	Only for 2006 to 2008
transitLapsed	Int	Only for 2006 to 2008
transitDecided	Int	Only for 2006 to 2008

The data for the United Kingdom is based on the 'entry clearance statistics' which were published by the Home Office and is now accessible via the national archives. There is data for the period 2001 to 2008. Unfortunately the data does not follow the calendar year, as is the case for the EU statistics, but the British government's financial year (1 April - 31 March). The data from 2001 to 2005 was published by the UK Visas agency. The 2006, 2007 and 2008 information was made available by the UK Border Agency. The raw data is grouped per consulate ('diplomatic post').

I imported the data in a series of steps. First, I isolated the information on visas related to short-stays and converted these tables to excel format. I then inspected the files and cleaned up the data, for example removing superfluous header lines. The precise changes are detailed in the source files. For 2001, 2002 and 2003 the data contained information on 'visit' visa applications received, issued and refused. From 2004 and onwards the files also list how many applications concerned family visits; from 2005 there is data on transit visas. The data for 2006, 2007 and 2008 maintained the basic structure from 2005 but now also added data on how many applications were withdrawn, lapsed and decided upon by the different diplomatic posts.

Second, I imported the visa practice data using a computer script. I did this for each year in turn and compared the sum totals in the database with the raw data. For 2006, 2007 and 2008 minor deviations (usually 5 or 10) started to appear. These were seemingly due to the fact that the original data was now rounded in 5s. I also inspected the data for any apparent major deviating trends that might reflect coding errors. This did not reveal any apparent problems.

After having imported the data I calculated the totals for the short stay visas applied for, issued, refused and the refusal rate. For the period 2001 to 2004 this was straightforward. For 2005 and onwards, when applications became divided into family and other visits, I used a sum of the two and calculated the refusal rate accordingly. For 2006 and onwards I also included withdrawn or lapsed in the estimation. This was done to improve comparability with the EU statistics.

In total, the database contains 1.210 observations on the UK case. The number of visas applied for is available from all observations. Refusal rates are missing from 14 cases because the number of visas issued and refused at the diplomatic posts was zero.

4.3 United States

The table on US visa-issuing practices is structured as follows:

Column name	Data type	Comments
visaPracticeUsID	Int	Primary key (unique identifier)
rcID	Int	Receiving country. FK reference to countries table
scID	Int	Sending country. FK reference to countries table
dYear	Int	
shortStayAppliedFor	Int	Calculated based on issued and refusal rate
shortStayIssued	Int	
shortStayRefused	Int	Calculated based on issued and refusal rate
shortStayRefusalRate	Decimal	In percentage (%)
typeB1issued	Int	B1
typeB2issued	Int	B2
typeB1comb2issued	Int	B1,2
typeB1comb2BCCissued	Int	B1,2/BCC
typeB1comb2BCVissued	Int	B1,2/BCV

I constructed the US data entries using an overview over visa refusal rates and a detailed background table setting out the number of visas issued per nationality. The time period covered is 2006 to 2012. The data relates to ‘B’ visas issued for visits for business or pleasure. This is by far the most widely used visa for temporary entry to the US. It can be issued either solely for business (B-1) or pleasure (B-2) or as valid for both purposes (B-1,2).

There is raw data on visas issued for earlier years but here data on refusal rates are missing. The US data differ from the UK and EU figures in important ways. It is, firstly, grouped per nationality. Hence, the data relates to all citizens of a given country regardless of where they submitted their application. The data, secondly, follows the US fiscal year which runs from 1 October to 30 September. For example, fiscal year 2006 runs from 1 October 2005 to 30 September 2006. Thus the data for a given year also includes information on some of the previous months.

