FVAP RESEARCH NOTE

Comparing Models for Estimating the Total Overseas Citizen Population, the Overseas Citizen Voting-Age Population, and Voting Rates

Executive Summary

The Federal Voting Assistance Program (FVAP) continues its exploration of better methods to estimate the population of U.S. citizens living overseas. FVAP has calculated these estimates since 2014 as part of its biennial Overseas Citizen Population Analysis (OCPA). Historically, FVAP relies largely on statistical modeling using country-level estimates produced by foreign governments in part, but recently explored alternative methods that rely more on foreign government estimates (FGEs) and less on modeling.

FVAP's estimates are the most widely referenced measures of the size of the overseas U.S. citizen community. They enable FVAP to provide estimated overseas voter registration and voting participation rates.

For the 2020 OCPA, FVAP did not update its population estimates but instead used the 2018 estimates to calculate the registration and participation rates of overseas citizens for the 2020 General Election. This research note provides updated 2020 estimates of the number of U.S. citizens living overseas, and rigorously assesses FVAP's statistical estimation process and the data sources used. This assessment is based on lessons learned since 2014, new 2020 FGEs, and other data that have become available

In recognition that at times FVAP's population estimates for certain countries differ substantially from FGEs of the number of U.S. citizens residing within their borders, FVAP's methodological assessment examines whether these estimates might be improved by giving greater weight to FGEs within statistical models.

Within this research note, the estimation approach giving greater weight to FGEs is referred to as the Calibrated method, while FVAP's traditional approach is referred to as the Model-Based method. FVAP compared the results of these methods and found that when greater weight is given to FGEs, validity does not improve and the margin of error increases. Even if the estimations produced seem to be more in line with external sources for certain countries (e.g., Mexico and Canada), FVAP's traditional Model-Based method is considered more reliable than the Calibrated method.

Estimation Methodology The U.S. Government does not track the locations of U.S. citizens when they travel, live, work, or study overseas. Since definite overseas citizen population data does not exist, FVAP uses a model of U.S. and foreign government data to estimate the number of U.S. citizens living in the different countries and regions of the world as well as the number of those citizens who were of voting age (over 18 years old) at the time of the General Election (18 years or older). The estimates for U.S. citizens who were at least 18 years old at the time of the General Election are referred to as the Overseas Citizen Voting-Age Population (OCVAP).

This Model-Based methodology uses a series of models that draw some data, but not all, from the often-sparse set of country-level estimates produced by foreign governments. This approach relies largely on statistical models and discards much of the information contained in FGEs of overseas U.S. citizen populations. As a potential update to the methodology for generating estimates of the overseas population, FVAP examined Calibrated estimates, which incorporate information directly from the FGEs for those countries in which they are available.

Testing Each Method To examine the relative performance of the two methods, FVAP ran a series of validation tests to see if the differences between them were the result of measurement error or a meaningful improvement using the Calibrated method. To do this, FVAP looked at two correlations:

1) The correlation between estimates and the number of ballot requesters within each election cycle;

2) The correlation between the estimated population growth rate within a given country and the growth rate in ballot requests within that country.

Results and Conclusions Due to the limitation of not being able to capture every potentially relevant feature of a country, we cannot confirm that the estimates produced for every country using the Model-Based estimates are more accurate than those of the Calibrated method. However, the estimates using the Model-Based method provide more reliable inferences about differences in the size of the U.S. citizen population and OCVAP voting rates across countries. Therefore, the analysis concludes that the estimates produced using the Calibrated method are not as reliable as the estimates produced by the Model-Based method.

In part, calibrating estimates to FGEs may not improve models because of measurement error in the FGEs themselves, especially those from developing countries. FGEs are not available for each country, and many release estimates in a cycle of only every 5 or 10 years. Furthermore, there is a great deal of variation in how foreign governments define their resident U.S. citizen populations. These FGEs are calculated using a variety of different methods (e.g., censuses and registries) and are reported through different sources (e.g., United Nations Statistics Division, foreign countries' national statistical agencies, U.S. Census Bureau, etc.). Countries may report their overseas population by nationality but not define how this is determined, for example, whether U.S. citizens, U.S.-born individuals, or dual citizens are included in these counts. These issues significantly limit the utility of FGEs in improving overall population estimates.

While the analysis conducted as part of this research note does not support changing FVAP's strategy for estimating the overseas citizen population, FVAP continuously seeks to improve the models and methodology used in its biennial OCPA reporting. As FVAP learns more about data on the relative sizes of different countries' U.S. citizen populations, this information may be useful in improving Model-Based estimates.

Comparing Models for Estimating the Total Overseas Citizen Population, the Overseas Citizen Voting-Age Population, and Voting Rates

Introduction

As part of the Overseas Citizen Population Analysis (OCPA), FVAP has developed estimates of the number of U.S. citizens living overseas and the size and voting rate of the Overseas Citizen Voting-Age Population (OCVAP) for the 2014, 2016, 2018, and 2020 elections. For every election, this has involved collecting data on absentee ballots requested by and the vote history of U.S. citizens living outside the United States from state and local election officials, using this data to count the number of votes recorded that can be attributed to the OCVAP both by country and in total, and dividing these counts by estimates for the size of the OCVAP.

In anticipation of the 2022 OCPA, FVAP is investigating updates to the methodology used to generate these population estimates. Specifically, the current methodology involves estimating a series of econometric models of an often-sparse set of country-level estimates of the overseas citizen population produced by foreign government statistical agencies or Foreign Government Estimates (FGE) and taking a weighted average of the predictions of these models for every country-year between 2000–2020. The resulting estimates are entirely based on models and thus inevitably discard information contained in the FGEs. As a potential update to the methodology for generating overseas citizen estimates, FVAP examined leveraging information from the Model-Based estimates while also incorporating information directly from the FGEs for those countries for which they are available.

This research note compares results using the two methodologies. It is organized as follows:

- Description of the current methodology used to generate estimates (the Model-Based estimation methodology);
- Description of the proposed updated methodology (the Calibrated estimation methodology);
- Two sets of OCVAP estimates, updated through 2020, with comparisons of estimates and the resulting OCVAP voting rates generated using each methodology;
- Comparisons of the validity of the two estimating methodologies; and
- Summary of the findings and potential avenues for future research.

Current Overseas Citizen Population Estimation Methodology

In general, the U.S. Government does not keep track of where U.S. citizens are located when they travel, live, work, or study overseas. For some nations, it is likely that data on the number of U.S. citizens currently in their country do exist; countries with visa requirements for entry and exit, such as China, should be able to provide information on the number of U.S. citizens in their country at any given time. However, it is not always possible to access such data. Thus, there is no exact count of the total number of citizens overseas; nor do many nations produce a consistent enumeration of the number of U.S. citizens who live within their borders.

Because of these issues and others discussed below, we must estimate the number of overseas citizens in each country to accurately measure voter participation among overseas U.S. citizens. These estimates have been generated using three primary data sources: foreign country data on the number of U.S. citizens living within foreign countries' borders, U.S. Government administrative data on overseas citizens, and data from academic studies that have examined factors that affect the number of U.S. citizens living in any given country around the world.

The groundwork for this analysis was laid in 2015 when FVAP conducted this analysis for the 2014 election; it was refreshed to produce the updated estimates for the 2016 election. This section discusses the data collection, imputation, and estimation methodology from 2017, as well as how it was updated to produce new estimates for the 2018 election.

Foreign Government Estimates (FGE)

The term "foreign government estimate" (FGE) is used throughout this report. This term refers to the data that foreign governments have, through registries and/or censuses, on the number of U.S. citizens living in their country.

Census versus Registry

This report also uses the terms "census" and "registry." And it is important to understand the distinction between the two.

- A **census** is a country-wide, periodic data collection that tallies all residents.
- A **registry** is a compilation of administrative records from numerous sources.

Registries may provide more complete counts than a census if they are updated often and are drawn from several different sources (e.g., tax records, visas, school forms). One major disadvantage of registries is that U.S. citizens may continue to appear on a foreign registry for several years after they no longer reside in that country.

Current Overseas Citizen Population Estimation Methodology

There are several sources for Foreign Government Estimates (FGEs) of the U.S. citizens living in each country. The FGEs used in the analyses come from: (1) the United Nations Statistics Division, which collects data on migrant stocks from the statistical agencies of many countries; (2) census microdata collected and standardized by the Minnesota Population Center's Integrated Public Use Microdata Series (IPUMS) International; (3) documents released by countries' national statistical agencies; (4) the Organisation for Economic Co-operation and Development (OECD) International Migration Database, which provides data on the number of U.S. citizens during the years 2000 to 2020 for most OECD countries; and (5) a U.S. Census Bureau internal document titled, "Estimating Native Emigration from the United States," which was compiled as part of a project to estimate U.S. net emigration.

The primary methods that foreign governments use to track the population of U.S. citizens in their country are censuses and registries. Both census and registry data, in addition to an indicator variable, were used to account for the difference in collection method. Countries vary in who they consider a U.S. citizen for purposes of a census or registry. Some countries count only U.S. citizens and others count only individuals born in the United States. The groups defined by these two criteria have significant overlap, but a small proportion of individuals belong to only one of those groups. This discrepancy was accounted for by having an indicator variable for whether the country uses U.S. citizens or U.S.-born individuals, ultimately allowing for the estimation of the number of U.S. citizens, despite this variation by country. Because countries that allow dual citizenship may undercount resident U.S. citizens by counting dual citizens as their own, a variable was created to indicate countries that allow their citizens to maintain dual citizenship with the United States.

