

CSSS 594 / POLS 559: Time Series and Panel Data for the Social Sciences

Problem Set 3

Professor: Chris Adolph, Political Science and CSSS
Spring Quarter 2015

Due in class on Tuesday 12 May 2015

General instructions for homeworks: Homework can be handwritten or typed. For any exercises done with R or other statistical packages, you should attach all code you have written and all (interesting) output. Materials should be stapled together in order by problem. The most readable and elegant format for homework answers incorporates student comments, code, output, and graphics into a seamless narrative, as one would see in a textbook. Working in groups on R code is allowed, but (1) each member of the group must provide his or her own writeup and (2) you must list all members of your group.

Problem 1: Race and Wages in Post-Recession America

[100 points total.] This problem considers the relationship between the median wages of white and black full-time employees in the United States using time series stretching from 1979 to the present. While this can be seen simply as an exercise in using our time series analysis tools, it also may help inform debates over the effects of the Great Recession on different demographic groups. Specifically, there is concern that the recession disproportionately affected certain groups – e.g., non-whites, men, or people without college degrees. One way to assess these claims is to look at changes in em-

ployment, income, and wealth during the recession by group. Another way is to ask whether short-run differences emerging between groups are likely to persist in the long-run.

- a. **[10 points.]** In the file `wagesA.csv`, you will find annual data on the median weekly wage in constant 1982–1984 dollars for Whites and Blacks. (Data are taken from the Current Population Survey.) For each of these two variables, use the diagnostic tools we have learned to investigate whether the variable appears to include a deterministic trend, AR terms, MA terms, and/or unit roots. Explain what you find and how you found it. (Do not fit an ARIMA model yet.)
- b. **[25 points.]** Choose and estimate appropriate ARIMA models for White and Black based on your findings in part a. Each model should control for the other time series. Note whether and why you chose to difference either or both time series. Next, broaden your search to other plausible models. Collect goodness of fit statistics from these models including at a minimum the standard error of the regression (in-sample), AIC, and at least one measure of out-of-sample fit. Present the goodness of fit measures across your candidate models in a clear and concise format facilitating intermodel comparison. Discuss the results and select the model you consider to fit the best.
- d. **[20 points.]** Interpret the results from the best models selected in part c. Explain – to the best of your ability – the estimated coefficients. Then present a graphical view of the conditional forecast of White and Black for the next k periods under two scenarios: a “factual” scenario in which any covariates are held to their observed range and/or values, and a “counterfactual” scenario in which the covariate is allowed to vary substantially from its typical values. Explain your choice of k based on your assessment of the reliability of your ARIMA models for short- and long-run forecasting. Discuss the implications of the results for concerns about racial inequality in the economic recovery.
- e. **[10 points.]** Fit error correction models for both White and Black (you may use either the Engle-Granger or Johansen estimators; note your choice). Explain what the statistical and substantive implications of your model coefficients (Box-Steffensmeier *et al* may be helpful here). Discuss the implications of the results for concerns about racial inequality in the economic recovery.

f. [35 points.] The analysis so far assumes that the same data generating process produced wages over the past three and a half decades, despite many debated social and economic changes. It is reasonable to wonder how important this assumption is to our substantive findings, especially if the dynamics of wages and the correlation of wages across racial groups has changed in recent years. To investigate this possibility, we now focus our attention on the period from 2000 to the present, a period in which quarterly data are available from the the Current Population Survey. (These data have not been seasonally adjusted and are in constant 1982–1983 dollars.¹) Two quarterly time series, `WhitesQ` and `BlacksQ`, can be found in `wagesQ.csv`. Repeat the above analysis – making changes as necessary to the specification of models, e.g., regarding stationarity, seasonality, and stochastic processes – and assess whether your substantive findings about the short- and long-run relationship between white and black earnings would change if you could not assume that the relationship between these variables before 2000 has persisted to the present day. Keep your writeup as brief as possible, highlighting differences and similarities to the results on the longer annual time series.

1 I apologize for the mismatch in base years between the two datasets. If you are confident in your ability to do so, feel free to adjust the *quarterly* data to match the 2010 base year of the annual data.