Response to Correia, Luck and Verner

Andrew Lilley, Matthew Lilley, Gianluca Rinaldi

May 18, 2020

We thank Correia, Luck and Verner for their response (hereafter, CLV2). We provide a brief reply.

1 Population Estimates

A. Both manufacturing employment and population estimates are affected by redistricting.

We first correct a misunderstanding in CLV2 surrounding the 1917 population estimates. CLV2 states *"the 1917 value is purely based on a linear extrapolation from the 1900 and 1910 census."* This is not correct. The 1917 estimates also take into account city incorporations, and therefore can better explain employment growth from 1914 to 1919. As we explained in our original comment:

> "Our city level population data for 1910 and 1917 comes from a Census Bulletin published in June 1918 [7]. As explained on page 5 and following of this document, the 1917 estimates are computed by linearly extrapolating the growth at the city level recorded from 1900 to 1910, while 1910 estimates are based on Census enumerations conducted that year. Crucially, the 1917 estimates are also adjusted by the Census for incorporations of cities in that period - the redefinition of city borders. This adjustment is very important as some cities in our sample increased both in population and manufacturing capacity substantially due to the incorporation of neighboring cities and towns. For instance South Omaha was annexed into the city of Omaha in 1915. The Bulletin estimate therefore takes into account the 1910 populations of South Omaha and Omaha to obtain estimates of 1917 population but keeps the 1910 population as only the Omaha estimate. This ensures calculated population growth from 1910 to 1917 is consistent with the reporting of manufacturing statistics which are not backward adjusted for incorporations."

As an example, see Table 50 at page 231 of the 1920 *Census of Manufactures*:

CITY.	Census year.	Number of establishments.	Wage earners (average number).
Omaha	1919	561	21,304
	1914	417	8,922
	1909	432	8,023
South Omaha[1]	1914	61	6,063
	1909	71	6,306

[1] Annexed to Omaha in 1915.

Figure 1: 1920 edition of the *Census of Manufactures*. Omaha's manufacturing employment increases by 140% from 1914 to 1919, which is mostly due to its absorbing South Omaha in 1915.

Consider the example of Omaha and South Omaha. In the 1920 *Census of Manufactures*, Omaha and South Omaha respectively have employment of 8023 and 6306 for 1909, and 8922 and 6063 for 1914. In 1919, post-annexation, Omaha has employment of 21304. The 1924 *Statistical Abstract* lists identical values for both 1914 and 1919. The 1910 *Census of Manufactures*, published before Omaha incorporated South Omaha in 1915, lists identical values (8023 and 6306) for 1909. Clearly, the 1909 value in the 1920 *Census of Manufactures* was not revised for the subsequent annexation, contrary to the claim in Footnote 5 of CLV2.[1]

B. The population growth estimate from 1910 to 1917 can be separated into a linear extrapolation component, and a component due to redistricting. Both predict 1914-1919 manufacturing employment growth. Both are also spuriously correlated with NPIs. Failing to control for either incorrectly attributes their effects to NPIs.

The population growth estimate from 1910 to 1917 can be separated into two components - growth due to linearly extrapolating growth observed between the 1900 and 1910 censuses, and growth due to redistricting.[2]

[1] More generally, CLV2 claim that that the 1909 data in the 1920 *Census of Manufactures* is backwardly revised whenever annexations occur. Unfortunately, this is not a rule. In addition to Omaha, the other cities with NPI data and large incorporations between 1914 and 1919 - Los Angeles, Portland, Richmond, and Toledo - have identical 1909 manufacturing data in the 1910 and 1920 *Censuses*. We use the 1909 value from the 1910 *Census of Manufactures* to obtain the most consistently defined and historically accurate series available.

[2] Due to data limitations, particularly the issue that manufacturing employment estimates are intended to be an annual average of the fiscal year of each business within a city, which differs by business and by city, we cannot always know whether these redistrictings affect employment growth between 1909-1914 (i.e. the pre-trend period in CLV1) or 1914-1919 (the main outcome variable of CLV1). In rare cases, such as the Omaha's incorporation for 1915 outlined above, it is made clear in the *Census of Manufactures*. We can confirm that for the five cities which grew by more than 5% overnight due to incorporations - Omaha, Los Angeles, Portland, Cincinatti, and Richmond - all but Cincinatti had their largest incorporations after 1914.