Some countries use ambiguous terminology, meaning it could not always be determined if a country was measuring U.S. citizens or U.S.-born individuals. The country of Kiribati in the Central Pacific serves as such an example. In Kiribati's census questionnaire, individuals are asked to list their "home country," but further clarification is not offered on whether the term refers to the individual's country of birth, country of citizenship, or an alternative definition. Other countries instead ask for each individual's nationality, but again do not specify how they define nationality. When these cases could not be resolved with certainty, they were excluded from the analysis.

FGEs are not available for every country and many release estimates on a cycle of every 5 or 10 years. In addition, some countries with complete data—foreign government data on U.S. citizens in their country, U.S. administrative data, and all other variables—still have errors in their FGEs because of the differences between registries and censuses. To have a complete and accurate estimate of the total number of overseas U.S. citizens, models were estimated to generate FGEs for all countries—those with complete data including FGEs, and those without an FGE. To accomplish this, U.S. administrative data on overseas citizens were collected, as well as additional predictors that research has demonstrated to be correlated with migration.

U.S. Administrative Records on Overseas Citizens

Several federal agencies collect data on overseas citizens and release statistics about subsets of that population. FVAP used these data to estimate the total number of U.S. citizens in each country. The key administrative data used were:

Number of U.S. Exchange Students, 2000–2018: Total number of U.S. exchange students attending foreign universities in each country for each year during the period 2000–2018, which is derived from data from the Open Doors Report on International Educational Exchange.[1]

Number of Social Security Beneficiaries, 2000–2018: Number of overseas Social Security beneficiaries for each year during the period 2000–2018, as reported annually by the U.S. Social Security Administration (SSA).

Number of Foreign Earned Income Returns, 2000–2016: Estimated number of Internal Revenue Service (IRS) Form 2555 returns (used to declare foreign income) filed by U.S. citizens living in select countries in 2001, 2006, 2011, and 2016.[2] Each form represents at least one U.S. citizen residing in the country.

Number of Civilian U.S. Federal Government Employees, 2000–2018: The number of civilian U.S. Federal Government employees residing in each country each year, as reported to the Federal Voting Assistance Program (FVAP) by the Office of Personnel Management (OPM).

Additional administrative records exist, such as overseas deaths, consulate registrations, and counts of military personnel. These data sources were not incorporated into estimates for two reasons. Some of these data are classified, sensitive, or otherwise not available to the general

[1] https://opendoorsdata.org/data/us-study-abroad/all-destinations/
[2] For more information on methodology see: Hollenbeck & Kahr, 2009

public; including them would have precluded other researchers from reproducing the results and, thus, undermined the transparency of these analyses. Also, these sources are likely to be strongly associated with tourism or military presence, rather than resident citizens; including them would add error by overestimating the number of U.S. citizens in countries with a military presence or a high volume of tourists from the United States.

Filling the Data Gap—Imputation and Estimation

Most modeling techniques require the predictor fields to be completely populated. Therefore, to use administrative data to model the U.S. overseas citizen population, missing data had to be addressed. In countries with low government capacity and with smaller populations, FGEs may be incomplete or nonexistent. Data from smaller countries may not be available if too few people meet a certain criterion. For example, there may be such a small number of U.S. tax filers living in Timor-Leste that the U.S. Government does not release records for Timor-Leste because of privacy considerations. It is probable that missing data is thus also correlated with migration, meaning that simply dropping country-years with missing data or substituting the mean would introduce bias into the estimates.

To model the full set of country-years without biasing the estimates, additional data were collected to impute the missing data. As the OECD explains, "Imputation is the process used to determine and assign replacement values for missing, invalid or inconsistent data [...] This is done by changing some of the responses or assigning values when they are missing [...] to ensure that estimates are of high quality and that a plausible, internally consistent record is created."

Missing U.S. administrative data was imputed by creating a predictive model that relies on variables known to be associated with higher levels of migration between countries, including:

- **Difference Between Foreign Country and U.S. Gross Domestic Product (GDP) per capita at Purchasing Power Parities (PPP) (Constant 2011 international dollars):** This variable is the difference between the PPP-converted GDP per capita of the foreign country and the United States in a given year in constant 2011 dollars, as reported by the World Bank's World Development Indicators. Research shows that countries with more favorable economic conditions are more attractive to U.S. citizens and, thus, have larger U.S. citizen populations. For countries for which this variable was missing (Taiwan, Cuba, Somalia), the data was imputed by regressing the log of the World Bank GDP per capita on the log of the GDP per capita provided by the Penn World Tables for a sample of countries in which both estimates were available. The resulting model was then used to impute the World Bank estimate for those countries with only a Penn World estimate. Version 9.1 of the Penn World Tables was used for Taiwan, and version 7.1 was used for Cuba and Somalia. The resulting predictions for Cuba and Somalia were extrapolated to 2018.
- **Population:** This variable refers to the population of the foreign country, as reported in the World Bank's World Development Indicators. The literature on international migration has typically found that countries with larger populations and economies tend to attract more migrants (Lewer & Van den Berg, 2008).
- **Distance From the United States:** This variable is the distance between the closest foreign city and U.S. city that both have a population over 750,000. For countries that do not have a city with a population over 750,000, the distance between the capital city of the foreign country and the closest U.S. city with a population of at least 750,000 was used. Distance has typically been found to be associated with lower levels of migration between two countries (Lewer & Van den Berg, 2008), likely because the larger distance is related to higher costs of migration (owing to factors such as travel and moving expenses).

- **Trade with the United States:** This variable refers to the mean end-of-year product trade (imports plus exports) between the United States and the foreign country, limited to the years 2000–2018, as reported by the Census Bureau. Trade has been linked to migration between trading countries (Felbermayr & Toubal, 2012; Sangita, 2013).
- **Institutional Quality:** This variable is the average of the six World Bank Worldwide Governance Indicators (WGI)—Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption—averaged across the years 1996–2018. This variable serves two purposes: First, research has found that institutional quality, and particularly the degree of political stability, is a determinant of net migration to countries (Ziesemer, 2010). Countries with good institutional quality are expected to have higher numbers of U.S. citizens. Second, countries with low governance quality are also likely to have poor FGEs, because they are unlikely to invest in the human capital of their bureaucracy.
- **Number of Immigrants in the United States:** This variable is the number of immigrants from a foreign country ages 25 and up in the United States in the year 2000 as reported by Artuc et al. (2013). One type of potential out-migrant from the United States is an immigrant from a foreign country (or their offspring) who then decides to return to his or her country of origin (Scheuren, 2012). A more general justification for including this variable is that it may serve as a proxy for factors that promote or inhibit migration both to and from the United States, such as transportation costs. Consequently, countries with larger numbers of immigrants in the United States would be expected to have larger numbers of U.S. citizens. On the other hand, the number of immigrants in the United States from a country may also be negatively associated with the number of U.S. citizens in that country, if factors that affect migration flows asymmetrically (such as political instability) are salient. The uncertainty regarding relationship direction is not a limitation for this predictor because the estimation strategy does not require an assumption of a positive or negative relationship.
- **U.S. Military Aid:** This variable refers to the total amount of military assistance in constant dollars made by the United States to the foreign country between 1946 and 2015 as reported by the United States Agency for International Development (USAID). Aid to foreign countries by the U.S. Government, and the associated interaction between those governments, may promote migration from the United States to the foreign beneficiary countries by facilitating the transfer of information about the foreign country to potential U.S. migrants (Berthelemy, Beuran, & Maurel, 2009). In addition, aid may be a proxy for general diplomatic ties (Alesina & Dollar, 2000) associated with foreign government policies that are advantageous to U.S. migrants, leading to increased U.S. migration to the country.
- **English or Spanish:** Two dichotomous variables indicating whether English or Spanish is spoken in the foreign country. The information is taken from Ethnologue: Languages of the World (Lewis, Grimes, Simons, & Huttar, 2009). This variable may be a proxy for cultural distance between the United States and the foreign country as well as for the ability to succeed in the host country's labor market (Adsera & Pytlikova, 2012). Given that English and Spanish are the two most widely spoken languages in the United States, countries where these languages are commonly spoken are expected to attract more U.S. citizens.
- **Trend:** This is a linear trend variable that controls for trends in the size of the overseas U.S. citizen population common to all countries and not explained by other theoretical variables. It accounts for variation in factors that affect migration to all other countries, such as advances in communication technology, changes in transportation costs, or general geopolitical factors. These factors may include population growth through births of U.S. citizens, whether overseas or within the United States, which would be expected to affect the total number of overseas U.S. citizens. This variable may also capture changes in transportation costs over the 2000–2018 period of study, which would also be expected to affect the tendency of U.S. citizens to migrate.

To impute data on exchange students, log-linear interpolation and extrapolation methods were used to determine values for missing years, as needed. Countries without a count for any year were assigned a value of zero.

For the SSA and IRS data, the missing data were imputed for countries for which there were no data. For the SSA data, most years had very reliable administrative counts on the total number of beneficiaries from a region (e.g., Africa) and by country. To impute the number of beneficiaries for African countries without counts, the number of beneficiaries from those countries that had a country count from the SSA was subtracted from the region total. For example, if there were 10,000 beneficiaries for Africa, but only South Africa included a count, and 500 beneficiaries were listed from South Africa, 500 were subtracted from the 10,000 regional total. There would be a remaining 9,500 beneficiaries to allocate to the countries without specific counts. To allocate the remaining beneficiaries, a model was created using the variables listed above.

We used this model to generate predicted numbers for those countries without estimates and distributed the unassigned beneficiaries of a region in proportion to that prediction. For example, a highly-populated African country where English is the primary language that has a relatively high GDP has more beneficiaries allocated to it compared to a highly populated French-speaking country in Africa with a relatively low GDP. A similar methodology was employed to generate estimates for the number of IRS returns for those countries for which the IRS does not already provide estimates. Once all countries had an estimate for the years for which data are available, estimates for the remaining years were produced using log-linear interpolation or extrapolation.

