Overview

- State main result, the continue by answering the following questions which will likely arise directly afterwards: what? why? how?

Introduction

- I develop an algorithm for agnostic and data-driven causal discovery of a unique state-space model and associated family of DSGE models which are *valid* (in a sense that I shall make specific later) relative to some observational data.

- I provide proof that the solution to this procedure is asymptotically unique and consistent, along with empirical evidence that it performs well with realistic sample sizes.

Causal Discovery

- *Causal Discovery* is a field initiated by the work of Judea Pearl, Peter Spirtes, and others involving using large data sets and algorithms to develop models that can make counterfactual predictions.

- One approach to this involves graphical models such as directed acyclical graphs (DAGs).

- Algorithms seek to identify a DAG from data using either a score-based (likelihood maximisation) or constraint-based (conditional independence testing) approach.

DSGE Models

- Type of problem we usually consider has the household solving some utility maximisation problem and firms maximising profits.

- Assuming a competitive equilibrium and market clearing conditions we are able to find a solution to this.

State Space Representation

- Arrows denote vectors because bold not working.

- Usually the solution is found by taking a log-linear approximation to the model.

- For any DSGE model this will yield a solution of the form in equations (1) - (3).

- Assuming this, the only thing we need to know to estimate the solution is the partition of the w_t variables into endogenous states or predetermined variables (x_t), controls (y_t), and exogenous states (z_t).

- The algorithm and test identify this partition.

Results (I)

- Consider the following example.

- Simulate data from a baseline RBC model with exogenous states z (technology) and g (government spending), and endogenous state k (capital).

- When using 100,000 observations we can rule out every partition other than the correct one with a conditional independence test, out of 834 considered.

- The entire problem space contains nearly 20,000 models and we reduce this to just one.

Results (II)

- Now consider a realistic sample size of just 100 observations.

- We run the algorithm 1000 times to assess its performance in this setting.

- The algorithm picks out the correct model nearly 95% of the time, and every time it is not rejected. Nothing else comes close. Again problem space of $\sim$ 20k models.

- Correct model is rejected about 5% of the time because of significance level of the conditional independence test.

Context

- Guido Imbens recently wrote paper discussing the application of DAGs in empirical economics.

- He says "the most important issue holding back the DAGs is the lack of convincing empirical applications. History suggests that those are what is driving the adoption of new methodologies in economics and other social sciences, not the mathematical elegance or rhetoric."

- I aim to remedy this by considering a useful application.

- He is quite negative about DAGs, saying that they are "most useful in complex models with many variables that are not particularly popular in empirical economics."

- Papers like Smets and Wouters (2007) show that macro models can be quite large. I suggest that the problem is the converse: complex models are not popular *because* economists don't have tools like DAGs that make these easy to work with.

Reduction of Problem Space

- Given a set of observations over k variables there are 3^k possible state space representations.

  - Choice of model is *trinomial*

- Reducing this to just one is then a huge reduction in the space of models that need to be considered.

- Does not solve microeconomic dissonace — different microfoundations may give rise to the same state space representation.

Agnosticism

- Assumptions we use are very minimal:

  - The true DGP of the macroeconomy can be modelled as some sort of log-linear DSGE model.
  - *Casual Markov Assumption* — All variables are independent of their non-descendants given their parents.
    * This is the statement that causality is equivalent to conditional independence (as implied by a DAG).
    * DSGE models have this property. So while in general this is a separate assumption it is in this case implied by the first condition.
  - *Faithfulness* — observed conditional independence are real features of the DGP. Effects do not cancel out exactly.
  - Linearity, Gaussian shocks — these can be relaxed.

- As a result the solution produced reflects the data to the greatest extent possible.

Inference

- We cannot pin down a unique DSGE microfoundation given its state space form. However, we can make statements about certain kinds of assumptions.

- For example, if we find that consumption is predetermined, this is evidence that any sensible model should include habits in consumption (although you could come up with some other explanation).

