Producing Estimates of International Migration for U.S. States: A Structural Time Series Approach Using Administrative Flight Data*October 2025

Andrew C. Forrester[1,2]
Srijeeta Mitra[1]

[1]Joint Program in Survey Methodology
University of Maryland, College Park

[2] U.S. Bureau of Labor Statistics

Abstract

Recent efforts across the federal statistical system aim to produce more accurate population estimates that incorporate international migration. Research by the U.S. Census Bureau relies on new administrative data to measure international migration into the U.S.; however, such data are often unavailable for subnational geographies, such as states. In this note, we leverage administrative data on inbound flights from the Bureau of Transportation Statistics and airline passenger statistics from U.S. Customs and Border Protection to produce novel monthly, statewide estimates of immigrant admissions to the U.S. Our methodology utilizes structural time series models that directly model the trend and seasonal patterns of migration. Additional outlier detection enables us to adapt the estimates to real-time shocks to international travel. We conclude by discussing further work using the time series method, some potential pitfalls, and how researchers might adopt this approach when adapting administrative data to estimate U.S. immigration.

Keywords: immigration, demography, time series, state-space

*All views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.

1. Introduction

International migration flows are a difficult component of population change to measure, as both *recent* and *accurate* data are often scarce (U.S. Census Bureau, 2024; Jensen et al., 2024). Unlike fertility and mortality, which are typically captured through well-established vital statistics systems, migration remains the least well-measured demographic process worldwide. The movement of people across borders is influenced by a complex interplay of economic, political, environmental, and social factors, each of which can change rapidly, making timely and reliable measurement especially challenging. Demographers must often rely on survey data and administrative data to overcome challenges ranging from underenumeration in both surveys and censuses, duration and nature of residence, national policy towards immigration, and exogenous events impacting migration (United Nations, 2022).

In recent years, statistical agencies and researchers have increasingly turned to data integration and model-based approaches to address these gaps. By linking administrative, survey, and auxiliary data, these methods capture short-term migration dynamics while maintaining consistency with long-term demographic trends. Recent work by the U.S. Census Bureau aims to improve international migration estimates by synthesizing administrative data across the U.S. Federal Statistical System (FSS) to improve prior survey-based estimates (U.S. Census Bureau, 2024). These efforts reflect a broader shift toward leveraging administrative big data to modernize population estimates and produce nearreal-time demographic indicators.

Drawing on similar data sources, we propose a model-based framework to improve the forecast quality of international migration inflows by modeling monthly nonimmigrant (e.g. business travelers, tourists, and students) arrivals by country of origin with a related administrative series: inbound airline passenger flight data collected by the Bureau of Transportation Statistics (BTS). Our approach borrows strength from a related and likely comoving time series using a structural time series approach that also flexibly accounts for trends, seasonality, and other exogenous shocks to migration (such as natural disasters or policy changes). The time series approach enables us to produce more reliable international migration forecasts by leveraging historical migration patterns and related auxiliary data.

Building on these recent developments, this note specifically explores whether Air Passenger Traffic (APT) data can strengthen model-based projections of nonimmigrant arrivals (as measured by I-94 records). APT data provide a near census of airborne travel to the United States from abroad, containing both origin and destination information, making them an ideal administrative covariate for modeling migration inflows. Moreover, because APT data are released more regularly than I-94 records, they offer the potential to enhance both the timeliness and accuracy of concurrent international migration estimates. Ultimately, improving migration forecasts has important implications for demographic estimation, economic planning, and policy formulation, particularly in contexts where migration constitutes a major driver of population change. Subsequent work will integrate these projections with demographic detail to produce alternative migration forecasts for the U.S. and states.

2. Literature Review

Early approaches to estimating international migration within the U.S. Census Bureaus Population Estimates Program (PEP) relied primarily on survey-based measures, with only limited use of administrative data. Historically, administrative sources were incorporated sparingly, with one key exception being the use of Air Passenger Traffic (APT) data to directly estimate migration flows between the mainland United States and Puerto Rico (U.S. Census Bureau, 2024). For the continental United States, however, the PEP traditionally derived international migration estimates from the residence one year ago (ROYA) question in the American Community Survey (ACS). While this approach ensured consistency with other Census Bureau estimates, it was constrained by sampling error,

reporting biases, and delays in reflecting emerging migration dynamics.