The collected and imputed data yield the final set of variables that were used to model the foreign country population estimates.

Estimating the Overseas Citizen Population

The data collected, along with the data imputations, yield the final set of variables used to model the foreign country population estimates. As noted above, FGEs are only available for some countries for some years and counts of demographic subgroups are available for even fewer countries and years. In addition, some countries with complete data—foreign government data on Americans in their country, U.S. administrative data, and all other variables—will still have errors in their FGEs because of the issues associated with registries, censuses, and other factors. Therefore, models were run to generate FGEs for all countries: those with complete data, including FGEs, and those without FGEs.

Several possible models and approaches can be used to develop this type of estimate. These models differ both in the underlying mathematical algorithms and in the choice of variables used to create the predictions. In an effective predictive model, the outcome variable (in this case, the population of U.S. citizens) is related to the predictor variables in a systematic way. Because the FGE is strictly positive and bounded from below at zero, each model was estimated using a Poisson regression. This model was run for every combination of predictor variables and then derived an average prediction.

The N models take the form:

$$\mathbf{FGE_{it}^{m}} = \mathbf{e}^{\beta C_{it} + \beta X_{it}^{m} + \gamma 1 REGISTRY_{it} + \gamma 2 CITIZEN_{it} + \gamma 3 DUAL_{it} + \gamma 4 (DUAL_{it} * CITIZEN_{it}) + constant}$$

FGE is the foreign government estimate of the size of the U.S. citizen population in country **i** in year **t** (i.e., there is at most one estimate for every country-year for the period 2000 to 2018).

C is a vector of variables reflecting the natural log of the size of particular subpopulations of the U.S.

citizen population and is thus highly likely to be correlated with the FGE. For this reason, these variables are included in every model. In these models, these variables are all of the U.S. Government administrative data for each country for each year.

X is a vector of predictor variables that are likely to explain variations in the U.S. citizen population of country i included in model m. These include the mobility variables described in the previous section. Because it is unknown which, if any, of the mobility variables improve model fit most effectively over a model with just subpopulation counts, models were run for every combination of mobility variables (including one specification with no such variables).

REGISTRY is a dichotomous variable that takes a value of 1 if the country's FGE is based on a registry count, and 0 otherwise.

CITIZEN is a dummy variable that takes a value of 1 if the FGE pertains to the number of U.S. citizens in the country, and 0 otherwise.

DUAL is a dummy variable that takes a value of 1 if the country allows dual citizenship with the United States, and 0 otherwise.[3]

DUAL * CITIZEN is an interaction variable that takes a value of 1 if the country both allows dual citizenship and has an FGE that counts U.S citizens, and 0 otherwise. The goal is to estimate the difference between the number of overseas U.S. citizens in countries that both allow dual citizenship and count the number of U.S. citizens, and countries that do not meet one or both conditions. Specifically, predictions are generated under the assumption that no country meets both conditions (i.e., DUAL*CITIZEN = 0) as it is under such circumstances that one is most likely to encounter citizenship misclassification and, thus, inaccurate citizen counts. In other words, citizenship-based FGEs for countries that allow dual citizenship are adjusted in such a way that the prediction incorporates dual citizens. To generate these predictions, REGISTRY is assumed to equal 0, CITIZEN is assumed to equal 1, and (DUAL * CITIZEN) is assumed to equal 0 for all countries.

The constraints applied to REGISTRY, CITIZEN, and the DUAL*CITIZEN product make the final predictions more comparable with respect to the population. To be specific, a count of U.S. citizens (i.e., CITIZEN = 1) is enumerated using a census (REGISTRY = 0).

[3] "Dual Citizenship" in this case means individuals can be citizens both of the country and the United States. Consequently, this variable is also coded as 1 for countries with that allow for citizenship for more than those two countries.

Averaging Across Models

Estimating the overseas U.S. citizen population was complicated because it was not clear which variables—and which combination of variables—should be used to model this population. To address this uncertainty, a variant of a method called ensemble Bayesian model averaging (EBMA) was used. EBMA has been found to yield more accurate predictions than using a single model when applied to predict armed conflict or the outcome of presidential campaigns (Montgomery et al., 2012). The general approach of EBMA is to take predictions from multiple models (i.e., ensembles) and create an average of all the estimates weighted by the model's fit to the data in combination with each model's correlation or redundancy with predictions derived from other models. The resulting estimate is designed to be more accurate than the estimates derived from any single model by minimizing the effects of overfitting the data resulting from individual model specifications. At the same time, this method also allows the final estimate to incorporate as much information as possible from the predictor variables.

Models

For the estimates of the overseas U.S. citizen population, the baseline model includes (1) all U.S. Government administrative data, (2) data about whether a country has a registry or census, (3) how that country counts a U.S. citizen, and (4) if the country allows dual U.S. citizenship. Additional models that include every combination of the migration research variables are also estimated.

The model space from which this average prediction is derived takes the form of all possible combinations of predictor variables. For k predictors, the number of models, N, equals 2^ (k) (including the model with no theoretical predictors, as described above). As applied to the estimation of overseas U.S. citizens, this approach is not likelihood-based (instead, it is based on root mean square error; see below) and, therefore, is not Bayesian. Consequently, the modeling approach is simply referred to as Ensemble Model Averaging (EMA).

The final estimate of the overseas U.S. citizen population for country i in year t is:

$$\exp\,(\mathrm{P_{it}}) = \exp\Big(\sum_{\mathrm{m=1}}^{\mathrm{N}} \mathrm{w^m P_{it}^m}\Big)$$

or the anti-log of the average of all linear predictions for the country across N models, weighted by model validation metric w.

The model validation metric w can be expressed in reduced form as:

$$\mathrm{w^m} = \frac{\mathrm{f^m} * \mathrm{c^m}}{\sum_{\mathrm{m=1}}^{\mathrm{N}} \mathrm{f^m} * \mathrm{c^m}}$$

In which fm is the component of the metric that indicates how well model m fits the data. fm can be written as:

$$f^m = \frac{(\frac{1}{MSE^m})}{\sum_{m=1}^{N}(\frac{1}{MSE^m})}$$

in which the MSE is the mean squared error. The MSE is determined through K-fold cross-validation (Stone, 1977); each observation in the sample is randomly assigned to one of K subsamples, the model is estimated using the K – 1 subsamples, predictions are produced for the excluded validation sample, and the MSE (weighted by the selection bias weight α_i, from above) is generated for that subsample. The cross-validation procedure is repeated K times, with each subsample acting as the validation sample in turn. The cross-validation step is then repeated S times, with the average of the S * K MSEs used as the model MSE. In this application, K = 5 and S = 10.

Overfitting and In-Sample Data

Overfitting often occurs when a model is made overly complex so that the results best fit the data being used for estimation (the "in-sample" data). This overfitting can affect the quality of the forecasting and prediction. The approach used here helps alleviate concerns about model overfitting by using model averaging and cross-validation.

Each model's contribution to the final estimate is therefore determined by its out-of-sample predictive ability, minimizing overfitting that could result from determining model performance based only on in-sample fit. Testing the model using countries that were not used to build the model allows for a more robust test, as its predictive power is more likely due to variation in the U.S. citizen populations in these countries and not random measurement error (Hawkins, 2004; Ward, Greenhill, & Bakke, 2010).

The other component of the model validation metric, cm, captures the degree to which the predictions generated by a model are correlated with predictions generated by other models. Specifically:

$$c^m = \frac{1/\sum_{j=1}^{N-1} \mathrm{Corr}\left(P^m, P^j\right)}{\sum_{m=1}^{N}(1/\sum_{j=1}^{N-1} \mathrm{Corr}(P^m, P^j))}$$

in which Corr is the correlation coefficient between models m and j. In other words, c^m is larger when a model is relatively uncorrelated with other models. The model validation metric wm is larger when models simultaneously (1) make relatively accurate out-of-sample predictions, and (2) are uncorrelated or not redundant with predictions made from other models. The validation metric, therefore, focuses on the models that are best at prediction, while also being sure to include a diverse set of model specifications rather than just minor variations of the same model. The proposed validation metric thus rewards accuracy and penalizes redundancy.

Mitigating Selection Bias

One potential issue with the modeling strategy outlined so far is that countries for which FGEs are available may have different characteristics than those where FGEs are not available. In particular, countries without FGEs tend to be poorly governed and tend to have relatively low economic output.

To account for this potential selection bias that may result from countries with FGEs being different in ways that may also affect the size of their overseas U.S. population, each country is given a weight for the purpose of model estimation:

$$\alpha_i = \frac{1}{\Pr(\text{FGE})_i * n_i}$$

in which Pr(FGE) is the predicted probability that a country has an FGE during the years 2000 through 2018 based on its observable characteristics and n is the number of years for which country i has an FGE. The predicted probability of having an FGE is generated using a logit regression in which the sample is all countries for which predictions are made. Predictor variables include all variables in vectors C and X in the estimation equation along with U.S. State Department region dummy variables. Data for the predictor variables for this selection equation were obtained for the year 2000. The result of the weighting is that countries with FGEs that have a low probability of having an estimate (based on the selection bias equation) have more weight when generating model parameters and predictions, resulting in more accurate EMA predictions for countries without estimates, and more accurate parameter estimates than those that would be generated in an unweighted model. This mitigates selection bias when there is not an unobserved factor (i.e., one not included in the model) that affects both the size of the FGE and whether a country has an FGE (Wooldridge, 2002). Including n in the denominator of the weight accounts for the overrepresentation of some countries in the sample because they have had FGEs for multiple years.