- If we find that inflation is a endogenous state then we have evidence of persistence in inflation, which might be explained by for example indexing. If it is a control we have evidence that inflation is strictly forward-looking.

- These are actual results from the paper. For US data results imply habit formation and inflation persistence should be features of any DSGE model.

- These types of inferences are particularly meaningful in the context of the previous slide.

DAGs

- This captures all of the relationships implied by equations (1) - (3).

Is this sensible?

- Depends on what we are interested in.

- Can still identify the effect of exogenous shocks without modelling endogenous dynamics.

- Cannot identify "deep parameters" in this kind of model.

- We cannot circumvent the Lucas critique in this way. Relates to failure to identify unique DSGE microfoundation.

Conditional Independence (I)

- These relationships are implied by standard theory for DAGs (see paper).

- The conditioning set needs to "d-separate" the targets. Basically contain something from every "backdoor path" and nothing in any "frontdoor path".

- In the paper I show that under the stated assumptions if all conditional independence relationships are known there is exactly one state-space model that satisfies (4), (5) and the minimum state variable criterion (MSV). I refer to the satisfaction of these criteria as *validity*.

Conditional Independence (II)

- In reality conditional independence is not known, so we implement a feasible test.

- Given that we are dealing with Gaussian variables independence is equivalent to zero correlation/covariance, and likewise conditional independence is equivalent to zero correlation/covariance between residuals, having partialed out the conditioning set.

- With some slight modification all 4 conditions can be combined into a single test on a covariance matrix.

- In particular, test whether $\vec{x}_{t-1}$, $\vec{y}_{t-1}$, $\vec{z}_t$, and (optionally) $\vec{z}_{t-2}$ are *completely independent* (covariance matrix is diagonal), conditional on $\vec{x}_{t-2}$ and $\vec{z}_{t-1}$.

- Many statistical tests are available for this purpose, I implement one found in Srivastava (2005).

Scoring

- In finite samples, it is not uncommon to find that more than one model is *valid*, however, it is usually a small number compared to the search space.

- In order to differentiate between this valid models I maximise a score function, which in most cases is the likelihood function of the model. I also implement penalised scores (AIC/BIC), but these are less useful because the models that survive the CI testing tend to have very similar structures/complexity.

- In general, one can just maximise the score function over the set of all models, however, I argue against this approach because it is not necessarily the case that the model with the best predictive accuracy gives the most reasonable causal explanation.

Algorithm

- To find the valid model(s), iterate over all possible models and apply the test (brute force).

- We can make this less onerous by applying MSV and stopping once we have found a valid model.

- If we find more than one valid model choose a winner using the score function.

- This algorithm will consistently identify the unique valid model as $n \to \infty$ with the number of variables fixed because the power of the CI test to reject incorrect models goes to 1.

  - We will still have a significance level α probability of rejecting the correct model, and thus finding no solution in the asymptotic case.
  - To counter this we can lower α as n increases.
  - α is the only tuning parameter of this algorithm.

Limitations

- Does not solve microeconomic dissonance

  - Researchers still need to specify a structural model if they wish to make inferences about the counterfactual effects of structural parameter changes.

- Type II error can be a problem on realistic data sizes

  - We rely on the rejection of incorrect models to a large degree
  - If there are models similar to the true solution and sample sizes are small we can end up with the wrong solution.

- Computational Complexity

  - Considering up to 3^k models can become very costly as k grows.
  - It is unclear the extent to which heuristic algorithms can improve the complexity as the search space is likely quite jagged and also this would threaten the agnosticism of the algorithm which is one of its primary benefits.

Conclusion

- I introduce a test and algorithm to identify a unique state-space model (and associated family of DSGE models) which is *valid* relative to some observed data.

- This procedure is asymptotically consistent and empirical results show it performs well on realistic sample sizes.

- One approach to applying causal discovery tools in the context of economics.