Recent demographic and policy shifts have highlighted the limitations of relying solely on survey data, particularly in capturing short-term or humanitarian migration flows. Large increases in refugee and humanitarian admissions in recent years were not adequately captured by ACS-based methods, resulting in systematic underestimation of international migration levels (Jensen et al., 2024). To address these shortcomings, the Census Bureaus new PEP methodology integrates multiple administrative data sources to derive more accurate and timely estimates. These sources, compiled into a consolidated Benchmark Database, provide a data-driven foundation for estimating international migration and improving the precision of annual population updates (Gross et al., 2024). There are important distinctions between an immigrant's class of admission and their current status. The focus of this note is to measure the inflow of *nonimmigrants*. Nonimmigrants are temporary visitors to the U.S. and include temporary workers and their families, students, tourists and business visitors, and diplomats among other categories (Bryan Baker, 2025). As of fiscal year 2024, the U.S. Department of Homeland Security (DHS) estimates that around 3.6 million nonimmigrants reside in the U.S. (Bryan Baker, 2025). However, as nonimmigrants are temporary residents by legal definition, flows into and out of the nonimmigrant population far exceed the level of current residents. Over the previous fiscal year, DHS granted admission to 132 million nonimmigrants (Rukh-Kamaa, 2024). The primary method used by the DHS uses administrative records, specifically I-94 records, to estimate this population and is the basis for our choice of methodology.

The shift from survey-based estimation at PEP to an administrative-data-driven framework represents a substantial methodological advancement. It reflects a broader movement in official statistics toward data integration and adaptive modeling, aligning with international efforts to leverage administrative and big data sources for improving migration measurement. Thus far, applications of administrative data to measure migration typically focus on internal migration. One key example is the matched tax return method used by the Population Estimates Program to measure domestic migration rates (U.S. Census Bureau, 2024). Emerging research has explored how high-frequency administrative indicators can be incorporated into a model-based framework to enhance the timeliness and forecasting accuracy of international migration estimates. Addressing this gap, the present study introduces a complementary approach that builds on existing Census Bureau efforts while extending them through the integration of additional administrative signals.

This paper makes two key contributions:

- **Use of high-frequency administrative data**: We incorporate Air Passenger Traffic (APT) data as a leading indicator to improve short-term projections of nonimmigrant arrivals, enhancing the timeliness and sensitivity of migration estimates.
- **Model-based integration through structural time series**: We apply a structural time series framework to jointly model visa issuances and APT data, capturing shared trends, seasonality, and shocks to produce more adaptive migration forecasts.

3. Data

As discussed, our analysis relies on two primary datasets capturing monthly international arrivals to the United States, spanning January 2015 through 2025. These sources provide complementary perspectives on migration and travel patterns, enabling the integration of administrative and survey-based information in a model-based framework.

1. **I-94/I-94W Arrivals:** The first data source is the monthly I-94 nonimmigrant arrival data collected by U.S. Customs and Border Protection (CBP) through Arrival and Departure Information System (ADIS)

and published by U.S. Department of State and the U.S. Department of Commerce. The dataset covers all origin countries and includes counts of nonimmigrant arrivals by nationality and intended destination. This information provides detailed insights into international travel and migration patterns to the United States over the study period (from 2015 to 2025). Notably, these data include *immigrant arrivals*.

2. **Airline Passenger Traffic by Origin:** To supplement the I-94 data, we incorporate inbound international airline passenger traffic from the U.S. Bureau of Transportation Statistics (BTS). This dataset reports monthly totals of passengers arriving in the U.S., disaggregated by destination state and country of origin, spanning January 2000 to the present. The high-frequency and near-complete coverage of this dataset makes it an ideal auxiliary source for modeling migration flows. The data is publicly available from the U.S. Bureau of Transportation Statistics.

In the next section we describe how these data sources will be combined and analyzed using time series methods.