Estimating the Eligible Voter Population

To estimate the number of U.S. overseas citizens who are eligible to vote, the modeled estimates needed to be filtered to include only individuals who were 18 years and older. We started the estimation process using data from the Database on Immigrants in OECD Countries (DIOC). This data set provides counts of international migrants 15 years of age and older in OECD and some non-OECD countries by country of origin, divided into demographic groups defined by age, education, and sex. There are three age categories (15–24, 25–64, 65 and older), three education categories (No Education/Primary Education, Secondary Education, Post-Secondary Education), and two sex categories, for a total of 18 demographic groups. The population of U.S. citizens under the age of 15 was estimated for a subset of the DIOC country-years by subtracting the total population aged 15 and older from an available FGE to get the population under age 15, resulting in a total of 19 demographic groups encompassing the entire U.S. citizen population in a country.

However, the DIOC has not released new estimates since 2014, so additional estimates from IPUMS International data were used. The IPUMS International website organizes census microdata from countries across the world; these data were collected and aggregated to mirror the same population categories as the DIOC data. In cases in which data were available from both the DIOC and IPUMS for a given country-year, the IPUMS data were used. Unlike the DIOC data, the under-age-15 population was available in the IPUMS data and did not require imputation.

The model-averaging methodology was used to obtain predictions for both the aggregate population as well as the sizes of each age-sex-education group for all countries in the frame for the years 1996 to 2020. The size of each stratum was then rescaled so that the total number of U.S. citizens in each country across all groups was equal to the total number of U.S. citizens in each country as estimated in the updated 1996–2020 populations. In practice, after allocating the population across groups for each country, the group of individuals who were under age 15 was removed first, as was a proportion of the age 15–24 group who were under age 18. This was done by removing a proportion of those who do not have a high school education, equivalent to the proportion of the relevant domestic U.S. population who are age 15–17. The estimated counts by demographic strata were then used to obtain an estimate of the size of the eligible population. This ultimately resulted in estimates of the number of voting-eligible U.S. citizens residing in each country from the years 1996 to 2020.

Proposed Updates to the Methodology

In this section, we describe the potential methodological update. As explained in the previous section, the estimations of the Overseas Citizen Population (OCP) that have been used by FVAP are model-based. They are ultimately functions of the set of predictor variables explicitly specified in the model. Other characteristics of the country, either observable or unobservable, are assumed to have zero relationship with the size of the OCP or a given OCP subpopulation. The limited set of predictors means that the model's FGEs will inevitably fail to explain all the variance in the FGEs. This underfitting means that some potentially useful information about the size of the OCP in a given country-year contained in that country-year's FGE is not being incorporated in the final estimates. For those countries that contain at least one FGE, the proposed methodological update involves calibrating the population predictions of each model to match an adjusted FGE for those country-years for which such FGEs are available while using the changes in the Model-Based predictions to interpolate and extrapolate from the known FGEs. For countries without at least one FGE, the uncalibrated Model-Based estimates will continue to be used.

Under the updated methodology, for each model in the ensemble, following the generation of the natural log of the predictions for that model, the following steps are taken to calibrate the predictions:

Step 1: Identify the latest available FGE for a specific country. Identify all FGEs in that country that count the same population in the same way (e.g., all count U.S. citizens using a registry). This subset of a country's FGEs is the FGEs used for calibration, or our *target* FGEs. This step was undertaken so that the trends in the Calibrated estimates are not influenced by different trends across FGEs derived from different sources. The source with the last available FGE is used because the focus of the OCPA has traditionally been on calculating voting rates for the most recent election year.

Step 2: Each target FGE is adjusted using the same measurement adjustment methodology used for the Model-Based estimates such that the adjusted FGE ($\widehat{\text{FGE}}$), represents the model's estimate of what the FGE would have been if it had been a count of all U.S. citizens (including dual citizens) taken from a census.

Step 3: For each country-year for which there is not a target FGE, the most recent adjusted FGE preceding ($\widehat{\mathrm{FGE}}_{Pre}$), and the nearest (in time) adjusted FGE following ($\widehat{\mathrm{FGE}}_{Post}$), are identified. Up to two sets of adjusted estimates are made for each Model-Based estimate for which there is not a target FGE:

a. $Ln\left(\widehat{\mathrm{OCP}}_{t,Pre}\right) = Ln(\widehat{\mathrm{OCP}}_t) - Ln\left(\widehat{\mathrm{OCV}}_{Pre}\right) + Ln(\widehat{\mathrm{FGE}}_{Pre})$
b. $Ln\left(\widehat{\mathrm{OCP}}_{t,Post}\right) = Ln(\widehat{\mathrm{OCP}}_t) - Ln\left(\widehat{\mathrm{OCP}}_{Post}\right) + Ln(\widehat{\mathrm{FGE}}_{Post})$

Where $\widehat{\mathrm{OCP}}_t$ is the Model-Based prediction (i.e., the estimate from the original methodology) for the country year, $\widehat{\mathrm{OCP}}_{Pre}$ and $\widehat{\mathrm{OCP}}_{Post}$ are the Model-Based predictions for the country-years of $\widehat{\mathrm{FGE}}_{Pre}$ and $\widehat{\mathrm{FGE}}_{Post}$ respectively. $Ln\left(\widehat{\mathrm{OCP}}_{t,Pre}\right)$ and $Ln\left(\widehat{\mathrm{OCP}}_{t,Post}\right)$ are thus the Model-Based predictions subject to an adjustment such that the trends implied by the Model-Based estimates are maintained, but levels of the Model-Based estimates are such that the trend passes through the adjusted FGE.

Step 4: For those country years that are between two target FGEs, and thus their associated $Ln\left(\widehat{\mathrm{OCP}}_{t,Pre}\right)$ and $Ln\left(\widehat{\mathrm{OCP}}_{t,Post}\right)$, the weighted average of the two Calibrated estimates is taken, where the weight increases with relative proximity (in time) of the Model-Based estimate with the nearest adjusted FGE:

a. $Ln\left(\widehat{\mathrm{OCP}}_{t,adjusted}\right) = \frac{t_{Post}-t}{(t_{Post}-t)+(t-t_{Pre})} Ln\left(\widehat{\mathrm{OCP}}_{t,Pre}\right) + \frac{t-t_{Pre}}{(t_{Post}-t)+(t-t_{Pre})} Ln\left(\widehat{\mathrm{OCP}}_{t,Post}\right)$

For those county-years which precede the first available target FGE, $Ln\left(\widehat{\mathrm{OCP}}_{t,adjusted}\right) = Ln\left(\widehat{\mathrm{OCP}}_{t,Pre}\right)$, whereas for those that follow the last available target FGE $Ln\left(\widehat{\mathrm{OCP}}_{t,adjusted}\right) = Ln\left(\widehat{\mathrm{OCP}}_{t,Post}\right)$. For those countries without any available FGE, $Ln\left(\widehat{\mathrm{OCP}}_{t,adjusted}\right) = Ln(\widehat{\mathrm{OCP}}_t)$. Finally, for those country-years that have a target FGE, $Ln\left(\widehat{\mathrm{OCP}}_{t,adjusted}\right) = Ln(\widehat{\mathrm{FGE}}_t)$.

Step 5: The Calibrated set of estimates for each model is averaged using the same methodology described in the previous section to arrive at the final, Calibrated set of estimates.

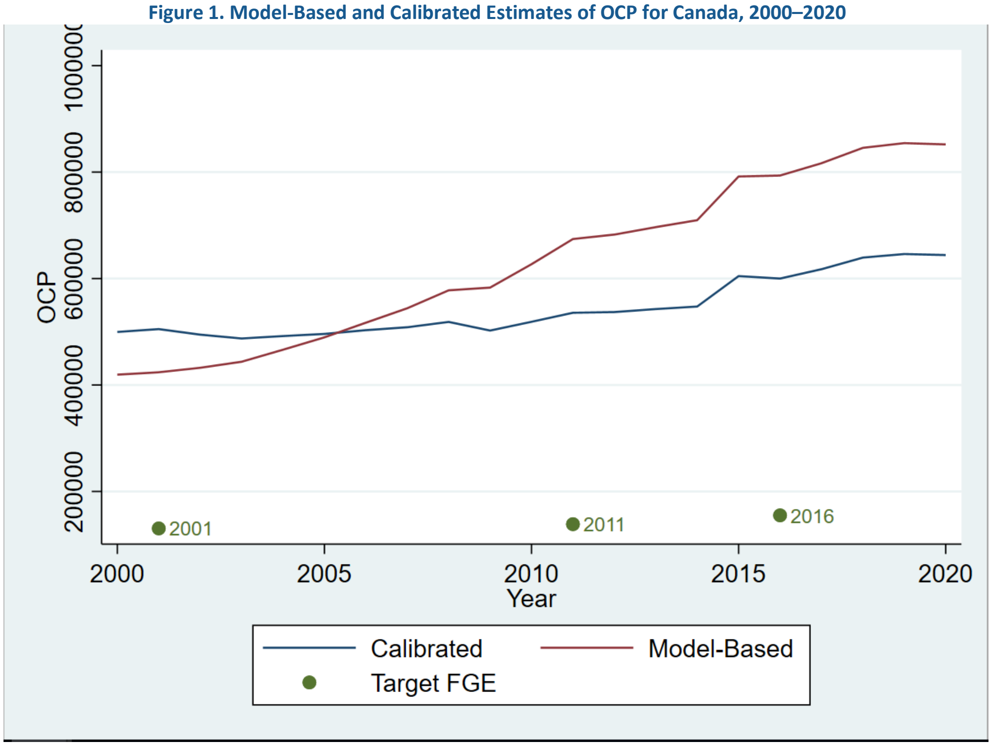
Step 6: Repeat steps 1–5 for each age-sex-education subpopulation. Apportion $\widehat{\mathrm{OCP}}_{P\ ,}$ across OCP subpopulations proportionate to the size of the estimated OCP subpopulation to arrive at the final estimate of the size of subpopulations. Sum OCP subpopulations age 18+ to arrive at the adjusted estimate of the Overseas Citizen Voting Age Population ($\widehat{\mathrm{OCVAP}}_{P\ ,}$).

