



# Correcting the record about post-2020 entrances to California

**NATALIE HOLMES AND EVAN WHITE**

In [March](#) and [December](#) of 2021, we published two reports about domestic migration in California. These reports discussed who was moving to California and who was leaving, and looked at regional patterns across the state.

Unfortunately, we have since discovered that some of those estimates were inaccurate due to an error in our calculations. The effect of this error is that we substantially underestimated domestic entrances to California in 2020 and 2021. The actual decline in entrances to California from other states was far less than we originally reported. Our error affects only entrances, and does not impact the estimates of the number of people leaving California or moving within California. This document describes the error, and the efforts we undertook to correct it. We have re-released a corrected version of the December 2021 report, which used data through Q3 2021.

We take full responsibility for this mistake. As we have noted in both publications, the implications of domestic migration are important. We deeply regret our dissemination of incorrect estimates and have implemented additional steps with our review process to prevent mistakes like this from happening again.

## More detail about the error

Our mobility reports use the [University of California Consumer Credit Panel \(UC-CCP\)](#), a quarterly panel (2004 to present) of anonymized credit report data from one of the three nationwide credit bureaus. We define quarters as the end of March, June, September, and December. We receive two extracts in the UC-CCP: one is a 2% national sample, and one focuses on California. For the December 2021 publication, we used only the California extract. For the March 2021 publication, we primarily used the national extract, and the error affects only estimates using the California extract.

When we first created the UC-CCP California extract, the credit bureau provided CPL with quarterly data on individuals that lived in California at any point from 2004 through 2019, even during the quarters in which they resided in another state. However, starting in Q1 2020, while we continued to receive information for new entrants to California, we stopped pulling data for those individuals back to 2004. In 2020, we conducted two pulls of monthly data: one in October that pulled data back to January 2020 for anyone who had arrived in the state since then, and one in December that pulled data from November and December 2020 but did not go back further. Starting in Q1 (March) 2021, we no longer pull historical data for new entrants to California.

A consequence of these different methods of pulling the data is that for anyone residing in California 2004 to 2019, we are provided with their location data even before and after they live in California. But for some people who began appearing in the California extract after Q4 2019, we do not go back to prior quarters to pull data on their prior location. This is important because it affects our ability to distinguish people who are new arrivals in California from those who have been living in California all along but have not previously appeared in the credit bureau data.

To date, our mobility reports have counted somebody as a “mover” to California only when we were able to observe them in two consecutive quarters, where their state code in the first quarter is outside California, and their state code in the next quarter is inside California. In this way, we do not inadvertently count the numerous first-time credit holders – appearing in the data for the first time – as movers. This approach works well for the 2% national extract. It worked well for the California extract prior to the 2020 changes in how we refreshed the data, because prior to 2020 we have the location of each individual in each quarter they are in the data. This approach also worked well for people in our dataset who lived in California after 2004 that later left the state and

subsequently returned, because we receive data on these individuals after they leave California, and thus were able to observe their intermediate location before they moved back to California. It also worked well for some new entrants to California in 2020, because we pulled prior locations during some months of that year.

However, this approach incorrectly excluded many people who never lived in California during the 2004-19 period, and who moved to California after Q4 2019. For most of those people, the first observation we saw was their first quarter in California, and because we did not observe their location in the prior quarter, we did not count them as a mover. Although we were aware of the complexity of the data design, we did not appreciate its impact on our estimates of moves into California in 2020 and 2021. As a result, we underestimated the number of people who moved into California after Q4 2019.

### **How we fixed the error**

We began our correction by using the national extract, which was unaffected by the sampling issue mentioned above. For anyone who appears in the national extract in any quarter, including 2020 and 2021, we see all available quarterly credit records since 2004. This enables us to clearly distinguish California entrants from first-time credit holders in the national extract. Though that extract is only 2% of credit holders, it gave us a good idea of the magnitude of our error and also confirmed that our estimates of exits, within-California moves, and move destinations were not affected. However, the 2% national extract does not have sufficient sample size to support region- and county-level estimates, so we could not rely on it alone.

Fixing our estimates of the number of entrances to California in the California extract was more involved. For periods in which we did not receive information on prior location, we needed to classify new arrivals in our sample either as movers to California or as new entrants to the credit system. Luckily, some characteristics of those two groups are starkly different. For example, movers usually had a longer credit history than new entrants to the credit system, and the origination dates on their accounts were usually further in the past. Age and credit score were also different between the two groups, because new credit-system entrants are often young and have thin credit files that result in lower credit scores.

For quarters in which we did not observe the prior locations of new individuals, our solution was to classify such people as a mover or as a new entrant to the credit system using a logit prediction model. We developed the model using the 2020-21 national data, where we could see if a person was either a mover or a new entrant to the credit system. We used variables that proved especially predictive: age, credit score, number of both total and open tradelines, oldest account (in years) among both all and open tradelines, and longest grid of payment history for both all and open tradelines. We also used indicators based on those variables: >1 tradeline, >1 open tradeline, >3 months of payment history across all tradelines, >3 months of payment history across open tradelines, >1 year account age across all tradelines, >1 year account age across open tradelines, each quarter-year, and age squared. Finally, because approximately one-third of individuals did not have any open tradelines in their first post-2020 appearance in California, we included an indicator for >0 open tradelines, as well as its interaction with the six total tradeline-based variables.