4. Methodology

We combine these two to obtain monthly incoming passengers as well as the volume of nonimmigrant arrivals broken out by origin country and the time period of arrival (year and month). Our final dataset includes monthly time series data spanning Jan. 2015 to Dec. 2024 that measure inflows into the U.S. by state and country of origin from two administrative data sources. Together, the datasets provide both official counts of nonimmigrant arrivals (I-94) and a near-census measure of international air travel (APT), allowing for a more comprehensive and timely modeling of migration inflows to the United States.

Given the time series nature of each measure of inbound movement (international migration and travel), we adopt a structural time series (STS) approach to model both the relationship between nonimmigrant arrivals (I-94) and airline passenger traffic (APT) and their respective time series components. The time series components include the general trend-cycle $T_t^{(k)}$ of each series, their seasonality $S_t^{(k)}$, relevant outlier regressors $O_t^{(k)}$, and irregular disturbances $I_t^{(k)}$. This approach borrows strength across two dimensions. First, the modeled covariate (APT) provides a measure of how much inbound travel is going into a state from abroad each month, of which international migration is a component. Second, the time series components help describe how each series evolves over time. Modeling the overall level of nonimmigrants entering a state provides a baseline against which seasonal and irregular components can be measured. Adding a seasonal component allows us to also model seasonal fluctuations in travel, such as travel from abroad for vacations. Combined, we are able to borrow predictive strength through both a covariate and the individual series themselves.

To formalize our empirical model, let the triple $\{s,o,t\}$ identify state s, country of origin o, and month t. Our response variable is the monthly flow of new nonimmigrant I-94 arrivals $\texttt{I94}_{s,o,t}$ and our covariate is inbound airline passenger traffic $\texttt{APT}_{s,o,t}$.

For a single state-origin pair, our bivariate system takes the form

$$\begin{aligned} \texttt{I94}_t &= T_t^{I94} + S_t^{I94} + O_t^{I94} + I_t^{NIV}, \\ \texttt{APT}_t &= T_t^{APT} + S_t^{APT} + O_t^{APT} + I_t^{APT}, \\ \left[I_t^{I94}\ I_t^{APT}\right]' &\sim \mathcal{N}(\mathbf{0}, H_t), \end{aligned} \tag{1}$$

where H_t is the error covariance matrix. Following the bivariate model described in Tiller (2006), we model correlated shocks in the level and slope terms as

$$
\begin{aligned}
T_t^{I94} &= T_{t-1}^{I94} + R_t^{I94} + \eta_t^{I94}, \\
T_t^{APT} &= T_{t-1}^{APT} + R_t^{APT} + \eta_t^{APT}, \\
\left[\eta_t^{I94},\ \eta_t^{APT}\right] &\sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{T^{I94}} & \sigma_T \\ \sigma_T & \sigma^2_{T^{APT}}\end{bmatrix}\right).
\end{aligned}
\tag{2}
$$

Correlated levels and slopes allow shocks to the level of overall airline passenger traffic to inform the level of I-94 admissions. For example, if a natural disaster resulted in a stronger air passenger outflow from a particular origin country, the inflow shock would propagate through the system to inform the level of I-94 admissions.

The seasonal component is specified independently for each series using a trigonometric formulation (Durbin and Koopman, 2012). This specification models seasonality through harmonic terms corresponding to each seasonal frequency, ensuring a smooth and cyclical pattern over time. Estimating the seasonal effects separately for each series allows for flexible dynamics without imposing unnecessary parameter restrictions across series.

One addition to the basic structural model is the inclusion of outlier regressors $O_t^{(k)}$ for each series and are modeled with fixed effects. Outliers were identified in a pre-processing step for each series using X-13ARIMA-SEATS. Specific outliers tested included Additive Outliers (AO), Level Shifts (LS), and Temporary Changes (TC). We expect outliers to be a modeling concern as our sample includes the 2020 Global Pandemic. During the Pandemic, the U.S. and other countries imposed a variety of travel restrictions that directly impacted global air travel—affecting both nonimmigrant arrivals (I-94) and total arrivals that includes USCs (APT).