As the US figures only include information on the refusal rate and the visas issued it is necessary to calculate the applications received and refused manually. In the Department of State's explanatory note to the overview of refusal rates it is clarified that the calculation is based on the number of decisions made and only include final refusals. I reversed this equation to identify the number of visas applied for and refused. For mathematical reasons, this approach cannot be used in cases where the refusal rate is 100. This is a very limited problem as such a high refusal rate is only reported in three cases: Micronesia 2006 and 2009 as well as Andorra 2010. Here the applications received and refused are unknown. Please note that since the US statistics apparently do not include withdrawn applications the number of visas applied for might be higher than calculated. Likewise, the US refusal rate could be somewhat underestimated in comparison with the UK and Schengen.

I coded the data in two main steps. I started out by converting the pdf files with refusal rates to Excel format to be able to import the data. The information on visas issued was already in Excel format.

In coding the data I made the following choices. For Hong Kong I used the refusal rate information for 'Hong Kong SAR' and not the separate figure for 'Hong Kong BNO HK passport'. Data on visas issued and refusal rates for unknown, no nationality or laissez-passer was ignored. Throughout the period I coded 'Serbia and Montenegro' under Serbia. For 2008 to 2010 this meant that the otherwise separate entries for 'Serbia and Montenegro' and 'Serbia' were merged into one. The data for 2006 included statistics on visas for Serbia without a corresponding record on refusal rates. I hence ignored this and only used the statistics on 'Serbia and Montenegro' in this year. The only missing cases are Montenegro for 2006 and 2007.

In relation to Mexico, importantly, the data is not valid. This is because a separate type of permit is widely used, the Border Crossing Card (BCC or BCV). The visa refusal rate published by the US does not include these, at least not for 2006-2009. This means that the estimate in the database does not reflect the actual number of permit applications and decisions for this country. The refusal rate is of course also of limited value.

After having thus inspected the data and clarified these issues I imported the data using a computer script designed for the purpose. I fetched and added the refusal rate and issued visas for each year and sending country. Then I calculated the number of visas applied for and refused based on the imported statistics. Afterwards, I randomly checked a set of the refusal rate figures. I also controlled that the sum total of visas issued in the database equals the sum in the original data. This control only revealed a deviation of 2 visas issued for 2006 relating to the excluded case of 'Serbia'.

Please note that in 2012 the US authorities change the way they calculate the refusal rate. Before, it was calculated using figures on all applications. In 2012 this was changed so that the estimate is only based on the outcome of the last visa request submitted by an application. Differences between 2011 and 2012 should thus be interpreted with caution as these might be a result of this change in administrative practice.

5. Consular representation

The information on consular services for visa-issuing purposes are stored in a single table structured as follows:

Column name	Data type	Comments
visaReprID	Int	-
rcID	Int	Receiving country. FK reference to countries table
scCityID	Int	Sending city. FK reference to countries_cities table
dYear	Int	
reprByRcID	Int	State representing the receiving country. FK reference to countries table
ExtSerPro	Tinyint	External Service Provider involved (outsourcing)

The database contains data on the diplomatic representation of the Schengen states. It includes information on their use of cooperative agreements between them. For example, Denmark might process and issue visa applications on behalf of Norway in several sending countries.

5.1 European Union (Schengen)

I coded the diplomatic representation of the Schengen states drawing on overviews put together by the Council General Secretariat in different years (annex 18 tables to the Common Consular Instructions). With the recent entry into force of the European visa code the task of compiling the annex was transferred to the Commission, and hence data for 2011 and onwards are based on Commission overviews (annex 28 to the EC visa code Handbook). These files detail the cities where the member states have independent consular representation for visa-issuing purposes, and where they rely on cooperative agreements with Schengen partner states. The pattern of diplomatic representation alters during years. I coded a consolidated version of the annex published in a given year as indicative of the representation pattern, but also made note of the precise date of the data so that it is possible to re-group otherwise where relevant. For example, in paper two I make use of a slightly different year coding.