These Calibrated estimates of the OCP by country-year will tend to be very similar to proximate FGEs, particularly when there is little growth or decline in a country's estimated OCP over time. The major discrepancies will tend to arise from the measurement error adjustments to account for who an FGE counts (i.e., U.S. born versus U.S. citizens) and how they are counted (census versus registry), which are a transparent outcome of the model. These Calibrated estimates thus will incorporate a large amount of information concerning the size of the OCP and OCVAP contained in the FGEs.

As an example of how the original Model-Based and Calibrated estimates of the OVP differ, Figure 1 presents the Calibrated and Model-Based estimates of the OCP for Canada for the period 2000–2020 along with the target FGEs used in the calibration. Note first the discrepancy between the

Calibrated OCP estimates and the target FGEs, where the former is significantly higher than the latter. This is because even the Calibrated OCP are subject to a model-based adjustment to account for who and how an FGE measures the population. In the case of Canada, this led to an upward adjustment to account for the fact that, as a country that allows dual citizenship with the United States, Canada may be failing to include these dual citizens in their counts. However, also note that, like the target FGEs, the trend in the OCP for Canada is relatively flat between 2001 and 2016. This reflects the influence of the Calibrated FGEs on Canada's OCP trend, as the calibration procedure forces the OCP estimates to pass through the *adjusted* target FGEs. By contrast, the non-Calibrated, purely model--based estimates show a stronger upward trend through the 2001–2016 period. In the case of Canada, this discrepancy in trends is likely partly because, despite the relatively limited growth in the OCP indicated by the Canadian Census, the number of American households in Canada reporting foreign income to the IRS roughly doubled from approximately 25 thousand in 2001 to approximately 50 thousand in 2016. Given that the number of IRS foreign income statements is an important determinant of the OCP Model-Based estimates, this resulted in relatively rapid growth in the latter. However, also note that the Calibrated and non-Calibrated OCP estimates are roughly parallel in the post-2016 period. This reflects that, even for the Calibrated estimates, the Model-Based predictions have a strong influence on extrapolating trends outside the range of the target FGEs.

Figure 1. Model-Based and Calibrated Estimates of OCP for Canada, 2000–2020



It should be noted that the FGEs are likely to contain some degree of measurement error. This was one reason why the original methodology averaged predictions from a large set of relatively parsimonious models rather than simply estimating one model with all predictors. Complex models may be subject to overfitting, whereby the resulting models are heavily influenced by idiosyncratic measurement error in the outcome variable, and thus do a poor job of predicting the outcome of interest out of the sample. Although the model averaging methodology was primarily motivated by the concern for maximizing out-of-sample accuracy (i.e., accuracy of predictions for country-years without FGEs), poor out-of-sample fit may also indicate measurement error in the FGEs. Our population of interest is individuals residing outside the United States who would have to vote absentee, and thus likely long-term residents of their host country. Foreign governments may over or underestimate this population based on how strictly they define residents. This error may be automatically incorporated into the Calibrated estimates given their close correspondence with the FGEs but may not be correlated with the Model-Based estimates if this error is uncorrelated with our predictors. Given this, it is unclear if the Calibrated estimates more accurately reflect the size and geographic distribution of the OCVAP than the Model-Based estimates produced under the original methodology. A later section of this report presents the results of a validation exercise to assess to what degree, if at all, one set of estimates better captures the OCVAP than the other.

Comparing the Model-Based and Calibrated Estimates

In this section, the Model-Based and Calibrated estimates of the OCVAP are briefly compared with respect to the OCVAP size and geographic distribution.

We start with a comparison of the OCP. In 2020, the non-Calibrated estimate of the OCP was 4,056,720, whereas the Calibrated estimate was 4,103,836. As can be seen in Table 1, the two sets of OCP estimates are largely similar with respect to the distribution of the OCP across world regions, though the Calibrated estimates assign a larger fraction of the OCP to North America than the Model-Based estimates. Additionally, the estimated growth of the OCP between 2000 and 2020 was generally higher in the Model-Based estimates than the Calibrated estimates, with the notable exceptions of Sub-Saharan Africa and North/Central/South Asia.

Table 1. Model-Based and Calibrated Estimates and Growth Rate of OCP by World Region

Region	OCP 2020		OCP Growth Rate 2000-2020	
	Model-Based	*Calibrated*	*Model-Based*	*Calibrated*
North America	1,382,801 (34%)	1,702,724 (41%)	75%	78%
South/Central America/Caribbean	408,651 (10%)	259,204 (6%)	71%	44%
Europe	1,087,420 (27%)	1,075,416 (26%)	77%	17%
Sub-Saharan Africa	96,646 (2%)	124,700 (3%)	146%	167%
Middle East/North Africa	266,643 (7%)	257,155 (6%)	142%	80%
North/Central/South Asia	128,185 (3%)	53,825 (1%)	183%	195%
East Asia	383,464 (9%)	311,470 (8%)	175%	138%
Southeast Asia	145,723 (4%)	136,910 (3%)	81%	29%
Oceania	157,189 (4%)	182,432 (4%)	149%	93%
Total	4,056,720	4,103,836		

This table provides Model-Based and Calibrated estimates of the 2020 total OCP and the growth rate of the OCP between 2000 and 2020. Proportions of world population by region are presented in parentheses in the first two columns. Model-Based estimates are derived from the original OCPA methodology. Calibrated estimates adjust the Model-Based estimates to match adjusted FGEs in country-years for which they are available. See the 2020 OCPA report for world regions definitions.[4]

The two sets of estimates are similar for countries with the largest OCP, though while Canada has the largest OCP under the Model-Based estimates, Mexico has the largest OCP under the Calibrated estimates. Another notable discrepancy is that the Model-Based estimate for France is more than double the country's Calibrated estimate.

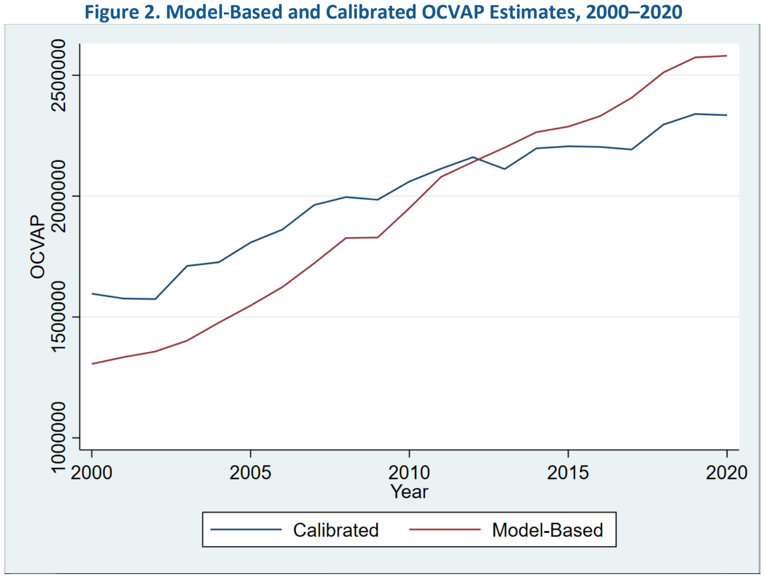
Table 2. Model-Based and Calibrated Estimates of Top 10 Countries by OCP

Rank	Model-Based		Calibrated	
	Country	*OCP, 2000*	*Country*	*OCP, 2000*
1	Canada	852,053	Mexico	1,058,483
2	Mexico	530,747	Canada	644,242
3	United Kingdom	314,008	United Kingdom	504,288
4	France	174,966	Australia	138,234
5	Israel	139,858	China	130,672
6	China	130,607	Israel	127,216
7	Australia	116,188	Germany	111,010
8	Japan	109,524	France	71,887
9	Switzerland	95,275	Ireland	71,295
10	Germany	90,583	South Korea	67,405

This table provides Model-Based and Calibrated estimates of the 2020 total OCP for the 10 countries with the largest OCPs. Model-Based estimates are derived from the original OCPA methodology. Calibrated estimates adjust the Model-Based estimates to match (adjusted) FGEs in country-years for which they are available.

The Model-Based and Calibrated estimates for the OCVAP were 2,580,690 and 2,334,653, respectively. Figure 2 presents the two sets of OCVAP estimates for the period 2000–2020. For the year 2000, the Calibrated estimate (1,596,408) starts higher than the Model-Based estimate (1,305,882), but by 2020 the Model-Based estimate exceeds the Calibrated estimates, reflecting the implied faster growth rate of the Model-Based estimate.

