

Exam 3: Making Hay While the Sun Shines

36-402, Spring 2016

Due at **10:30 am** on **Monday**, 9 May 2016

Instructions

Please read the problem background carefully, before beginning the data analysis. Adequate data analysis here *will* require you to go beyond what you know from linear regression, and use methods from this class. You will be graded not just on the technical correctness of your results, but also on the soundness of the reasoning you use to get to the results, and the clarity with which you communicate both your reasons and your results.

You have a maximum of 10 pages.

Allowed and prohibit resources You can use your notes, the textbooks, and anything other printed or electronic reference (with exceptions noted below), if it is properly acknowledged. However, **all your work must be your own**. You may not, under any circumstances, discuss the exam with anyone other than the professors and the teaching assistants. This prohibition specifically includes classmates, friends, relatives, strangers, and people online. You may freely use solutions provided *this semester* for previous homeworks and exams, with acknowledgment. You may not read, copy, “consult”, “study”, “check your work”, or otherwise have anything to do with solutions for this class in previous years. You also may not post this exam or any portion of it to any online forum.

Using prohibited resources, or any form of collaboration, is not just cheating but easily detected cheating. If you find that you have broken these rules, please contact the professors as soon as possible to arrange for an oral examination. Otherwise, being detected in cheating will result in formal academic disciplinary action under the university’s policy on academic integrity.

If you have any questions about what is and is not allowed, please ask the professors.

NOTE THE DEADLINE

Data

The data consists of five economic time series, collected for the United States from the beginning of 1947 to the start of 2016. The data are recorded four times a year, or “quarterly” (so the basic unit of time is a “quarter”). These are

- GDP (inflation adjusted)
- Value of goods consumed (ditto)
- Investment spending (ditto)
- Total hours worked
- Output per hour worked for all non-financial firms (inflation adjusted)

Because most macroeconomic modelers do not concern themselves directly with these series, but only with their fluctuations around their long-run trends, all the series in the data set have been de-trended, and the data file contains the logged fluctuations around the trend.

These five series are selected because of their importance to one of the foundational theories of modern macroeconomics, the “real business cycle” theory, which holds that the business cycle results from the economy reacting (optimally) to “real shocks”, which change what it is actually possible for the economy to produce, as opposed to monetary, financial or social causes. In the original and most widely-taught form of the theory, productivity is an exogenous variable, and the fluctuations in all other variables are ultimately driven by it. People (according to the theory) should work and invest more when they realize productivity has gone up, and more work and investment now should lead to more consumption later. When productivity falls, people should work and invest less.¹

1 Specific Problems

1. (5) Use the data up through 2005 to build a regression model for GDP at time t as a function of the other variables at time t . You should carefully justify your choice of regression model. Report the uncertainty in your estimates in a suitable manner, including a justification for why the way you are calculating those uncertainties is appropriate to this data. Describe how this model indicates GDP is related to the other variables at the same time, including which variables are the most important predictors; be specific in referring to your quantitative results. Finally, use the model you estimate for the data up through 2005 to predict the post-2005 data. How does the predictive performance compare? What does this tell you about the data and/or model?

¹GDP is, by definition, the sum of private consumption, private investment, government expenditures, and net exports.

