DAGs with Applications to Macroeconomic Timeseries

Emmet Hall-Hoffarth

University of Oxford

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

August 3, 2020

Overview

1. Background
   - DAGs
     - Representation
     - Structure learning
   - DSGE
2. Application
   - RBC
   - Real data
3. Papers to Write
   - DSGE-specific strucuture learning
   - Real data
   - Choose your own states
   - Lasso
4. Conclusion

DAGs

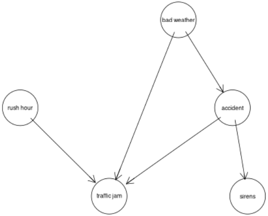
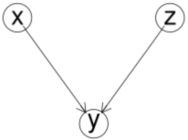


Figure: An example of a simple DAG

- Root nodes of the graph are *i.i.d.* shocks.
- Parent nodes are exogenous relative to their children.
- Each arc specifies a (generic) conditional probability relationship.
- I will assume that the data follows a joint normal distribution *s.t.* each conditional probability relationship is a linear function with a normally distributed error.

Structure Learning



$$x = \epsilon_x$$

$$y = \beta_{yx}x + \beta_{yz}z + \epsilon_y$$

$$z = \epsilon_z$$

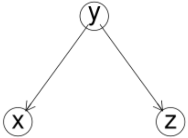
(a) Collider

$$x = \epsilon_x$$

$$y = \beta_{yx}x + \epsilon_y$$

$$z = \beta_{zy}y + \epsilon_z$$

(b) Chain



$$x = \beta_{xy}y + \epsilon_x$$

$$y = \epsilon_y$$

$$z = \beta_{zy}y + \epsilon_z$$

(c) Fork

Figure: The three possible v-structures of a 3 node DAG. Error terms ϵ are all i.i.d. Gaussian shocks.

Structure Learning

- These three V-structures are partially identifiable because they have different testable implications about conditional (in)dependence.
  - Collider $\implies x \perp\!\!\!\perp z$ but $x \not\perp\!\!\!\perp z|y$
  - Chain $\implies x \not\perp\!\!\!\perp z$ but $x \perp\!\!\!\perp z|y$
  - Fork $\implies x \not\perp\!\!\!\perp z$ but $x \perp\!\!\!\perp z|y$
- In the context of a linear Gaussian model these conditional (in)dependence relationships can be tested with a t-test.
- Many algorithms have been developed which use this basic idea to identify (to the greatest extent possible) the underlying DAG given observed data. This are *Constraint-Based Algorithms*.
- There are also *Score-Based Algorithms* that maximise a score (eg. likelihood) over the set of possible DAGs using gradient decent, and *Hybrid Algorithms* that combine both approaches.

DSGE

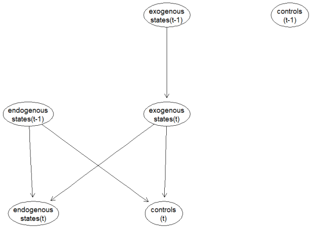


Figure: The general solution to a DSGE model as a DAG

- Exogenous states: Current value determines current state of model and depends only on its own past. (eg. technology)
- Endogenous states: Current value depends on states, but past values determine current state of model. (eg. capital)
- Controls: Current value depends on states, past values are irrelevant. (eg. consumption)

DSGE

- The DAG on the previous slide defines the *ground-truth* that we seek to identify from data generated by a DSGE simulation.
- Structure learning provide asymptotic theoretical guarantees that they can identify the ground truth given some observational data.
- This model is essentially a regularised ADL regression. Therefore, structure learning provides an agnostic way to regularise and choose contemporaneous regressors for the ADL model.
- So everything is in order then?

Application

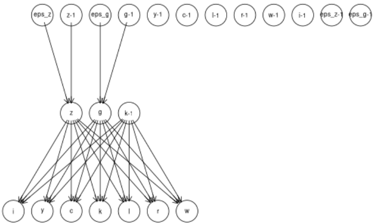


Figure: Ground truth solution to RBC model

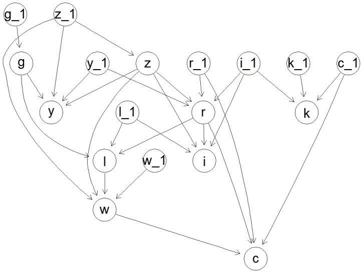
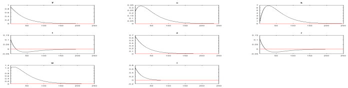


Figure: Estimated DAG for RBC data (hybrid algorithm, 100k observations)

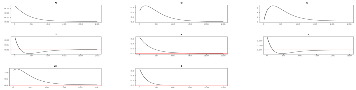
Application

- Not quite....
- There seeems to be a number of problems with these algorithms, especially using simulated data:
    - Faithfulness: We require that every dependency in the underlying DGP correspond to an observed statistical dependence. This is violated if some effects cancel out resulting in no observed correlation (eg mon. auth. chooses interest rate to cancel effect of inflationary shock on output)
    - Collinearity: Calculating conditional correlation requires calculating regression residuals. In simulation, these are very close to zero conditional on the correct states, so the computation is unstable.
    - Asymptotics: The problem space grows super-exponentially with the number of variables $\implies$ even a sample size of 100k may not be big enough for asymptotic arguments.