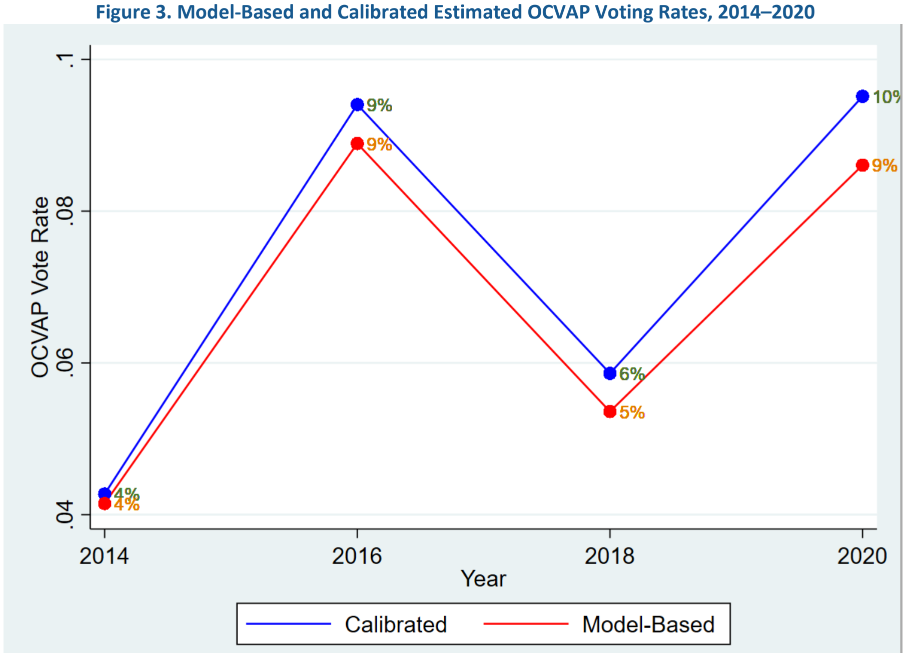
Figure 2. Model-Based and Calibrated OCVAP Estimates, 2000–2020



[4] https://www.fvap.gov/uploads/FVAP/Reports/OCPA-2020-Final-Report_20220805.pdf

The implications for these discrepancies on the OCVAP voting rate can be seen in Figure 3, which presents the OCVAP voting rates for the General Elections in the period 2014–2020 using both the Model-Based and Calibrated OCVAP estimates as the denominator. Both series are similar in levels and trend, though the Calibrated estimates consistently imply a higher OCVAP voting rate due to the fact that the Calibrated estimates are lower over the period for which voting data is available.

Figure 3. Model-Based and Calibrated Estimated OCVAP Voting Rates, 2014–2020



Turning to the geographic distribution of the OCVAP according to two sets of estimates, Table 3 presents estimated OCVAP sizes by world region for the period 2020. The largest absolute discrepancy in OCVAP between the two sets of estimates is in North America, which has 100,000 fewer OCVAP in the Calibrated estimates versus the Model-Based estimates. The largest proportional discrepancy is in North/Central/South Asia, which has a Calibrated OCVAP 50% lower than its Model-Based OCVAP. Overall, the two sets of estimates imply similar distributions of the OCVAP across regions, as indicated by the similar percentages of the population in each region (in parentheses). Although the two sets of estimates largely agree on geographic distribution of the OCVAP in 2020, there are large discrepancies in the implied growth rates between 2000 and 2020, with the Calibrated estimates indicating much slower growth than the Model-Based estimates in most regions, but particularly in North America, South/Central America/Caribbean, and Europe. Exceptions to this trend include Sub-Saharan Africa and South Asia, which saw faster growth under the Calibrated estimates than the Model-Based ones.

Table 3. Model-Based and Calibrated OCVAP Estimates by World Region

Region	OCP, 2020		OCP Growth Rate, 2000–2020		OCVAP, 2020		OCVAP Growth Rate, 2000–2020		Voting Rate, 2020	
	Model-Based	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*
North America	1,382,801 (34%)	1,702,724 (41%)	75%	78%	667,852 (26%)	555,291 (24%)	82%	39%	7.0%	8.4%
South/ Central America/ Caribbean	408,651 (10%)	259,204 (6%)	71%	44%	197,026 (8%)	128,079 (5%)	72%	30%	5.6%	8.6%
Europe	1,087,420 (27%)	1,075,416 (26%)	76%	17%	880,578 (34%)	876,421 (38%)	77%	21%	12%	12%
Sub-Saharan Africa	96,646 (2%)	124,700 (3%)	146%	167%	67,646 (3%)	84,614 (4%)	172%	202%	5.2%	4.1%
Middle East/North Africa	266,643 (7%)	257,155 (6%)	142%	80%	226,477 (9%)	216,656 (9%)	145%	80%	6.4%	6.7%
North/ Central/ South Asia	128,185 (3%)	53,825 (1%)	183%	195%	57,657 (2%)	26,459 (1%)	176%	220%	5.4%	11.8%
East Asia	383,464 (9%)	311,470 (8%)	175%	138%	256,601 (10%)	193,898 (8%)	177%	125%	5.5%	7.3%
Southeast Asia	145,723 (4%)	136,910 (3%)	81%	29%	84,427 (3%)	90,990 (4%)	108%	86%	10%	9.3%
Oceania	157,189 (4%)	182,432 (4%)	149%	93%	142,425 (6%)	162,245 (7%)	153%	99%	11%	9.5%

This table provides Model-Based and Calibrated estimates of the 2020 OCVAP, the growth rate of the OCVAP between 2000 and 2020, and the OCVAP voting rate in 2020 by world region. Proportions of world population by region are presented in parentheses in the first two columns. Model-Based estimates are derived from the original OCPA methodology. Calibrated estimates adjust the Model-Based estimates to match (adjusted) FGEs in the country-years for which they are available. See the 2020 OCPA report for world regions definitions.

Figure 4 displays a scatterplot of country-level OCVAP estimates, with the natural log of the 2020 Model-Based estimates on the horizontal axis and the natural log of the 2020 Calibrated estimates on the vertical axis. A 45-degree line where the two values are equal is also included. When a point is above the line, that indicates that the country has a higher Calibrated estimate than Model-Based estimate. Points further away from the line indicate a larger proportional discrepancy between the two sets of estimates. The scatterplot indicates a somewhat strong association between the two sets of estimates, with countries with a relatively high Model-Based estimate also having a relatively high Calibrated estimate (spearman rank correlation=.90, r=.87). One interesting tendency visible in the scatter plot is that proportional discrepancies tend to be larger for countries that both sets of estimates indicate have smaller OCVAPs. This may come from countries with larger true OCVAPs being more likely to have more FGEs on average and thus have a stronger presence in the sample used to estimate the OCVAP models, leading to the models being better able to predict the FGEs for countries with larger OCVAPs.

Figure 4. Calibrated Versus Model-Based Country Level OCVAP Estimates for 2020

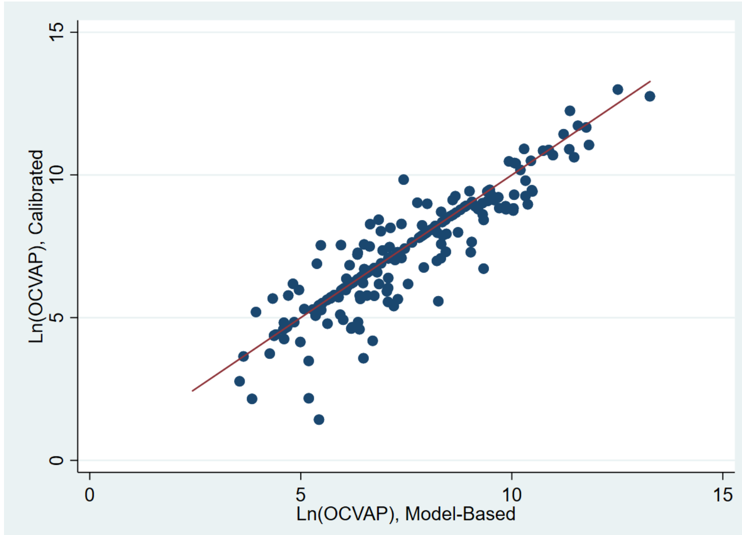


Table 4 lists the top 10 countries with the largest discrepancies between the Model-Based and Calibrated 2020 estimates. Although, as noted above, countries with likely smaller OCVAPs have the largest proportional discrepancies, the countries that host the likely largest OCVAPs still have the largest absolute discrepancies.

Table 4. Top 10 Countries with Largest Model-Based Versus Calibrated Discrepancies, 2020

Country	Model-Based OCVAP, 2020	Calibrated OCVAP, 2020	Difference (Calibrated – Model-Based)
Canada	580,475	346,795	-233,680
United Kingdom	271,652	439,968	168,316
Mexico	87,377	208,496	121,119
France	137,111	63,112	-73,999
Japan	96,457	40,997	-55,460
Switzerland	85,987	54,317	-31,670
Thailand	29,491	54,969	24,479
Poland	32,193	7,872	-24,322
Italy	35,892	12,372	-23,520
Greece	35,456	12,912	-22,545

This table lists the top 10 countries with the largest discrepancies between Model-Based and Calibrated estimates in 2020. The discrepancy/difference is provided in column 3. Model-Based estimates are derived from the original OCPA methodology. Calibrated estimates adjust the Model-Based estimates to match adjusted FGEs in country-years for which they are available.

Validating the OCVAP Estimates

In this section, the relative validity of the Model-Based and Calibrated estimates is assessed. To understand the logic behind the validation approach, consider two hypothetical countries whose OCVAP populations have an identical ballot request rate due to similar composition of the OCVAP with respect to perceived utility of voting, similar levels of infrastructure, and thus, similar obstacles to voting, etc. These two countries only differ in that one has a true OCVAP twice as large as the other. Consequently, we know that the number of ballot requests observed in one country will also be twice as large as the other. The validity of the two sets of estimates is thus measured by the degree to which a proportional increase in the estimated OCVAP is associated with an identical proportional increase in the number of ballot requests, as measured using the number of ballot requests in the ballot request file from which the vote counts used to generate the OCPA estimates of the OCVAP participation rates are drawn. More formally, this test is implemented by estimating the following linear model using Ordinary Least Squares (OLS) for each election (2014–2020) and each set of estimates:

$$X)\ Ln(Ballot\ Requesters)_i = \beta Ln(\widehat{OCVAP})_i + \delta X_i + e_i$$

Where $Ln(Ballot\ Requesters)_i$ is the natural log of the number of ballot requesters who requested their ballots from an address in country *i* in the given election; $Ln(\widehat{OCVAP})_i$ is the estimated OCVAP in country *i* during the given election; X_i is a set of country-level characteristics that potentially serve as a proxy for the determinants of the ballot request rate amongst OCVAP in that country, which include the World Governance Indicators, mailing times to the country. the interaction between the former, logged GDP per capita, logged distance to the United States, indicators for the predominant language in the country (English, Spanish, other) and region of the world fixed effects (see the 2020 OCPA report for more details on these variables); and e_i is an error term.