2. (10) Use the data up through 2005 to build a regression model for GDP at time t as a function of all variables at time $t - 1$ *other than* productivity². Again, carefully justify your choice of regression model, report inferences and uncertainties, and describe in words how the variables at time $t - 1$ are related to GDP at time t . Contrast these relationships to those you found in the previous problem. Finally, once again evaluate the model on post-2005 data.
3. (5) Repeat the previous modeling exercise, but now including productivity at time $t - 1$ as a predictor of GDP at time t . What changes? Which model predicts GDP better? Does the answer change depending on whether we look at data through 2005 or post 2005? What (if anything) does a comparison of the models suggest about whether productivity drives GDP?
4. (5) Fit additive³ regressions of GDP, consumption, investment, and hours worked at time t as functions of all five variables at time $t - 1$. (These are four separate models.) Describe, in words, what these models say about how these variables are related to productivity. Evaluate whether these findings are compatible with the theory sketched above.
5. (5) The original model for productivity in real business cycle theory was a linear first-order autoregressive process with Gaussian noise. Estimate the best first-order autoregressive model for productivity you can (which will not necessarily be linear or Gaussian).
6. (5) Estimate a regression model of productivity at time t as a function of all five variables at time $t - 1$. As above, justify your choice of model, perform inference and report uncertainties, and describe its predictive performance. Does it predict better than the purely autoregressive model from the previous problem? If so, what, if anything, does this suggest about whether productivity is exogenous?
7. (10) Using the analyses above, and any additional data analysis you care to make, provide evaluations of the following proposition: “Exogenous changes in productivity are the main driver of the macroeconomic fluctuations”. If you think that this question cannot be answered with the analytical tools or the data available, explain your reasoning.

Rubric

Words (5) The text is laid out cleanly, with clear divisions and transitions between sections and sub-sections. The writing itself is well-organized, free of grammatical and other mechanical errors, divided into complete sentences logically grouped into paragraphs and sections, and easy to follow from the presumed level of knowledge.

²That is, 2005-12-31 should be the date of the last value of t , not of $t - 1$.

³This is not a hint about what models are best elsewhere.

Numbers (5) All numerical results or summaries are reported to suitable precision, and with appropriate measures of uncertainty attached when applicable.

Pictures (5) All figures and tables shown are relevant to the argument for the ultimate conclusions. Figures and tables are easy to read, with informative captions, axis labels and legends, and are placed near the relevant pieces of text.

Code (10) The code is formatted and organized so that it is easy for others to read and understand. It is indented, commented, and uses meaningful names. It only includes computations which are actually needed to answer the analytical questions, and avoids redundancy. Code borrowed from the notes, from books, or from resources found online is explicitly acknowledged and sourced in the comments. Functions or procedures not directly taken from the notes have accompanying tests which check whether the code does what it is supposed to. All code runs, and the Markdown file knits.

Modeling (10) Model specifications are described clearly and in appropriate detail. There are clear explanations of how estimating the model helps to answer the analytical questions, and rationales for all modeling choices. If multiple models are compared, they are all clearly described, along with the rationale for considering multiple models, and the reasons for selecting one model over another, or for using multiple models simultaneously. Models beyond those covered in 401 are used, and used appropriately.

Inference (10) The actual estimation of model parameters or estimated functions is technically correct. All calculations based on estimates are clearly explained, and also technically correct. All estimates or derived quantities are accompanied with appropriate measures of uncertainty.

Conclusions (10) The substantive, analytical questions are all answered as precisely as the data and the model allow. The chain of reasoning from estimation results about the model, or derived quantities, to substantive conclusions is both clear and convincing. Contingent answers (“if X , then Y , but if Z , then W ”) are likewise described as warranted by the model and data. If uncertainties in the data and model mean the answers to some questions must be imprecise, this too is reflected in the conclusions.

Extra credit (5) Up to five points may be awarded for reports which are unusually well-written, where the code is unusually elegant, where the analytical methods are unusually insightful, or where the analysis goes beyond the required set of analytical questions.

Hints

Linear models In general, if you think the best regression model for a particular problem is linear and estimated by ordinary least squares, you should use it. But you also need to provide very good reasons why that is a better description of the data than non-linear alternatives; failing to do so will lead to very limited partial credit (at best).

Regressing X_t on X_{t-1} and Y_{t-1} See the chapter on time series for functions which can help re-shape data frames to make it easier to fit such models. The examples only illustrate how to use them to regress X_t on X_{t-1}, \dots, X_{t-p} , so you will need to modify the examples, not just copy them.

Use of the block bootstrap If you decide to use a block bootstrap (which is not necessarily the right choice), a reasonable block length is 24 quarters. (You can use this hint as your justification for picking that.)