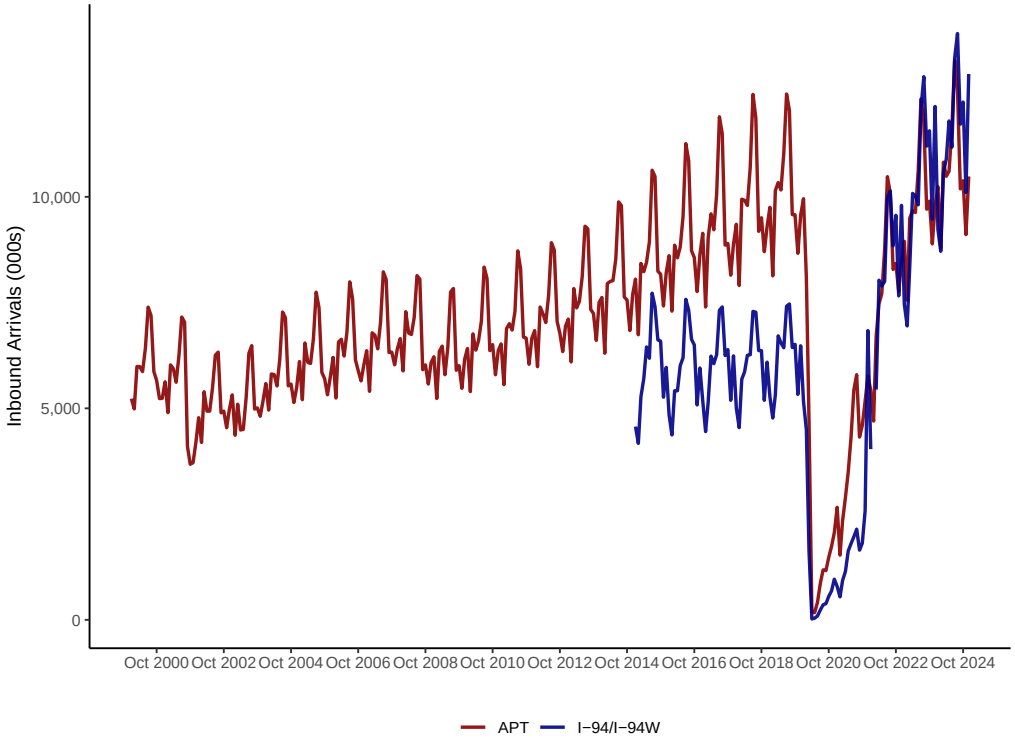
Parameter estimation proceeds by mapping the structural model into state space and maximizing the Gaussian log-likelihood via the Kalman filter with diffuse initialization, following Durbin and Koopman (2012). The covariance parameters are estimated by numerical optimization, while regression coefficients are obtained by generalized least squares conditional on the state parameters. Given the estimated parameters, filtered, smoothed, and forecast state estimates are obtained through the standard Kalman recursions and De Jong (1989) smoothing, providing point estimates and their associated variances for both the latent components and observed signals.

5. Results

Our main objective is to evaluate whether APT data might be a useful predictor to improve model-based estimates of nonimmigrant arrivals as measured by I-94 arrivals. In this exploratory analysis, we first compare the time series of I-94 arrivals against the comparable APT series on both an aggregate basis and state-by-state. We aim to examine whether the structural time series approach is appropriate, namely whether we can borrow predictive strength from the APT data as its time series components relate to I-94 arrivals. Recall, APT represents a near census of inbound air travel and, as such, nonimmigrant arrivals should be a fraction. Figure 1 plots the two time series aggregated to the nation (fifty states and the District of Columbia) to compare both the level and seasonality across each series.