The parameter of interest is β, which can be interpreted as the percentage change in the number of ballot requesters that is associated with a doubling of the estimate OCVAP, holding the control variables (X_i) constant. If the estimated OCVAP corresponded exactly with the “true” OCVAP, we would expect $\beta = 1$. In practice, to the degree that a given set of estimates suffers from measurement error, under standard assumptions about the nature of that measurement error, $\beta < 1$ and is biased towards zero. We thus identify the more valid set of estimates as that which results in β that is closer to 1, which will likely also be the one with the larger β. A higher β is interpreted as indicating that cross-country variation in the estimated OCVAP is driven more by variation in the “true” OCVAP and less by measurement error. See Appendix A for a more detailed description of the logic of the validation check.

Using the number of ballot requesters to assess the relative validity of the Model-Based and Calibrated estimates is motivated by 1) the fact that the number of ballot requesters is the only proxy for the OCVAP in the country for which data is available for all of our countries of interest and 2) it should be strongly related to the precise definition of OCVAP for which FVAP is interested, namely U.S. citizens residing abroad under circumstances where they would have to use an absentee ballot to vote in a U.S. General Election (i.e. not tourists or other very short term visitors).

However, this test of validity rests on a couple of strong assumptions. One is that there are no unobserved differences between countries that are correlated with both the estimated OCVAP and ballot request rate. If characteristics of countries that determine both the size of the OCVAP and their willingness and/or ability to vote are not included in (or very strongly correlated with) X_i, then β may deviate from 1 even absent measurement error. This source of bias may differentially impact the two sets of estimates, and thus lead to poor inferences about relative validity. For example, the Model-Based OCVAP estimates are ultimately functions of the numbers of American Social Security recipients, individuals filing tax returns with the IRS, U.S. federal government employees, and students residing in a country. Each of these groups is either abroad temporarily or maintains some formal connection to the United States, and thus may see more utility in voting in a U.S. election (and thus requesting an absentee ballot) than the general OCVAP population. Consequently, variation in the Model-Based OCVAP estimates across countries may more strongly reflect variation in the number of "motivated" OCVAP residing in the country than the total number of OCVAP, and thus be correlated with the level of motivation amongst a country's OCVAP. The estimate for β may thus exceed 1 for the Model-Based OCVAP because it is strongly correlated with the ballot request rate in the country as well as the size of the OCVAP. By contrast, the Calibrated OCVAP estimates are likely less correlated with the ballot request rate, and as a result, the β may be lower for the Calibrated estimates than the Model-Based ones even if the former is a more accurate measure of the "true" OCVAP. A related strong assumption of the validation test is that the size of the OCVAP does not affect the ballot request rate through, for example, social network effects or stronger outreach efforts by stakeholder groups; if this assumption is violated, then β could deviate even if a given set of estimates did not suffer from measurement error. These limitations should be kept in mind when assessing the results of these validity tests.

The validation tests assess the estimates as a whole, not those of particular countries. One set of countries may, on the whole, be more informative but have greater error for particular countries than the counterpart estimate in the other set. As an example, models of migration between pairs of countries often include as a predictor an indicator for whether the countries are contiguous (i.e., share a land border).[5] In the case of the models used to generate OCVAP estimates, such an indicator was not included due to only two countries (Canada and Mexico) sharing land borders with the United States. Including such an indicator would potentially overfit the model, as the relationship between contiguity and the true OCVAP cannot be distinguished from measurement error in the Mexican and/or Canadian FGE. However, to the degree that contiguity is an important predictor of the size of the OCP and/or OCVAP, conditional on the predictors in the model, the Model-Based estimates may be particularly inaccurate with respect to Canada and Mexico due to this exclusion, and consequently, the Calibrated estimates may be relatively more reliable for these two countries, even if the Model-Based estimates were more reliable on the whole due to the fact that the overwhelming majority of countries are not contiguous. This limitation should be kept in mind when interpreting the results of the validation tests. To partially account for potential differences in estimated validity for the sample as a whole and accuracy for particular countries, validation tests were also conducted on a weighted sample of countries, where greater weight is given to countries that are expected to have relatively large OCVAPs (i.e., countries where the two sets of estimates are in agreement with respect to the relative size of the OCVAP).

Results for the validation tests are presented in the top panel of Table 5. For ease of interpretation estimates of both β, the elasticity of ballot requesters with respect to estimated OCVAP, and $1 - \beta$, our estimate of the degree of measurement error, are presented along with associated standard errors and p-values. The results indicate that, across elections, the elasticities for both the Model-Based and Calibrated estimates are statistically significantly lower than 1, consistent with both sets of estimates suffering from some degree of measurement error. However, across all elections, the Model-Based estimates consistently show a lower degree of measurement error.

One limitation of these initial results is that they give equal weight to all countries. However, FVAP

[5] See, for example: Lewer, J. J., & Van den Berg, H. (2008). A gravity model of immigration. Economics letters, 99(1), 164-167. Ortega, F., & Peri, G. (2013). The effect of income and immigration policies on international migration. Migration Studies, 1(1), 47-74.

may be more concerned with the accuracy of OCVAP estimates (and thus vote rates) of countries with larger OCVAPs. Consequently, in the bottom panel of Table 5, the country-sample is weighted by the country's OCVAP size, which is measured as the geometric mean of the Model-Based and Calibrated OCVAP size. Countries are thus given more weight when both sets of estimates agree that they are relatively large. The estimated magnitude of measurement error is consistently lower for the weighted sample versus the unweighted sample, indicating that the estimated OCVAP is more accurately measured in both sets of estimates for countries where the OCVAP is likely the largest. This is not surprising, as 1) countries with smaller OCVAPs have on average fewer FGEs in the samples used to estimate the model, and thus the trained models are likely to generate less accurate predictions for such countries, and 2) these countries are disproportionately developing countries, where the FGEs that do exist may be less accurate due to limited statistical capacity of those country's governments. The results of the weighted estimates continue to be consistent with the Model-Based estimates having less measurement error than the Calibrated estimates.

Table 5. Validation Tests for Model-Based and Calibrated OCVAP Estimates

	2014 (N=158)		2016 (N=158)		2018 (N=158)		2020 (N=158)	
	Model-Based	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*
	Unweighted							
β	.77	.54	.74	.52	.73	.51	.70	.50
	(.05)***	(.05)***	(.05)***	(.06)***	(.05)***	(.06)***	(.05)***	(.05)***
$1-\beta$	.23	.46	.26	.48	.27	.49	.30	.50
	(.05)***	(.05)***	(.05)***	(.06)***	(.05)***	(.06)***	(.05)***	(.05)***
	Weighted							
β	.92	.70	.90	.71	.89	.73	.87	.72
	(.05)***	(.06)***	(.05)***	(.06)***	(.05)***	(.06)***	(.06)***	(.06)***
$1-\beta$	.08	.30	.10	.29	.11	.27	.13	.28
	(.05)	(.06)***	(.05)**	(.06)***	(.05)**	(.06)***	(.06)**	(.06)***

This table presents estimates of the fraction of cross-country variation of OCVAP estimates that reflect variation in the true OCVAP (β) versus the fraction of cross-country variation due to measurement error ($1-\beta$) by methodology (Model-Based versus Calibrated) and election year. Estimates of β are derived from OLS regression estimate of the elasticity of the number of UOCAVA ballot requesters with respect to the estimated OCVAP by country, controlling for potential determinants of the country's ballot request rate (e.g., mailing times between the country and the United States, infrastructure, language, distance to the U.S., world region). Higher values of β are interpreted as indicating more accurate estimates of the geographic distribution of the OCVAP. Model-Based estimates are derived from the original OCPA methodology. Calibrated estimates adjust the Model-Based estimates to match adjusted FGEs in country-years for which they are available. Unweighted estimates give equal weight to each country when estimating β; weighted estimates give more weight to countries that likely have large OCVAPs, which is measured by the geometric average of the Model-Based and Calibrated estimates. Robust standard errors are in parentheses. Asterisks indicate the level of p-values of observed values under the hypothesis that β or ($1-\beta$) equal zero; *p<.10, **p<.05, ***p<.01.

Under the logic of the validation test, if a country's ballot request rate did not change over time, but its OCVAP did, one would expect to observe a proportional change in the number of requested ballots. This prompts a second set of validation tests that involve estimating the relationship between the growth rate in the estimated OCVAP between elections and the growth rate in ballot requests. Formally:

$$X)\ \Delta Ln(Ballot\ Requesters)_i = \beta \Delta Ln(\widehat{OCVAP})_i + \delta X_i + e_i$$

Where $\Delta Ln(Ballot\ Requesters)_i$ is now the change in the natural log of ballot requesters between two elections (e.g., 2014–2016, 2014–2018) and $\Delta Ln(\widehat{OCVAP})_i$ is the change in the estimated OCVAP between the same two elections. β has the same interpretation as in the equation above.