The 1900-10 extrapolation component predicts manufacturing employment growth due to two possible channels, firstly by predicting actual population growth which then produces contemporaneous employment growth, and secondly through lagged effects of population growth on employment growth. The component due to annexations affects manufacturing employment growth mechanically.

To make this point, we decompose the estimated population increase $\Delta Pop_{1910}^{1917} = Pop^{1917} - Pop^{1910}$ for each city into these two components, in percentage growth terms: $\Delta Pop_{1910}^{1917} = \Delta Pop_{1910}^{1917,Extrap} + \Delta Pop_{1910}^{1917,Incorp}$. We then regress log manufacturing employment growth from 1914-1919 on log total population growth, and the logs of both components separately. Both components explain manufacturing growth. Moreover, NPIs are spuriously correlated with both, so failing to include these as controls will attribute these effects to NPIs.

Table 1: Regressions With 1910-1917 Population Growth and Its Components

	Log Manu Employment Growth (1914-1919)		Days Of NPIs	
Total population growth	1.8***		235.2***	
	(0.4)		(45.8)	
Extrapolated population growth		1.6***		199.4***
		(0.5)		(67.1)
Growth due to incorporations		1.8**		275.8***
		(0.7)		(75.7)
R^2	0.58	0.58	0.31	0.32
N	43	43	43	43

The dependent variable of the left two columns is the log growth of manufacturing employment at the city level from 1914-1919. The dependent variable of the right two columns is the number of days that NPIs were in place. All explanatory variables are estimates in log growth terms. Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

C. 1910 to 1920 population growth is an inappropriate estimate of population growth since it is affected by the pandemic.

CLV2 respond that our population estimate does not match an alternative based on interpolating between the 1910 and 1920 censuses. They then acknowledge that this alternative is problematic because it is endogenous to the pandemic. We agree. In particular, changes in migration after the pandemic (potentially caused by its economic effects) invalidates the interpolation estimate and the following analysis.

As an indicative exercise (though this does not resolve endogeneity concerns), a regression of $\Delta \ln Pop_{1910}^{1920} = \beta_0 + \beta_1 \Delta \ln Pop_{1900}^{1910} + \beta_2 Mortality_i^{1918} + \varepsilon_i$ for cities with NPI data shows that an increase in influenza-like mortality in 1918 of one death per hundred persons (approximately the range of mortality rates in 1918) is associated with a reduction in city-level population of 19%. The 1920 population estimates can

be substantially affected by the treatment variables of interest.

2 Pre trends

D. CLV2 use prior population growth as a control in the DiD regression. This is not a valid fix when the parallel trends assumption is violated.

As we showed above and in our comment, employment growth in this period is strongly driven by population growth. CLV claim that controlling for historical population growth during the pre-treatment period alleviates pre-trends concerns. However, when the parallel trends assumption is violated, this type of control is insufficient. Note that even in the presence of strong long-term trends, population growth covering any specific period contains random idiosyncratic variation. This idiosyncratic variation means pre-treatment population growth is a good match for variation in pre-treatment employment and output growth, but it is less able to explain counterfactual future growth. This time varying attenuation bias will then absorb the pre-trend, while only attenuating the post-treatment estimates. This problem is larger when pre-treatment population growth is more affected by idiosyncratic variation rather than long-term trends.

E. Unit-specific time trends are the preferred correction, and renders the result insignificant.

Instead, as discussed in LLR, the standard correction is to include unit-specific time trends, which does not suffer from the time-varying attenuation bias problem discussed above.[3] As we showed in the pooled specification in Table 2 of LLR, implementing this renders the effect of Days and Speed of NPI on manufacturing employment and output insignificant. The 95% confidence interval of implementing one NPI for a single day on manufacturing employment is from -0.2% to +0.2%. To put that in context, the estimated confidence interval for the marginal effect on employment for the city with the fewest NPI days (28 days) adopting the most stringent policy (170 days) ranges from -23% to 23%.

[3] A brief explanation of why this correction is appropriate is given in pages 238-240 of Mostly Harmless Econometrics (Angrist and Pischke, 2008).