To get a sense of how well our model performed, we checked it against the national sample, where we definitively know if a person was a mover or a person appearing in the credit system for the first time. We found that the model correctly classified people as a mover or a new entrant 95% of the time. We then applied the model to data in the California extract from Q1 2020, Q4 2020, and Q1-Q4 2021, the quarters affected by the change in definitions (and we will continue to apply it in the future as new quarters of data arrive). We weighted observations in these data based on their predicted probabilities of being movers, and tabulated results using those weights. We can assess the performance of our method by examining Q2 and Q3 of 2020, for which we can distinguish movers from new credit files in the California data. The number of California arrivals indicated by our method differs from that indicated by counting movers directly by just 1.2% in Q2 2020 and 1.7% in Q3 2020.

## The corrections

On April 14, 2022, we released an updated version of the [December 2021 report](#). Many of the key results regarding moves into California (entrances) are significantly different from what we initially reported. A summary of the major changes is described in the table below.

TABLE 1: [Corrections of Key Entrance Measurements](#)

| <b>ESTIMATE<br/>(Q1 2020 TO Q3 2021)</b>  | <b>INCORRECT<br/>ESTIMATE<br/>IN DEC. 2021 BRIEF</b>           | <b>CORRECTED ESTIMATE<br/>IN UPDATED<br/>APRIL 2022 BRIEF</b>  |
|---|--|--|
| Moves into California, from other states.   | 38% Decrease   | 8% Decrease  |
| Number of entrances, Q3, 2021   | 104,722  | 172,543  |
| Economic regions experiencing declines in entrances   | All regions saw declines, ranging from -45% to -26%            | Regions saw changes ranging from -17% to +1%                   |
| Number of California counties experiencing a decline in entrances from other states   | All California counties saw declines                           | 40 out of 58 counties saw declines                             |
| Decline in out-of-state entrances to San Francisco, Santa Clara, and San Mateo Counties   | San Francisco (-53%)<br>Santa Clara (-52%)<br>San Mateo (-48%) | San Francisco (-27%)<br>Santa Clara (-20%)<br>San Mateo (-19%) |
| State-wide net exits in Q1 2020<br>(Net exits are defined as the number of entrances from other states minus exits to other states) | 60,000   | 41,000   |
| State-wide net exits in Q3 2021   | 150,000  | 83,000   |

This chart shows the differences between the incorrect estimates (on the left) and the corrected versions (on the right).

FIGURE 1: Entrances to and exits from California (incorrect)

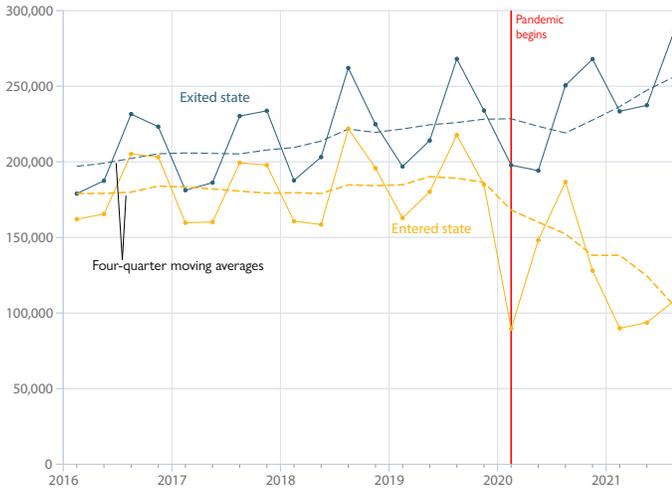
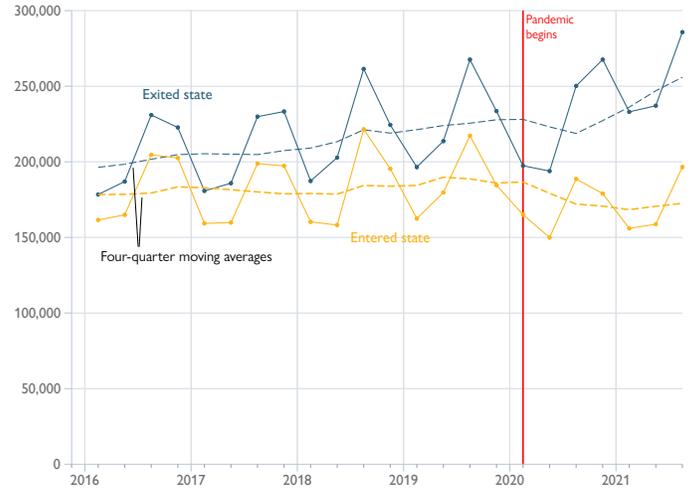


FIGURE 1: Entrances to and exits from California (corrected)



For a full accounting of the corrections, please refer to the corrected [December report](#). We have archived the previous versions of these reports here: [December 2021](#), [March 2021](#).

### Questions?

For any questions about this correction, please email the authors at [evanbwhite@berkeley.edu](mailto:evanbwhite@berkeley.edu) and [nholmes@berkeley.edu](mailto:nholmes@berkeley.edu), or the CPL Communications Director at [press@capolicylab.org](mailto:press@capolicylab.org).