The benefit of exploring changes over time in the same country is that, if the unobserved obstacles to voting and motivational factors that influence the ballot request rate change very slowly over time (or are constant), they should have a limited correlation with $\Delta Ln(Ballot\ Requesters)_i$, and thus will not bias the results. The limitation is that any measurement error in the levels of the estimated OCVAP will result in an even greater degree of influence of that measurement error on the changes in the OCVAP. Consequently, β will likely be a very conservative test of the reliability of the OCVAP estimates, particularly if one is mostly concerned with comparing the OCVAP (or voting rates) between countries in the same election.

Table 6 presents differences for pairs of temporally adjacent elections and pairs of midterm and presidential elections ending in 2018 and 2020. The results indicate a much larger degree of measurement error in the changes in estimated OCVAP relative to the size of the OCVAP for both sets of estimates, with most of the variation in the election-to-election growth in OCVAP being noise. These results indicate that any estimated changes in the OCVAP or OCVAP vote rate for countries should be interpreted with extreme caution. Neither estimate consistently outperforms the other when it comes to trends in the OCVAP, though the point estimates indicate that the Calibrated estimates may be more informative about changes ending in 2020 (2018-2020, 2016-2020). This may reflect that some of the key predictors in the model (e.g., IRS tax filers) do not have recent updates and are imputed through extrapolation. These extrapolated values may be poor predictors of the change in 2020 OCVAP relative to previous years given the disruptions caused by the COVID-19 pandemic. By contrast, the Calibrated estimates may have incorporated some of the information about a sudden change in 2020 OCVAPs contained in available 2020 FGEs.

Table 6. Validation Tests for Model-Based and Calibrated OCVAP Estimates Using Changes Over Time

	2014–2016 (N=158)		2016–2018 (N=158)		2018–2020 (N=158)		2014–2018 (N=158)		2016–2020 (N=158)	
	Model-Based	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*	*Model-Based*	*Calibrated*
Unweighted										
β	.25	.16	.34	.16	-.19	.04	.28	.27	.11	.19
	(.20)	(.15)	(.11)***	(.08)**	(.16)	(.08)	(.10)***	(.10)***	(.11)	(.12)
$1-\beta$	.75	.84	.66	.84	1.19	.96	.72	.73	.89	.81
	(.20)***	(.15)***	(.11)***	(.08)***	(.16)***	(.08)***	(.10)***	(.10)***	(.11)***	(.12)***
Weighted										
β	.12	.24	.33	.08	.15	.27	.28	.16	.33	.43
	(.31)	(.23)	(.13)**	(.11)	(.24)	(.05)***	(.18)	(.20)	(.15)**	(.07)***
$1-\beta$	.88	.76	.67	.92	.85	.73	.72	.84	.67	.57
	(.31)***	(.23)***	(.13)***	(.11)***	(.24)***	(.05)***	(.18)***	(.20)***	(.15)***	(.07)***

This table presents estimates of the fraction of cross-country variation of the percentage change in OCVAP estimates between given elections that reflect variation in the percentage change in the true OCVAP (β) versus the fraction of cross-country variation in the percentage change due to measurement error ($1 - \beta$) by methodology (Model-Based versus Calibrated) and election year. Estimates of β are derived from OLS regression estimate of the elasticity of the percentage change in the number of UOCAVA ballot requesters between two elections with respect to the percentage change in the estimated OCVAP by country between the same elections, controlling for potential determinants of the country's ballot request rate (e.g., mailing times between country and the United States, infrastructure, language, distance to the U.S., world region). These variables do not change over time. Higher values of β are interpreted as indicatingmore accurate estimates of the relative growth rate of the OCVAP between two elections across countries. Model-Based estimates are derived from the original OCPA methodology. Calibrated estimates adjust the Model-Based estimates to match (adjusted) FGEs in country-years for which they are available. Unweighted estimates give equal weight to each country when estimating β; weighted estimates give more weight to countries that likely have large OCVAPs, which is measured by the geometric average of the Model-Based and Calibrated estimates in the end year. Robust standard errors are in parentheses. Asterisks indicate the level of p-values of observed values under the hypothesis that β or $(1 - \beta)$ equal zero;– *p<.10, **p<.05, ***p<.01.

Overall, the results of the validation tests indicate that both the Model-Based and Calibrated estimates suffer from measurement error. Accepting the assumptions of the validation test, the results indicate that the Model-Based estimates are likely more informative about the difference in the size of the OCVAP at any point in time, but neither set of estimates is informative about country-level trends in the OCVAP.

Summary and Conclusion

This research note has two primary goals. The first is to create updated 2020 estimates of the Overseas Citizen Voting-Age Population (OCVAP) in anticipation of the 2022 OCPA report. The second is to examine whether changing the OCVAP estimation methodology to bring the OCVAP estimates closer to those of overseas U.S.-origin populations reported by foreign government statistical agencies would lead to more reliable estimates than the current, strictly Model-Based methodology. The current Model-Based methodology uses those FGEs to fit the model but relies solely on model predictions for the final estimates. The relative validity of the two sets of estimates was assessed by estimating the proportional increase in the number of ballot requesters that was associated with a doubling of an OCVAP estimate for each set of estimates. A set of estimates that demonstrated a greater association (closer to 1) than the other is interpreted as being more informative about the size of the "true" OCVAP.

The key results of this analysis are as follows:

- The size of the global 2020 OCVAP is similar in the Model-Based (2.6 million) and Calibrated (2.3 million) estimates. This results in roughly similar estimates of the OCVAP voting rate, with a 2020 Model-Based OCVAP voting rate of 9% and a Calibrated OCVAP voting rate of 10%.
- The two sets of estimates of the global OCVAP demonstrate different dynamics, with the Model-Based estimate increasing faster between 2000 and 2020 than the Calibrated estimates.
- Regarding the geographic distribution of the OCVAP, the two sets of estimates correspond somewhat closely in their distribution of the OCVAP across world regions and the ranking of countries by OCVAP size, though there are notably large absolute discrepancies.
- The validation tests indicate that both sets of estimates are at least somewhat informative about the relative size of country-level OCVAPs within a given election, particularly with respect to countries that are likely to have large OCVAPs, but with the Model-Based estimates generally being more reliable.
- Neither set of estimates are found to be reliable concerning election-to-election changes in the size of the OCVAP.

Accepting the caveats and assumptions of the validity tests, the conclusion of these results is that calibrating the OCPA Model-Based estimates to more closely match the estimates of the U.S. citizen populations given by foreign governments is unlikely to result in more reliable counts of the OCVAP and voting rate. This may be due to measurement error in the FGEs, particularly those from developing countries, as well as idiosyncrasies in how the statistical agencies of different governments define their resident American population. By contrast, the original, more model-based approach appears to be successful in extracting most of the information about the size of the "true" OCVAP while discarding much of the measurement error. It is therefore recommended that FVAP continue using the original OCPA methodology.

In the future, finding additional sources of information for the relative sizes of different countries'

OCVAP may help improve the econometric models underlying the OCPA estimates. Even if these alternative sources of data cannot be used for such a purpose due to limited geographic or temporal scope, they may still be used as an alternative to the number of ballot requests for the purpose of validating the number of ballot requesters. Web analytics and social media activity attributed to U.S. citizens residing abroad may be one source of such data.

Appendix A: Formalization of Validation Test

The number of ballot requesters equals the size of the true OCVAP multiplied by the true ballot request rate.

$$1)\ \#BallotRequesters = OCVAP * BallotRequestRate$$

Taking logs gives the following linear equation:

$$2)\ Ln(\#BallotRequesters) = \beta Ln(OCVAP) + Ln(BallotRequestRate)$$

The elasticity of ballot requesters with respect to the true OCVAP (β), after controlling for the ballot request rate, thus equals 1. In a sample of countries, where $Ln(OCVAP)$ are uncorrelated with $Ln(BallotRequestRate)$, the simple OLS regression of $Ln(\#BallotRequesters)$ on $Ln(OCVAP)$ would yield:

$$\beta = \frac{Cov(Ln(\#BallotRequesters),\ Ln(OCVAP))}{Var(Ln(OCVAP))}$$

Because $\beta = 1$:

$$3)\ Cov\big(Ln(\#BallotRequesters),\ Ln(OCVAP)\big) = Var\big(Ln(OCVAP)\big).$$

The (log of the) estimated OCVAP, as opposed to the (log of the) true OCVAP, can be written as:

$$Ln\big(\widehat{OCVAP}\big) = Ln(OCVAP) + e$$

Where e is random measurement error. The elasticity of $(\#BallotRequesters)$ with respect to $\big(\widehat{OCVAP}\big)$ is:

$$4)\ \hat{\beta} = \frac{Cov(Ln(\#BallotRequesters),\ Ln(\widehat{OCVAP}))}{Var(Ln(\widehat{OCVAP}))} = \frac{Cov\big(Ln(\#BallotRequesters),Ln(OCVAP)\big)+Cov\big(Ln(\#BallotRequesters),e\big)}{Var\big(Ln(OCVAP)\big)+Var(e)+2Cov\big(Ln(OCVAP),e\big)}$$

Under the assumption that e is uncorrelated with both the number of ballot requesters and the true OCVAP, Equation 4 can be simplified to:

$$4a)\ \hat{\beta} = \frac{Cov\big(Ln(\#BallotRequesters),Ln(OCVAP)\big)}{Var\big(Ln(OCVAP)\big)+Var(e)}$$

And substituting Equation 3 into Equation 4a yields:

$$4b)\ \hat{\beta} = \frac{Var\big(Ln(OCVAP)\big)}{Var\big(Ln(OCVAP)\big)+Var(e)}$$

The elasticity of the number of ballot requesters with respect to the estimated OCVAP is thus strictly bound between 0 and 1 and can be interpreted as the fraction of variance in the estimated OCVAP is attributable to variance in the true OCVAP, whereas, conversely, $1 - \hat{\beta}$ is the fraction of variance in the estimated OCVAP that is due to measurement error.