Joint Replication of Hagedorn, Karahan, Manovskii, & Mitman

Joint Online Appendix for:

Unemployment Insurance Generosity and Aggregate Employment

Christopher Boone, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan

Revisiting the Effects of Unemployment Insurance Extensions on Unemployment: A Measurement-Error-Corrected Regression Discontinuity Approach

Steven Dieterle, Otávio Bartalotti, and Quentin Brummet

D Online Appendix D: Joint Replication of HKMM

In this joint appendix, we describe our replication of the main estimate in Hagedorn, Karahan, Manovskii and Mitman (2015), hereafter HKMM. We face two challenges in the replication: (1.) properly constructing the data set and (2.) executing the estimation correctly. Though proper execution of the estimation is potentially challenging due to the non-linearity of the model, replicating the data set is ultimately more difficult. While we are unable to simultaneously reproduce the exact sample size and point estimate, our preferred replication is very close: we come within 0.002 of the estimate and our data set contains about 1% fewer observations (385 out of 37,177). In this description of our replication of HKMM, we discuss not only our final specification but also the choices we made and the reasons for those choices. In particular, we traded off sample size with estimate closeness. We considered specifications with closer estimates where the gap in the sample size compared to HKMM's sample was substantially larger; we also considered samples which matched more closely the sample size of HKMM but where the estimates were not as close.

Estimation Equation and Method

The HKMM estimation equation, which uses data at the pair-by-time level, is as follows. Prior to estimation, the data for a given pair p at time t is spatially differenced.

$$ln(u_{pt}) - \beta(1 - s_t)ln(u_{pt+1}) = \alpha * ln(D_{pt}) + \lambda_p' F_t + \epsilon_{pt}$$

In this expression,u_{pt} is the unemployment rate from LAUS (as calculated prior to the March 2015 redesign of the LAUS program), β is the discount factor equal to 0.99, s_t is the separation rate, andD_{pt} represents weeks of UI benefits available. $\lambda_p' F_t$ represent the interactive effects: F_t is a time-specific vector of length K of common factors, while λ_p (also of length K) represents the pair-specific factor loadings. HKMM determine, by minimizing an Akaike information criterion, that the optimal K is equal to 2. We replicate their minimization, also obtaining two factors as optimal. All of our estimates estimate with two factors in both space and time. We follow HKMM in estimating the model using the method of Bai (2009). In the April 2016 version of HKMM, the authors report a main estimate of 0.049.

Sample

The biggest challenge in replicating this result is determining precisely which pair-time observations were used in the sample. HKMM report using an unbalanced panel of quarterly LAUS unemployment data spanning 32 quarters from 2005q1 to 2012q4 with a sample size of 37,177 county-pair-by-quarter observations in their baseline regression. Dividing the number of observations by the number of quarters indicates that

this sample size is similar to a balanced panel of 1,162 county-pairs (37,177/32=1,161.78). Our initial sample of pre-revision LAUS data yields an unbalanced panel for 1,171 county-pairs and a total number of observations of 37,464. This is a nearly balanced panel. It only drops data for the four quarters following Hurricane Katrina (2005q3-2006q2) for the two border pairs that include St. Tammany, LA (paired with Hancock, MS and Pearl River, MS). The missing counties for these quarters range from small to above mean county size. In 2005, these three counties had populations of 217,407, 46,097, and 51,764 respectively, according to the U.S. Census Bureau. Dropping these counties fully to create a balanced panel is essentially inconsequential to our estimates. The estimate in the balanced panel is 0.0527 and the estimate in the unbalanced panel is 0.0529.[73]

Though estimating on the above unbalanced sample yields estimates which are close to those reported in HKMM, in order to more closely match HKMM's reported number of observations, we consider an additional sample restriction to the unbalanced panel. In particular, we note that HKMM draw employment data for auxiliary specifications from the Quarterly Workforce Indicators (QWI). In an earlier draft from October 2013, HKMM report a sample size of 30,988 county-pair-by-quarter observations covering the period from 2005q1 to the beginning of 2012. Over the 28 quarters covered, this sample size would be consistent with a nearly balanced panel of 1,107 county pairs (30,988/28=1,106.71), which aligns with the number of pairs for which HKMM report having "complete data" in the October 2013 and April 2016 drafts. We believe that the phrase "complete data" likely refers to the presence of unemployment data in the LAUS and employment data in the QWI in a given quarter.

So, we consider the possibility that the choice to use the panel of "complete data" in the October 2013 draft may have carried over in some form to the sample used in the April 2016 draft. Specifically, since QWI data only cover Massachusetts beginning in 2010q1,[74] we exclude county pairs that include a Massachusetts county from the sample to generate an unbalanced panel of 1,150 county pairs. This sample restriction leads to a sample size of 36,792 county-pair-by-quarter observations. While this restriction leads to a smaller sample than reported in the April 2016 version of HKMM, it is simpler and yields an estimate closer to HKMM than other QWI-based sample restrictions we considered. In particular, using this sample, we find an estimate of 0.0510 (to four deminal places), compared to the 0.049 (to three decimal places) reported by HKMM.[75]We use this sample for the replication estimate used both by Dieterle et al. (2018) and Boone et al. (2018): an unbalanced panel which (1.) keeps counties which temporarily did not report in the aftermath

[73]See https://www.bls.gov/katrina/lausquestions.htm for a discussion of the impact of Hurricane Katrina on LAUS.

[74]The QWI includes both beginning-of-quarter and end-of-quarter statistics. Since the beginning-of-quarter statistics are rolled over from the previous quarter's end-of-quarter numbers, some of the QWI data for Massachusetts does not begin until 2010q2.

[75]We estimate this model using the user-written Stata command "regife" (Gomez 2015). We have also written our own simplified version of this command and are able to obtain identical estimates.

of Hurricane Katrina and (2.) drops all counties in Massachussetts.

Possible Reasons for Remaining Discrepancy

Lastly, we note that there may be other minor specification choices that prevent us from replicating the results of HKMM. First, we obtain the dependent variable (unemployment rates as estimated by LAUS prior to the March 2015 redesign) through a FRED API. While the original LAUS dataset (which HKMM presumably used) includes the estimates of the raw counts of unemployed persons and the size of the labor force, the FRED API reports only the unemployment rate to the nearest tenth of a percentage point. Thus, HKMM may have been using unrounded unemployment rates while we are using rounded unemployment rates. Second, there may be differences in how we aggregate weeks of benefits from the weekly level (at which they are reported) to quarters. We calculate the weeks of benefits available on a given calendar day, and then aggregate to the month level. We then aggregate to the quarterly level using an unweighted average of the three months within the quarter. It is possible that HKMM performed this aggregation somewhat differently. Third, it is possible that we use a different separation rate than HKMM. We use the non-seasonally-adjusted total separation rate as reported by JOLTS. Other possibilities include the seasonally-adjusted version or the version which includes only private employment. In any case, while these uncertainties might prevent us from replicating HKMM's result exactly, the fact that our replication is within 0.002 (to three significant digits) of HKMM's estimate suggests that these minor differences do not matter qualitatively. [76]

[76] Since the Bai (2009) estimator is non-linear, an additional possibility is that the likelihood function used in the optimization has multiple local optima and that HKMM and our replication of HKMM are at different optima. We do not, however, think this is likely given (1.) that we are able to exactly replicate the optimality of two factors and (2.) that our estimates are so close to those of HKMM.