Suggestions from Robert Waldmann

Waldmann's suggestions un-emphasized; commentary by CRS indented and italicized.

1   E-mail of 2015-10-03

1) permutating variables is absolutely brilliant. I think the approach of considering all permutations is valid – no one can suspect that you chose a permutation which makes SW look horrible. But it makes the results abstract.

I really liked the idea of substituting investment and hours worked. A few permutations which you can describe might be much more striking than the calculation of distributions over the full set of permutations.

The problem is that the few permutations to present must be chosen ex ante, publicly and not by you three. I think the thing to do is issue a public challenge – ask cyberspace to propose permutations (with survey monkey or something) then report results with the 5 most often proposed permutations. I think it would be interesting to ask respondents to say if they are pro DSGE, anti DSGE, or open minded. Yes most will claim to be open minded, but I suspect that people will have a strong guess about a permutation which will work OK (say permute capital and labor) and one which will work very badly (permute hours worked and the interest rate). I also guess that the effort to find a permutation which makes the model work worse than the SW model will not be very successful (I am assuming that people won't bother doing simulations to find the permuted models which work even worse).

> *Since we look at all permutations, I don't see the point to this.*

2) estimating parameters. Here, once upon a time, DSGE modelers (who were then all RBC not New Keynesian) claimed they got the parameters from micro data or long term trends. Estimating with macro series and a "prior" which has worked well in the past and is popular in the literature is a degeneration. I am confident that DSGE fans will argue that they have long known that one shouldn't do what you did. They have (also they do it). I don't know how to deal with people who do things they know are invalid and reject the same things when done by critics because they are invalid.

> *In a way our stuff here, and in the risk-bounds paper, does provide a sort of back-handed support for calibration (or despair of ever estimating macro models); I guess we should own this.*

3) the point of micro foundations is that they are supposed to give good conditional forecasts conditional on a change in policy. I think in the simulate, estimate, test out of sample effort it would be interesting to simulate, estimate, change policy parameters and test. The problem with ad hoc reduced form models is supposed to be that they include parameters which are not deep and structural and the same for different policy regimes but which depend on the policy. Changing the policy rule is supposed to illustrate the advantage of DSGE

> *Potentially interesting. Presumably the model we use to generate the new-policy trajectories would be SW. How hard would this be to do for a few of the best permutations?*

4) well just simulate and estimate. You note that the parameter estimates do not converge to the parameters of the DGP. I think it would be interesting to see if the estimates of taste and technology parameters depend on the policy (Taylor rule) parameters. The whole point of micro foundations is that deep parameters are supposed to be policy invariant. I think it is worth checking if the allegedly deep SW parameters are policy invariant.

> *Again, potentially interesting. Look at covariance between these parameters over simulation runs?*

5) point 2 is a council of despair – you won't win because they will cheat as they always do. Points 3 and 4 are proposals to make the manuscript longer. Wrong Robert. Another barrier to publication is that the manuscript is much too long. I suspect it has a better chance as one permutation manuscript and one estimate with more data than we really have manuscript.

> *Not sure what to do about that.*

2   E-mail of 2015-10-05

1) Some possible snark. People don't argue that models are The Truth but rather that they may be the best available approximation to the truth. If so, the SW model must be the best approximation among the 7! = 5040 (if I calculate correctly) permutations of the SW model.

This is a finite mechanically defined set of models one of which is much more highly thought of than the other 5039. Best available =¿ best out of 5040 randomly generated models.

Somehow some statement along the line of the above might be fun.

> *I don't follow this well enough to have an opinion.*

2) I continue to applaud the decision to look at all the permutations, because of the total elimination of any risk of data dredging, data snooping and ex post choice of the silly model which performs better than SW. But it is abstract. I can't help wanting to see what happens with one (or a few) specifications. I

think choosing a specification based on within sample performance is sort of like a kind of estimation (sort of because I couldn't tell calculate degrees of freedom). I am very interested in a comparison of the out of sample performance of the permutation with the highest within sample likelihood. I am also interested in the out of sample performance of the one with the highest within sample penalized likelihood (but more nearly able to fight the fascination).

The fact that well over half of permutations work better out of sample suggests that checking this won't be too painful.

The point (if any) of this exercise is to get something more concrete than uh those numbers which you graphed (distribution of the calculated posterior to within sample estimation probability density of out of sample observations ??? I don't have confidence in my memory or even my understanding of the legends of the figures). I really really want to see time series of forecast real GDP hours worked and price inflation calculated by the computer when it was told that real GDP is price inflation and price inflation is the nominal interest rate and hours worked is wage inflation. That is I want to visually compare the performance of the totally confused computer which has been deceived by obviously arrant nonsense and the computer which has been programmed with the best available DSGE model (that is arrant nonsense which isn't obviously nonsense to many macroeconomists).