The coding was done in two steps. First, I converted the raw pdf data to excel and inspected the data for immediate conversion errors. Having done so, I went through detailed footnotes in the originals providing additional comments on individual cases. For example, a note might indicate that a representation agreement only last for a specific time period or that a consulate is currently not accepting visa applications. When an embassy was noted to be in practice closed or not accepting applications I registered this and did not import it as a case of visa representation. I removed notices on representation agreements if they lasted less than half of the year in question. When the Commission took over responsibility of the annex they also started to collect information on the cities where applications are processed in cooperation with a private company (‘external service provider’). I decided to also code this, as the outsourcing of the visa process might be interesting for other researchers to explore. The list now also indicates cities where a member state does not have consular representation as such but the private firm has an office. I did not code these cases as instances of representation to ensure consistency with earlier years.

In the second step I imported the information to the database making use of the automated computer script. Of these, data for 2011 and onwards also contain a coding of the use of private firms.[4]

5.2 United Kingdom

Information on the UK case is not contained in the overviews produced by the Commission and the Council. Data on the consular services of the United Kingdom are thus indirectly coded via the information on the visas issued. As this data is supplied by consulate it provides information on the diplomatic posts where the UK handles visa applications. And, reversely, it indicates where the UK is not represented for visa-issuing purposes.

5.3 United States

Information on US consular services is missing from the database. These cannot be inferred from the visa practice data as this is grouped per nationality and do not state in what country or city the applications were lodged.

6. Visa policy and practice: Mobility Barriers Index

The database contains a main mobility barriers table drawing together the information from the sub-tables on visa requirements, issuing practices and consular representation data for the different receiving countries. The content of this table is shown as default on the database website. The structure of this table is as follows:

Column name	Data type	Comments
evdID	Int	Primary key (unique identifier)
rcID	Int	Receiving country. FK reference to countries table
scID	Int	Sending country. FK reference to countries table
dYear	Int	
rcSchMember	Tinyint	
visaRequirement	Tinyint	
rcReprType	Int	
rcReprOthers	Int	
rcReprByRcID	Int	Representing state. FK reference to countries table
visaAppliedFor	Int	

[4] External service providers, the data indicates, is only used in a relative small amount of cases. They are, however, often involved in major sending countries such as Russia, Turkey, India and China.

visaIssued	Int	
visaRefused	Int	
visaRefusalRate	Decimal	In percentage (%)
mobBarIndex	Int	Mobility Barriers Index

The analytical unit in the table is country-pairs in different years. It measures the mobility barrier of a receiving state towards a sending country in a given year. The table was put together in a series of step. I started out by adding the basic information countries and years. I then added on the information on visa requirements, issuing practice, consular representation and the mobility barrier index. After having done so, I corrected the table for missing data entries. For example, I made sure that the records accurately reflected that there is no data for the UK for 2009 and 2010. As a final step I coded cooperating countries as having the same refusal rate and mobility barrier. The next sections describe in further detail the different variables contained in the table and how they were computed.

Starting with the diplomatic representation, the “rcReprType” variable measures whether a receiving country has an embassy in a sending state, relies on a cooperative agreement or is not present at all in the location. If a state relied on both (for example had an own embassy in one city and cooperated in another) I coded it as being independently represented. The “rcReprOthers” variables measure whether or not the receiving country represents partners in the sending state. Finally, the “rcReprByRcID” variable is used to identify what member state a receiving country is represented by. For the US I simply coded information on these dimensions as missing. I coded the UK as represented in states where it had processed visas. Reversely, I assumed that it had no representation in the countries were it did not process visas. I followed a similar logic for the countries only partially participating in Schengen.

On the visa statistics, I aggregated the consular data to the country level where necessary. For example, I calculated the total number of visas applied for, issued and refused at all French consulates in China. These sums were then inserted in the table. Note that I also use these sums to calculate the refusal rate. That is, the refusal rate is not an average of the practice at the individual consulates in a state but for the sending country measured as a single unit.