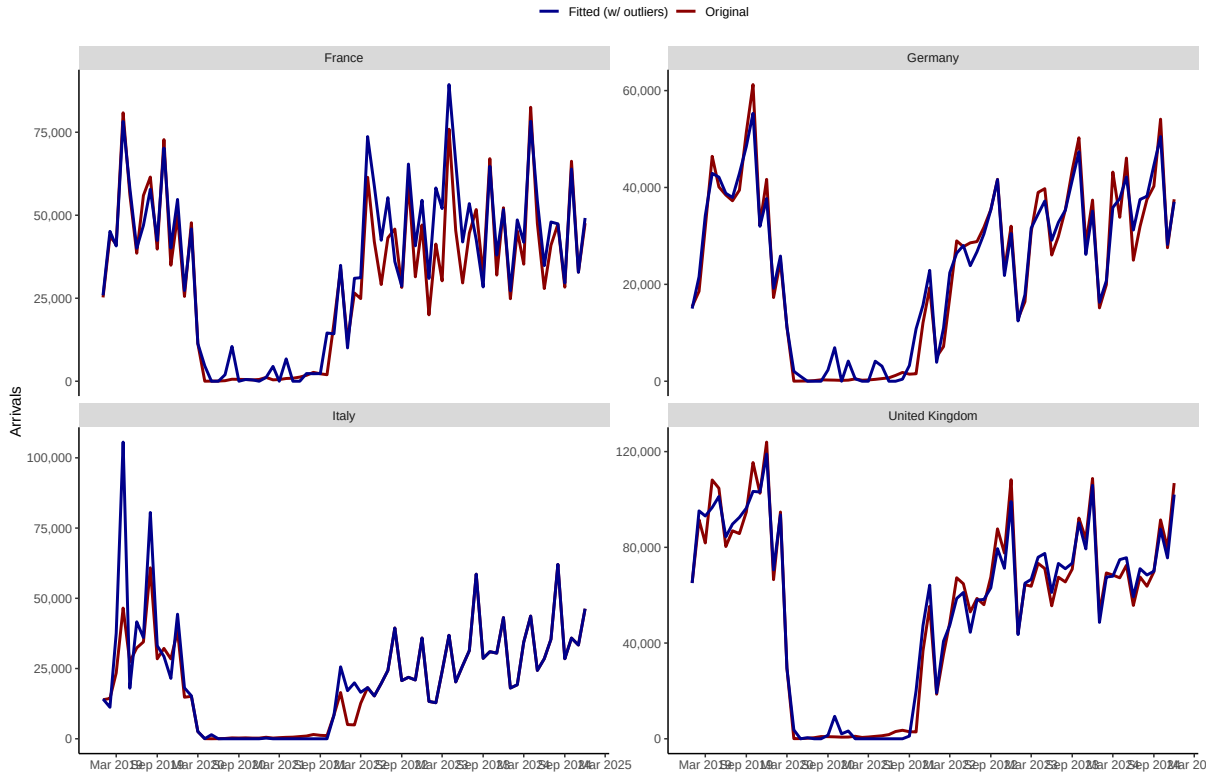
As expected, each series follows a similar seasonal pattern, albeit at a different overall level. This is mostly explained by the conceptual difference between the two measures—only a subset of inbound APT must submit and I-94 form for entry into the U.S. As such, we expect large level differences between the two series, thus motivating the inclusion of a local level component. Interestingly, following the COVID-19 Pandemic (indicated by a dashed line) we notice that I-94 arrivals recovered at a slower pace than overall APT. A likely explanation for this is the slowdown of global tourism following the Pandemic (Yepez and Leimgruber, 2024). For example, looking at I-94 arrivals broken out by type of visa, the number of arrivals on tourism visas decreased 35.6 percent

Figure 1: Aggregate I-94 (air mode only) Arrivals and Airline Passenger Traffic (APT) by Month



Source: Authors calculations from U.S. Bureau of Transportation Statistics and U.S. Customs and Border Protections

Figure 2: I-94 Arrivals and Inbound APT for Selected Countries, New York



Source: Authors calculations from U.S. Bureau of Transportation Statistics and U.S. Department of Commerce

comparing 2019 to 2022. Accordingly, the slower catch-up relative to overall APT motivates the inclusion of a slope term to allow the trajectory of each series to differ across each series. With evidence that the structural time series approach is reasonable at the national level, we now apply the proposed model to statewide data.

