Causal Discovery of DSGE Models

Emmet Hall-Hoffarth

University of Oxford

emmet.hall-hoffarth@economics.ox.ac.uk

October 26, 2020

Overview

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.

What does this mean?

Causal Discovery

- *Causal Discovery* is a field initiated by the work of Pearl, 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.
- Imbens (2019) calls for more applications of graphical models to be demonstrated within the field of economics.

DSGE Models

Consumer problem

$$\max \sum_{t=0}^{\infty} \beta^t U(C_t, L_t, ...) \; s.t.$$

$$P_t C_t + ... \leq W_t L_t + ... \; \forall t$$

$$\lim_{T \to \infty} \beta^T \lambda_{x,T} x_T \to 0 \; \forall x \in \{K, B, ...\}$$

Firm problem

$$\max \; P_t f(L_t, ...) - W_t L_t \; \forall t$$

Competitive equilibrium, market clearing conditions ...

State Space Representation

- Solve out these equations and take a log-linear approximation to the model that takes on the following general form:

$$\vec{y}_t = \vec{A}\vec{x}_{t-1} + \vec{B}\vec{z}_t \tag{1}$$

$$\vec{x}_t = \vec{C}\vec{x}_{t-1} + \vec{D}\vec{z}_t \tag{2}$$

$$\vec{z}_t = \vec{E}\vec{z}_{t-1} + \vec{\epsilon}_t \tag{3}$$

- This is a partition of the variables $\vec{w}_t$ in the model into three categories:
  - $\vec{x}_t$: predetermined or endogenous state variables
  - $\vec{y}_t$: control variables or "jumpers"
  - $\vec{z}_t$: exogenous state variables
- The algorithm identifies this partition (and the matrices $\vec{A}$ - $\vec{E}$).

Why is this useful?

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.
- Note that this does not mean that we identify a unique DSGE model — 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 (Spirtes & Zhang, 2016).
    - 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 indepedence 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.

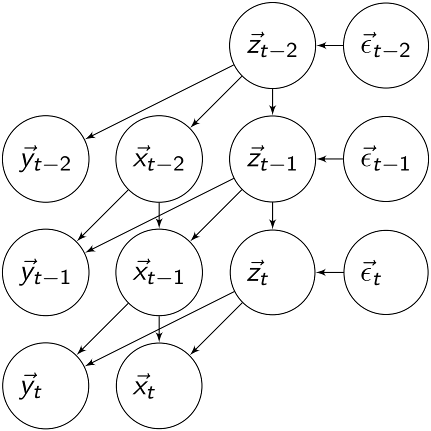
Inference

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

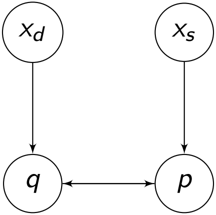
How is this even possible?

DAGs

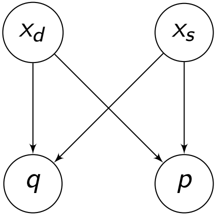
- We can represent the state space representation as a DAG.



(Aside) Is this sensible?



$$\iff$$



$$p = \alpha_p + \beta_{ps}x(s) + \beta_{pq}q + \epsilon_p$$
$$q = \alpha_q + \beta_{qd}x(d) + \beta_{qp}p + \epsilon_q$$

$$p = \frac{1}{1 - \beta_{pq}\beta_{qp}}[(\alpha_p + \beta_{pq}\alpha_q) + \beta_{ps}x(s) + \beta_{pq}\beta_{qd}x(d) + (\epsilon_p + \beta_{pq}\epsilon_q)]$$
$$q = \frac{1}{1 - \beta_{qp}\beta_{pq}}[(\alpha_q + \beta_{qp}\alpha_p) + \beta_{qd}x(d) + \beta_{qp}\beta_{ps}x(s) + (\epsilon_q + \beta_{qp}\epsilon_p)]$$

- Understanding response to shocks does not require inference of "deep" parameters (β_{xy}).

Conditional Independence (I)

- Given the Causal Markov assumption we can infer the following relationships from this DAG:

$$x_t \perp\!\!\!\perp x'_t \,||\, [\vec{x}_{t-1}, \vec{z}_t] \text{ for all } (x_t, x'_t) \in [\vec{x}_t, \vec{y}_t] \tag{4}$$

$$x_{t-1} \perp\!\!\!\perp z_t \,||\, \vec{z}_{t-1} \text{ for all } x_{t-1} \in \vec{x}_{t-1} \text{ and } z_t \in \vec{z}_t \tag{5}$$

$$x_t \perp\!\!\!\perp z_{t-1} \,||\, [\vec{x}_{t-1}, \vec{z}_t] \text{ for all } x_t \in [\vec{x}_t, \vec{y}_t] \text{ and } z_{t-1} \in \vec{z}_{t-1} \tag{6}$$

$$z_t \perp\!\!\!\perp z'_t || \vec{z}_{t-1} \text{ for all } z_t \neq z'_t \in \vec{z}_t \tag{7}$$

- 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) (McCallum, 1999). 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.

Results (I)

- Baseline RBC model.
- 9 Observables, true partition is $\vec{z} = [g \ z]$, $\vec{x} = [k]$, $\vec{y} = [y \ c \ i \ r \ w \ l]$.
- With full sample of 10^5 observations the only valid partition is the ground truth.
- 834 models considered (models with $1 \leq \#$ state variables ≤ 3).
- 19683 possible models reduced to 1.

Results (II)

- 1000 iterations, 100 sample size
- Wins = number of times model selected, Valid = number of times model is valid relative to CI test

Index	Exogenous States	Endogenous States	Wins	Valid
1	g z	k	944	944
2	g w	k	27	729
3	g y	k	27	571
4	c g	k	2	8
5	g l y		0	340
6	g r y		0	421
7	g r	k	0	576
8	g l z		0	716
9	g i r		0	781
10	g i l		0	629
11	g i	k	0	867
12	g r w		0	609
13	g r z		0	858
14	g k l		0	625
15	g l w		0	603
16	g k r		0	779
17	c g w		0	1

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.

Citations

Imbens, G. (2019). *Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics* (tech. rep.). National Bureau of Economic Research.

McCallum, B. T. (1999). Role of the minimal state variable criterion in rational expectations models. *International finance and financial crises* (pp. 151–176). Springer.

Spirtes, P., & Zhang, K. (2016). Causal discovery and inference: Concepts and recent methodological advances. *Applied informatics*, *3*(1), 3.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, *35*(2), 251–272.