Online Appendix: The Importance of Immigrant Workers for American Homebuilding*

Troup Howard† Mengqi Wang‡ Dayin Zhang§

February 2024

*Acknowledgments. All remaining errors are our own.
†University of Utah. Email: troup.howard@eccles.utah.edu.
‡University of Wisconsin-Madison. Email: mengqi.wang@wisc.edu.
§University of Wisconsin-Madison. Email: dayin.zhang@wisc.edu.

1 Sample Construction: CoreLogic Microdata

All quantity results are based on 4.22M records of new construction transactions, representing the near-universe of residential construction between 2005 and 2012. Most price results are based on a subset of 2.50M records that contain fully populated attribute data, so that we can adequately adjust for (observable) quality. This section details each step of the data build.

We assemble data on all new residential construction completed in local markets from CoreLogic deeds data. We start with all transaction records flagged as new construction and with a built year between 2005 and 2012. Notably, this includes records with a sales year through 2022. We also permit the recorded sales year to be one year before the built year, to include preconstruction sales. We remove nonresidential properties by restricting to property types of single-family homes, condos, and duplexes. We also remove approximately 100k records that despite being flagged as new construction have a prior transaction associated with the same property PIN. Finally, we keep only properties which we can match to a county launch date for Secure Communities. Practically, this means excluding US Territories and nine small counties with irregular governance structures. In total, this yields 4.22M records. This is the core dataset for all quantity results.

For tests of transaction prices, we impose four further filters. First, to ensure an accurate signal about market value, we restrict attention to arms-length transactions, which removes 107k observations. We also remove 50k observations with a second transaction occurring within 60 days of the focal transaction, to avoid inadvertently including partial-interest transactions. Second, we remove 63k transactions with sales price greater than \$5M or less than \$10,000. Third, we remove 323k properties which sell more than 10 years after their built year for two reasons: (i) at this age, it's no longer clear that such a transaction accurately reflects "new construction" pricing, and (ii) as such delays are unusual, we are concerned about large property-specific unobservables. This leaves 3.68M observations.

Our fourth filter concerns the availability of hedonic attributes. We use these attributes to provide both direct and indirect tests of endogenous shifts in building characteristics. As usual in real estate microdata, there is only partial coverage of structural characteristics. We focus on standard hedonic attributes in the real estate literature: square footage, age, bedrooms, bathrooms, and census tract. Of these, square footage is the most consequential because we also use this when creating quantity measures. Property attributes are contained in CoreLogic tax assessor files; each observation in the deeds data contains a unique PIN that permits straightforward linking to the assessor data. Because it is quite common for assessors to receive building information at a meaningful lag, we allow a look-ahead period of five years. That is: for a property transacting in year t, we take the first set of fully populated hedonic controls in any year between t and $t+5$. After this, for records missing attribute data, we match to MLS records and augment them with listing characteristics. This allows us to recover attributes for another 78k records.

In total, within the subset of 3.68M observations that pass our first three price filers, we have a full set of hedonic attributes for 2.50M records (68%), have at least square footage (but are missing some other attributes) for 866k (24%), and are missing all or most attributes for 312k (8%). To hold the sample constant between regressions that do and do not use quality controls, we use the set of 2.50M observations as our core dataset for price regressions. In addition, within the set of 866k properties that have (only) square footage measurements, we can show that the quality adjustments evident in our core price-sample of 2.50M do also appear to be present in this ancillary set of 866k.

2 Empirical Specification: Population and Workforce Normalizations

The purpose of this section is to expand on several considerations relating to specification and weights, especially with respect to population and workforce regressions. We believe there are multiple choices which would be reasonable from an ex-ante standpoint. Most, but not all, of these choices have little impact on our results. This discussion unpacks why we consider the choices in the paper to be optimal and highlights important sensitivities.

At its core, our paper exploits the setting of an immigration shock to explore what happens as a consequence of a plausibly exogenous reduction to regional construction workforces. Accordingly, we begin by empirically validating that Secure Communities – the shift in immigration enforcement – does have the anticipated effect at the county level. We explore the effect on group-level populations, and on group-level workforces.

