

1. Use the Penn World Table Data as specified:
  - (a) Create a dummy variable for Mexico.
  - (b) Use *rgdpo* and *emp* to construct output per worker. What is the mean of this variable for Mexico in 1980?
  - (c) Limit the sample to 1980 and regress output per worker on the Mexico indicator. Show the mean in (b). Is the sum of the coefficient and the constant.
  - (d) Repeat (c), but omit the constant term from the regression. What is the coefficient on the Mexican indicator?
  - (e) Create a dummy variable for France. Limit the sample to 1980 and regress output per worker on the Mexican and French indicators. What does the constant term now represent?
  - (f) Using a consistent set of countries for 1980-2019, create an indicator for post-1984 years and regress output per worker on this post-1984 indicator. What does this dummy variable estimate?
  - (g) Using the sample from (f), include country fixed to estimate the post-1984 effect? If you add in year fixed effects as well, what happens to the coefficient of post-1984? Give a one sentence explanation for why this happens.
  - (h) Follow from (f) and (g) and interact the post-1984 measure with the indicators for Mexico and France. Use country and year FE, what do the interactions tell us? Contrast with the estimate in (e).
2. China formally joined the World Trade Organization in December 2001. Assuming that 2002 is the first “treatment” year for China, use the Penn World Table dataset to estimate the following:
  - (a) Restrict the sample to 1990-2019, and estimate the post-2001 difference in the natural log of output for China relative to other countries—i.e., estimate the interaction between a post-2001 dummy and a Chinese dummy while controlling for country and year FE.
  - (b) Interpret the coefficient. What is the relative difference for China after 2001 compared to other countries? Be precise.
  - (c) Plot the event figure for the pre/post estimation in (a).
  - (d) What can we say about pre-trends and the parallel trends assumption?
  - (e) Extend the event figure back to 1950. Can we say anything about the timing?

3. Please use the provided data from Nunn and Qian (2011; hereafter NQ) to answer the following questions. All estimations should include NQ’s baseline set of flexible controls (i.e., interacted with year dummies), which include the log of Old World crop suitabilities, the log of elevation, the log of tropical percent, and the log of ruggedness. Country and year FE should also be included, and standard errors should be clustered at the country level.
- (a) The year 1750 is an arbitrary “turn-on” date for the treatment in NQ. Instead use the true post-Columbian date of 1500 CE to estimate the difference in population and urbanization by potato suitability.
  - (b) Interpret the coefficient above. How does a standard deviation in potato suitability affect population and urbanization after 1500 CE? Be precise.
  - (c) Control for a post-1500 European difference. How does that affect the coefficient of interest?
  - (d) Estimate the event figure from (a), using 1500 CE as the treatment date.
  - (e) Merge in lactase persistence frequencies (LPF) by countrycode from the provided LPF data.
  - (f) Again, using a 1500 CE treatment date, jointly estimate the interaction post-potato, post-LPF, and post-potato\*LPF. What is the effect of potato suitability on population and urbanization when LPF=0? LPF=1?
  - (g) From (g), include a post-European control. Is the interaction between potatoes and milk affected?
  - (h) Plot the event figure for the interaction coefficient from (g)—i.e., the coefficient for post\*potato suitability\*LPF.
4. Please use the provided data from Nunn and Qian (2011; hereafter NQ) to answer the following questions. All estimations should include NQ’s baseline set of flexible controls (i.e., interacted with year dummies), which include the log of Old World crop suitabilities, the log of elevation, the log of tropical percent, and the log of ruggedness. Country and year FE should also be included, and standard errors should be clustered at the country level.
- (a) Assume potatoes entered through Palos, Spain. Find the distance from Palos for each country using the country’s centroid.
  - (b) Assume potatoes diffused linearly across space, beginning in the 1600 CE. What method would you use? Estimate this method.
  - (c) Following from (b), assume also that agricultural conditions favorable to potatoes also matter. Estimate a model that accounts for linear distance, a turn on date for potatoes beginning in 1600 CE, and potato suitability.
  - (d) Interpret coefficients in (c), and argue for/against the spatial diffusion from Palos.
  - (e) What are some critiques of this approach?