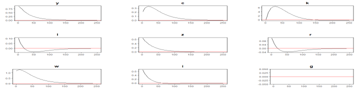
Application



(a) Original Simulation



(b) Ground Truth DAG



(c) Fitted DAG (Hybrid Algorithm)

Figure: IRFs to a one standard deviation technology shock.

- Even if structure learning does not return the ground-truth it does (sometimes) return a model that is causally equivalent in the sense that correctly identifies the states and replicates the true IRFs.
- Furthermore, when using real data concerns about faithfulness, collinearity, and even asymptotics (if we consider a small number of variables) are less pronounced...

Application

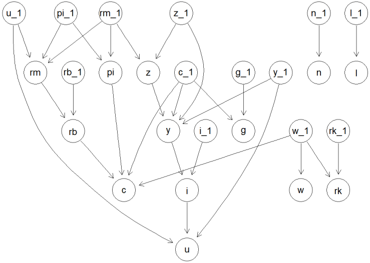


Figure: Estimated DAG for US economy 1980-2005 (hybrid algorithm)

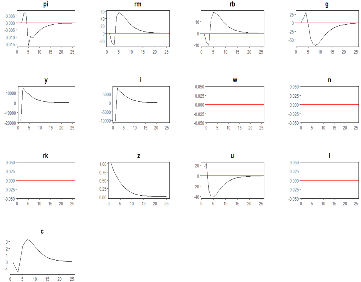


Figure: IRFs to TFP shock for estimated DAG

Application

- Since there is no ground-truth availible we can consider this DAG against commonly accepted stylised facts.
- The parents of rm, other than its own lag are u_{t-1} and π_{t-1} suggesting Taylor Rule like behaviour. There is a causal pathway from past monetary policy into current unemployment implying that monetary policy is a viable stabilising tool.
- n and l are separate from the rest of the graph suggesting the exogeneity of these variables, which is entirely reasonable.
- z_t feed directly into y_t, w_{t-1} into c_t, and i_t into u_t, all of which seem like plausible causal linkages.
- The IRFs seem to reflect many stylised facts about the macroeconomy in terms of direction of effect, even if the implied responses leave something to be desired, especially with regards to their jagged shape.

DSGE-Specific Structure Learning

- Another problem with existing structure learning algorithms is that they are *too agnostic* relative to what we are willing to assume in this context.
- If the DAG is a solution to a DSGE model it must take on a specific form, as illustrated in Figure 3.
- By explicitly assuming the resulting DAG takes on this form we can drastically reduce the size of the search space and likely greatly improve the resulting estimated DAG.
- I have already implemented this, and the only significant roadblock seems to be the collinearity issue as discussed earlier.

Focus on Real Data

- The original idea of this research was to use simulated data to give credence to the application to real data.
- As it turns out there are many problems with simulated data the appear to not be an issue for real data.
- This suggestion, therefore, is to ignore the simulations and instead provide a number of real data applications in the paper and evaluate those models against well-understood intuitions.

"Choose your own states adventure"

- If a DSGE-like solution is assumed, a researcher needs only to choose what they assume to be the exogenous and endogenous states in order to estimate a model.
- I propose a methodology whereby researchers first test different combinations of states to see whether they produce expected IRFs, and only then go to develop a structural model that contains those states.
- This approach is data-driven and mostly agnostic, as apposed to the *ad hoc* approach that is commonly employed.

Lasso

- Assuming a DSGE solution the estimated VAR should be sparse (the only relevant regressors are the states). Therefore, it makes sense to use penalised regression to identify these states, which can then be fed into a DAG.
- This approach seems to work very well even on simulated data.
- Does not differentiate between exogenous and endogenous states (because it is a VAR instead of an ADL). That is left to the researcher to decide.
- Potentially not a very original research contribution.

Conclusion

- DAGs can allow for agnostic, data-driven discovery of causal models, and there are likely many applications in economics where this will prove useful.
- My research aims to demonstrate one such applicaion by utilizing macroeconomic timeseries data.
- While initial results have been mixed, DAGs do so some promise, especially as a regularisation technique.
- The exact direction of the paper(s) that I will write on these topics has yet to be determined.