The first issue concerns the choice normalization for the dependent variable. As discussed in the paper, there are two natural choices. We can estimate impact in group-level population shares (simple share), or in log population share.[1] The former implies an additive impact, and the latter implies a proportional impact. The reason this becomes consequential is heterogeneity in the baseline share of undocumented residents. ACS-reported county-level LEFB share in our data ranges from 1% to 30%. In one of these counties approaching 30%,

[1] Because we use constant 2005 population figures for the denominator, using log population shares is mathematically equivalent to using log population levels – the denominator is absorbed in the county fixed effect. For clarity of exposition, we frame the discussion in terms of simple share versus log share.

often counties bordering Mexico, an increase in immigration enforcement might shift undocumented share by 1 or 2pp; but this might represent deporting nearly every undocumented resident in an area with a much lower baseline. For this reason, we tend to mildly preference a representation where immigration enforcement has proportional impact. However, it is not clear that this must be the case. LEFB share correlates very highly with population ($\rho = .94$). Because enforcement operates through interaction with law enforcement, a careful model of additive versus proportional impact would require thinking about crime and arrest propensity as a function of population. In turn, this raises complicated issues of policing, community demographics, and even potentially local compliance with ICE detainers that are all far beyond the scope of this paper. So, we do not take a firm stand that the true model is necessarily proportional.

As a follow-on, one main purpose of estimating population-level impact is to demonstrate that the effects captured by our empirical strategy are reasonable when compared to known administrative statistics on Secure Communities. As we discuss in the paper at length, our estimates will catch adjustments on several margins beyond direct deportation, which is both reasonable and desirable – but if our empirical design has us estimating an impact of, for instance, 10 million people while SC official data reflect an impact of a quarter million, this discrepancy would be a meaningful warning signal about our use of the setting. If the true model lies somewhere between purely additive and purely proportional, then our regression estimates – as usual – will be a convex combination of heterogenous treatment effects. To effectively compare our in-sample statistical estimates with the population parameters, therefore, it is essential to weight regressions with county-level population.

When we estimate the impact on workforces, we make two changes in specification. First, we estimate in simple shares rather than logs. The largest factor motivating the use of log shares in the population regressions is the wide heterogenity in baseline. The distribution of LEFB construction share (and of all other groupings conditioned on occupation) is much compressed, simply because we're looking at a subset. LEFB construction workers in all county-years are less than 5% share of total population and in the vast majority are less than 1.5%. This reduction in variance will also reduce the potential difference between an additive and a proportional model. In addition, the use of simple shares allows the estimation to handle zeros. There are an appreciable number across all county-years. A log specification would drop counties with any single zero or rely on some other choice about handling zeros. We view all of these choices to be suboptimal relative to estimating in simple shares. This does mean, however, that our estimation places weight on zeros rather than, for instance, viewing zeros as a consequence of ACS sampling variation. We have tried filling single zeros in the time series with a midpoint imputation and find this does not meaningfully affect estimates, suggesting low sensitivity to such occurrences.