For this evaluation, we fit four separate bivariate models for matched I-94 arrivals and inbound APT for selected countries in the State of New York. This selection allows us to test two aspects of our method. First, New York's nonimmigrant travelers from the four selected countries was relatively consistent (except during the Pandemic) and large in level over time. As such, we expect the structural time series approach to perform well at detecting and modeling level, slope, and seasonality components. Second, we are able to leverage the Pandemic period to test our outlier pre-testing capacity using X-13ARIMA-SEATs. For this procedure, we first perform automatic outlier detection for each series independently, including each of the identified outliers from X-13 as regressors in our model. The model fits were very good, as shown in Figure 2.

The models performed less well during the COVID-19 pandemic; however, including outlier treatments for the pandemic-induced lull significantly improved performance. Taken together, for relatively large arrival levels with higher frequency data, the time series approach is conceptually well-suited to migration trends and seasonality and produces well-fitting estimates using our proposed model. While the time series approach appears to work well for larger inflows, future research will examine the models' performance with smaller counts and consider whether additional distributional assumptions are methodologies would be more appropriate.

6. Discussion and Future Work

In this analysis, we introduce Airline Passenger Traffic (APT) data as a predictor to improve time series model-based forecasts of nonimmigrant admissions to the U.S. We proposed a bivariate structural time series model that models each series jointly and incorporates each series level and slope, seasonality, and outlier effects. We find that the bivariate model fit the I-94 arrivals data well—both at the national and state levels. Including outlier pre-processing further enables the models to flexibly respond to large-scale shocks to migration, as seen with travel restrictions implemented during the COVID-19 Pandemic. A few avenues of additional work remain on the time series approach. For one, our research focused on states that traditionally have larger immigrant populations and, hence, draw larger immigrant inflows. Where nonimmigrant inflows are smaller or less frequent, other demographic approaches may be more appropriate. Second, we have yet to consider adding demographic detail, namely sex and age, to our estimates. Future research may wish to examine the best ways to obtain demographic characteristics of the new arrivals and estimate their expected length of stay. Finally, a blended approach of time series modeling for large immigrant flows and demographic methods for smaller flows may be warranted. As more administrative data on immigrant flows become available to the public, we suggest that researchers conceptualize a broad approach to incorporating these data into both statistical and demographic methods for measuring migration.

References

Bryan Baker (2025). Population Estimates for Nonimmigrants Residing in the United States: Fiscal Years 2019 to 2024. Technical report, Office of Homeland Security Statistics; Office of Strategy, Policy, and Plans; U.S. Department of Homeland Security.

De Jong, P. (1989). Smoothing and Interpolation with the State-Space Model. *Journal of the American Statistical Association*, 84(408):1085–1088.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press.

Gross, M., Lamas, J., Mayol-Garcia, Y. H., and Jensen, E. (2024). Census Bureau Improves Methodology to Better Estimate Increase in Net International Migration. Section: Government.

Jensen, E., Gross, M., and Johnson, S. (2024). Crossing Borders: Exploring the Federal Governments Estimates and Projections of International Migration. Working Paper No. 107, U.S. Census Bureau, Washington, DC: U.S. Department of Commerce.

Rukh-Kamaa, A. (2024). U.S. Nonimmigrant Admissions 2023 Annual Flow Report. Technical report, Office of Homeland Security Statistics, Washington, DC: U.S. Department of Homeland Security.

Tiller, R. (2006). Model-Based Labor Force estimates for Sub-National Areas with large Survey Errors. Technical report, U.S. Bureau of Labor Statistics, Washington, D.C.

United Nations (2022). Handbook on Measuring International Migration through Population Censuses. Studies in Methods No. 115, United Nations Statistical Division, New York, NY.

U.S. Census Bureau (2024). Methodology for the United States Population Estimates: Vintage 2024. Technical report, U.S. Census Bureau, Washington, DC: U.S. Department of Commerce.

Yepez, C. and Leimgruber, W. (2024). The evolving landscape of tourism, travel, and global trade since the Covid-19 pandemic. *Research in Globalization*, 8:100207.