If a member state was represented by another, I coded the number of visas applied for, issued and refused as missing but copied over the refusal rate from the representing state. I thus assume that when countries share embassies their mobility barrier is the same. In the (few) situations where a receiving country was represented by more than one partner state in a sending country I selected one of them randomly.

Moving on to the Mobility Barriers Index, the overall idea behind this indicator is to provide a single restrictiveness score for a country-pair in a year taking into account both visa requirements, issuing practices and consular representation. I set up the indicator as a four point scale ranging from no mobility barriers to low, medium and high.

The index was constructed using the following rules. Firstly, if no visa requirement is in force I code the mobility barrier as none. Although the control at the territorial border is a hurdle to movement, it is here assumed to have a very limited impact compared with the obligation to obtain a visa before even embarking on the trip.

Secondly, if a visa requirement is in force I use the refusal rate to group a case as either one of a low, medium or high barrier. To do so, I first inspected the interquartile ranges in the dataset and used these as a starting-point for the classification. Here I only considered applications lodged in a visa-list country (EU, UK) or by nationals on a visa-list (US). The first range of observations in the dataset (0-25%) covers refusal rates from 0 to 4, the second and third (25-75%) 5 to 21 and the fourth (75-100%) captures rates from 22 and above. I then decided to deviate a little from the interquartile ranges and use a more easily communicable range. I thus coded a refusal rate of below 5 as low, between 5 and 20 as medium and above 20 as high. These values thus approximate but do not strictly follow the interquartile ranges.

Thirdly, when receiving countries cooperated in a sending state by sharing consulates I coded them as having the same mobility barrier.

Fourthly, if a receiving state was not represented at all in a sending country I coded the barrier as medium. It might be argued that this score should be higher. Why not code the impossibility of lodging a visa application as a high mobility barrier? I decided not to go this route as in several

cases the absence of an embassy or consulate need not be a major obstacle to travel as applications can be forwarded and processed in a nearby state, and here issued leniently. Hence, the medium score is a compromise between this consideration and rival cases (such as Somalia, Sierra Leone) where the absence of a consulate could well be interpreted as a high barrier.

A key feature, finally, of the index is that it tries to take into account the many options receiving states have for preventing applications from being lodged in the first place. It does so by comparing the actual number of visas applied for with a model estimate. If the application figures are very low, below 20% of estimated, it adds a penalty score to the index. For example, a score of "1" (low barrier) is lifted to "2" (medium). The model is simple so as to not bias later analytical result and ensure the transparency of the indicator. It only uses the population sizes of the receiving and sending countries, as well as the travel distance, to estimate application numbers. Adding on for example income (GDP per capita) to the model would undoubtedly increase explanatory purchase. But it would at the same time risk biasing later analytical results investigating how wealth influences mobility barriers. I ran the model as an ordinary linear regression analysis predicting the number of visas applied for based on the predictors. All the variables were transformed using the natural log to better approximate a normal distribution. This worked well for all variables but the size of the receiving countries. The distribution of this variable was not optimal. The regression only seeks to predict the amount of applications received for countries facing a visa requirement. The key results of the model are as follows:

Table 1

Estimating the number of visas applied for

Main model	Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
Constant	5,60	0,23		23,89	0,00	5,14	6,06
Population size (sending country) (ln)	0,47	0,01	0,44	41,58	0,00	0,45	0,49
Population size (receiving country) (ln)	0,78	0,01	0,59	52,87	0,00	0,75	0,81
Travel distance (ln)	-1,23	0,03	-0,51	-47,55	0,00	-1,28	-1,18

Notes: r square = 37%, n = 6.806

As shown in the table, the model has a good overall explanatory purchase (37%). All the predictors are significant at the 0,01 level. Let us look at a couple of examples. The regression predicts that Austria (population size 8,2 million) should receive about 3.600 applications from Albania (3,1 million) with a travel distance of approximately 800 kilometres. In practice, Austria received on average 3.800 applications annually. A contrasting case is Germany in Iraq in 2005. Here the model predicts about 11.000 visa requests annually yet only 1.100 was received this year. In general, the regression identifies cases of few applications across most sending countries. The instances with very low figures concentrate, however, in a set of countries such as Algeria, Zimbabwe, Pakistan, Nigeria, Iraq and Afghanistan.

