

External validity is determined at the county-year level by comparing internal data with statistics gathered from external sources. Our process consists generally of four criteria, outlined below.

Filing Coverage

We first directly compare the caseload recorded in our data with an external source for any given county-year. The variable *cov_ln* is the percentage of courts filings covered in our data. The threshold we use is:

$$86\% \geq cov_ln \leq 114\%$$

In other words, the filings in our data cover at least 86% of the filings in the courts data but no more than 114% (over-coverage).

In areas that have overall lower case counts, we apply an additional threshold based on the difference in the number of filings (*dif_ln*) in the courts data and our data:

$$-10 \geq dif_ln \leq 10 \ \& \ f_ln > 0$$

For example, if there are 4 courts filings and 3 internal filings, our data covers only 75% of the courts data. But practically, this is roughly representative of the case volume in this county-year. Only county-years with relatively low filing volumes will be marked in this scenario. This marking excludes county-years with no data (*f_ln* == 0). We deal with zero counts separately.

Extrapolated Comparison

For counties in which courts data are only available in some years, we extrapolate coverage comparisons. This is done through the following steps:

1. We created a new variable that holds the available *ext_comp* values.
2. We replaced missing values of *ext_comp* with the value from the most recent year observed within the same county. For example, if *ext_comp* was observed in 2015 but not 2016, replace the new variable with the 2015 value for 2016.
3. Mark the county-years in which courts data are unavailable, but our data on filings meet the threshold of the extrapolated courts filings. We only mark county-years in which courts data are unavailable and no source for external comparison exists.

Zero Counts

There are county-years in which there were no eviction case filings. It is likely that a small percentage of the county-years with 0 case filings in our data are true zeros (i.e. no eviction cases were filed). Verifiable true zeros can be marked when both $ext_comp == 0$ and $f_ln == 0$.

We marked plausible zero filing counts with the follow steps:

1. Within each county, we counted the number of years with $ext_comp == 0$ (no filings reported in the courts data).
2. Within each county, we counted the number of years with $ext_comp != .$ (non-missing courts filings).
3. We divided the number of years with $ext_comp == 0$ by the number of years with $ext_comp != .$. This is the proportion of county-years with true zero filing counts.
4. In counties in which at least 50% of the reported courts filings were 0, we mark the filing counts in our data without available courts data as reliable.

Case Outcome Quality

There are areas in which we have good filing coverage, but limited information on how cases were resolved. We use the percentages of unresolved cases ($pbns$) and dismissed cases ($pcds$) to determine if case outcomes in a given county-year is reliable:

$$pbns + pcds \leq 60$$

Combined rates of dismissed and truly unresolved cases over 60% is atypical and likely indicates that we do not have adequate information on outcomes, setting aside variation in the likelihood of dismissed cases across states.