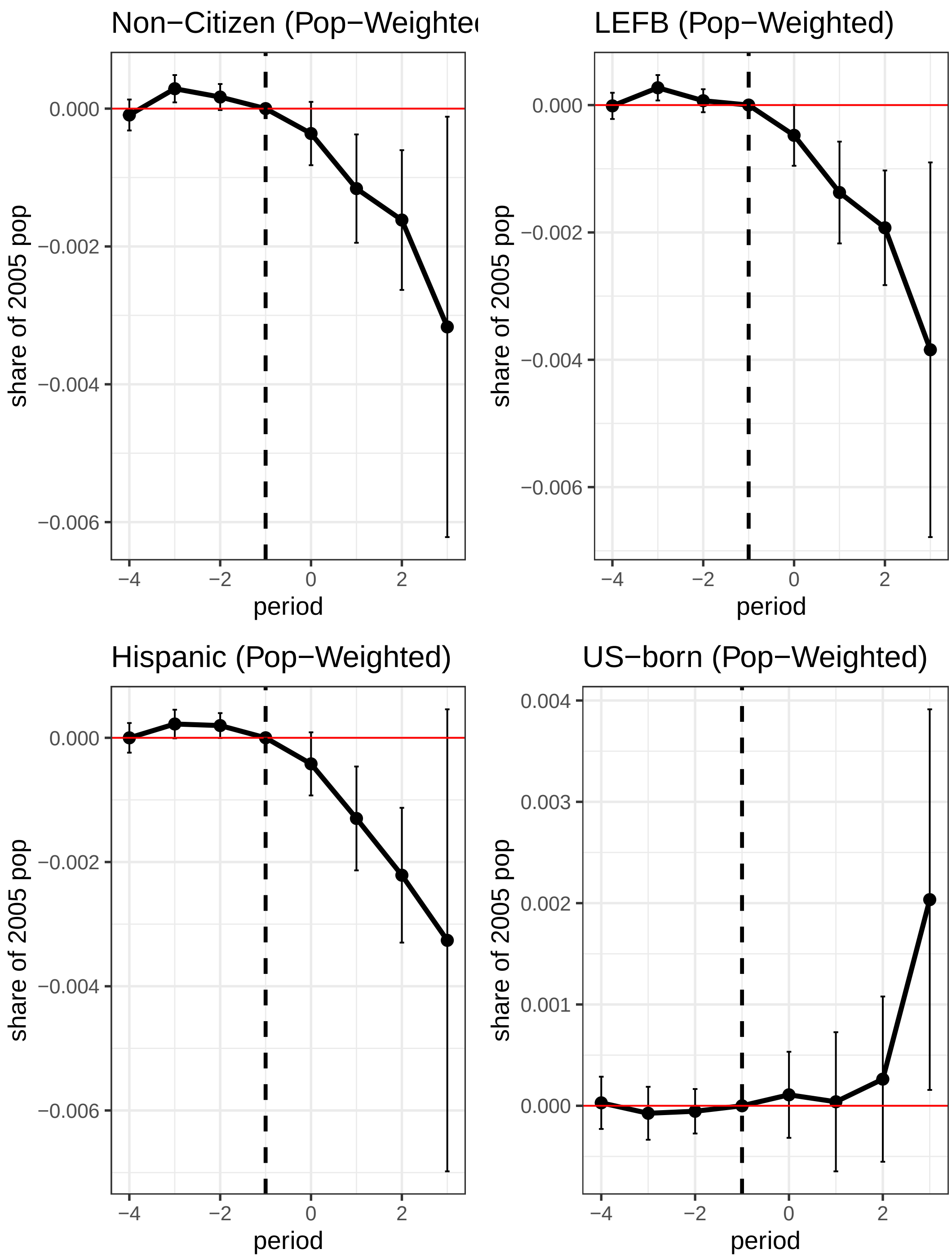
The second salient choice concerns the use of weights. While weights are essential in the population estimates for ensuring that we can meaningfully compare magnitudes between our coefficients and external data, our analysis of workforce impact is fundamentally concerned with regional housing markets as the random unit. That is, the causal response we are interested in occurs within a county-level housing market. We want to document that construction workforces are shocked by SC, and then explore what happens thereafter to regional homebuilding and home prices. Since the random unit is the local housing market, the use of population weights is actively undesirable. We do not want to place a large amount of additional weight on how SC impacts LA county. We want to estimate the average impact of a negative labor supply shock on US counties, which is the statistic from an unweighted regression. As Solon et al. (2015) note, discrepancy between weighted and unweighted regression coefficients denotes causal heterogenity with respect to the weights. In this setting, that's heterogeneity of with respect to population, which is interesting but not the primary focus.

Practically speaking, the use of weights makes little difference with respect to measuring SC's impact on the undocumented population. There is one consequential difference, however, which is the impact on US born workforce. Figure 1 of this appendix repeats the estimations of Figure 4 in our paper, but weights observations with 2005 county population. We find very similar impact for groupings of noncitizens, LEFB, and Hispanic workers. Reductions are slightly smaller than the unweighted versions. This suggests that immigration enforcement has different impact by population and in turn that the sensitivity of homebuilding to labor supply may vary between small and large regions–neither of which is surprising. The bottom right graph in Figure 1 shows the response of US born workers. In contrast to Figure 4 in the paper, the weighted regressions show no change in US-born construction workers until the last period, where we find an increase instead of a decrease. Paired with Figure 4, this suggests that domestic labor may replace immigrant labor in construction more readily in highly populated areas, and/or that complementarities between domestic and immigrant labor are stronger in more rural areas and weaker in urban areas.

References

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human resources*, 50(2):301–316.

Figure 1: SC Workforce Impact: With Population Weights



Note: Note.