I checked the stability of the coefficients, significance levels and model explanatory purchase (r^2) by running the same model separately for all available combinations of receiving countries and years. For some states (Iceland, Estonia, Latvia, Lithuania, Luxembourg, Malta, Slovenia) the number of observations per year was too low to carry out such an analysis. In total, I ran 123 separate regressions checking for changes in significance levels and shifts in the direction of the coefficients. Doing so identified, first and foremost, that the Portuguese application figures are not captured by the indicators. They are not significant and the effect of sending country population size even drops slightly below zero. In the case of Slovakia size was not significant in any years. For the remainder, there were four years where population size dropped out as insignificant and seven where distance did the same. Thus, in terms of significance levels and the direction of effects the predictors are very stable. Turning to the explanatory purchase of the model we do see some variation. In most cases the r^2 is between 30 and 50%. The model is particularly strong in the case of the US (r^2 = 66%). At the bottom end we find Spain, the UK and the Netherlands (r^2 = 18-23%).

The overall size of the coefficients did not, at least apparently, vary. However, as the variables are measured on a logarithmic scale small changes can have a major impact. Minor alterations matter a lot for small countries but less so for larger ones.

Still, a relatively strong and transparent model I used it to identify the cases where the model estimate is much higher than the applications actually received. I assigned a penalty score when the number of visa application received was below 20% of the estimated. This should ensure that the

ability of destination states to prevent applications from being lodged is adequately captured. Where to draw the line is of course debatable.

A closer look at the cases where the index score was adjusted reveals a set of interesting patterns. Firstly, two main sending countries where the barrier index is moved upward are Burma and North Korea. Other key cases are Zimbabwe, Afghanistan and Tunisia. The first two suggests that the model also captures cases where it might be debatable to what extent the barrier is due to policies and practices of the receiving or sending countries. In these two states exit is tightly controlled by the sending state governments. In terms of receiving countries, the adjustment affects the observations from Romania and Poland far more than for other states. This implies that these states make more use of the options for preventing applications being lodged in the first place than other receiving states. It could also reflect that the shared model is not well adapted to these cases. That is, the coefficients are not so well-suited to capture the dynamics of these cases.

7. Equations

7.1 European Union (Schengen)

(1) $applied\ for = \ ABC\ visas\ applied\ for + \ VTL\ visas\ issued$

(2) $issued = \ ABC\ visas\ issued + \ VTL\ visas\ issued$

(3) $refused = ABC\ visas\ not\ issued$

(4) $$refusal\ rate = \frac{ABC\ visas\ not\ issued}{ABC\ visas\ not\ issued + ABC\ visas\ issued + VTL\ visas\ issued}$$

7.2 United Kingdom

(5) $$refusal\ rate = \frac{refused + withdrawn + lapsed}{issued + \ refused + withdrawn + lapsed}$$

Note 1: Calculation is based on ‘visit’ visas (2001-2004) and family plus other (2005-2008)

Note 2: Data on withdrawn and lapsed only available from 2006 and onwards.

7.3 United States

$$(6) refusal\ rate = \frac{refused - overcome}{issued + refused - overcome}$$

$$(7) Applications\ refused = \frac{issued}{1 - \frac{refusal\ rate}{100}}$$

$$(8) Applications\ received = issued + refused$$

8. Database diagram

The diagram below is a technical overview of the structure of the database (referred to as an 'E-R' diagram). This version mainly shows the data-columns that are used to cross-reference the content in one table with another. For example, the figure highlights that the table 'visaPractice_EU' is linked with the countries table (for identifying the receiving country) and the table over cities in different countries (for identifying the location of the diplomatic post).