In any case, I think even semi sane readers would like to know which permutation gives the highest likelihood (and I guess the highest penalized likelihood and the best out of sample performance).

> *Time series of predictions for a few of the best permutations, plus the baseline, seems reasonable (and fast).*

3) I can't banish the thought that the manuscript is better as 2 papers (so the original project will become 3). The lack of consistency results are as devastating as the permutations can improve the model results. But they are much less funny and novel. There are lots of journals with extremely strict page limits (also ones where long manuscripts aren't banned but are penalized).

Also on consistency, it is clear that there is a pseudo Bayesian estimator which will give the parameters used to generate the pseudo data as pseudo estimates, if the pseudo prior is a degenerate prior with probability 1 on those values. Similarly there are pseudo Bayesian estimators which will give estimates close to the true values if the prior is highly concentrated on those values. (my frequent use of the word pseudo shows that I will not just will not use "scare quotes," but I admit I was "tempted").

In the DSGE business a lot of theory is used to justify the specification but the priors are pulled out of hats. I haven't heard any references to introspection and sincere belief that the prior mode is the most likely value. I think it is obvious that no human being has a prior probability distribution for the parameters.

The older calibration literature included the claim that the macroeconomic time series weren't used at all to estimate parameters. I think this claim was

3

false (the chosen parameters had nothing to do with micro data based estimates and were clearly chosen to make the simulated time series act like the observed time series). But the literature, including the current literature, is based in some sense on the recognition that there aren't enough observations to estimate the parameters. I think it is based on the realization that, since there aren't enough data to pin them down, it is possible to fit summary statistics by claiming to have the right prior.

No actionable suggestion, except for splitting into more papers.

4) I will now run out on thin ice and make a (frozen) fool of myself by typing about the abuse of Bayesian statistics (a topic which you understand uh shall we say better than I do and avoid quantification of my relative ignorance). The following text is especially a waste of time.

a) a "Bayesian model" is a contradiction in terms. Bayesian priors and posteriors are beliefs. Models are things we don't believe are true, but think might be useful. the prior density on any vector of parameters in a model is zero - we are sure that they don't describe the data generating process. We can have a prior on parameters of a parametric class of hypotheses, but we have to believe that The Truth might belong to this class for any prior probability density or prior probability to be positive.

b) a lot of the problem is that people call maximizing penalized likelihoods "Bayesian. " This is only true if the penalty is based on sincere a-priori probability densities. Otherwise the estimator is a Stein class estimator. There is no need to sincerely believe in a penalty function in order to use one. This is a twitty argument about semantics, but I can't resist.

c) I think the problem of choosing the best pseudo prior using an within sample statistic is interesting – the idea is to find a statistic analogous to the Akaike information criterion or the "B"IC (note the scare quotes) so that the expected value of the out of sample loss is monotonic in the statistic. I don't think this is a very hard problem. I assume solutions are published somewhere in the statistics literature, but I don't see them applied in the econometrics literature.

d) I am sure that the so called priors in the macroeconomics and time series econometrics literatures are not prior at all. the models are so far abstracted from our actual experience that I just don't see how priors could be elicited. The specifications have been modified in order to deal with anomalies – the models aren't at all the models which economists found attractive back at the dawn of the literature (in the 70s). I don't doubt at all that the so called priors were chosen to give good performance out of the so called samples, this so called out of sample performance evaluated by looking at different models including different pseudo priors and always the same allegedly out of sample observations. That is clearly data snooping. This must have happened, because there is no other plausible way to chose the so called priors.

How can anyone claim to have a prior on the parameters of a Calvo Poisson alarm clock ? The prior can't be innate and it can't be based on personal

experience with Poisson alarm clocks. Why is it that macroeconomists have completely different beliefs about the Frisch elasticity of labor supply than labor economists ? How can one think about someone's labor supply without knowing if that one is a single man or a married woman ? I think the answer to the last question is to put all common sense and real world experience out of mind, but then one can't have a prior.

> *Sensible, as far as I can follow, but not actionable. At most it calls for a haughty footnote.*

3   E-mail of 2015-10-31

1) how about a test of how likely it is that the SW model would perform so badly if it were The Truth (AKA the data generating process). That is simulate data with the SW model and compare the performance fitting and out of sample forecasting those data with the SW model and permuted models.

The mechanical way to do this is 10000 simulations then (I think) estimation of 5040 permutations of each. I think it might make sense to do a lot of simulations and estimate only, say 1 permuted model per simulation (or 10 chosen at random). Since there are choices of exactly how many permutations, I think it would be better to announce the experiment publicly before setting the computers to work (to avoid the suspicion of dredging of simulated pseudo data).

> *This seems (unfortunately) like it'd be desirable in principle. But I'm sure it'd take a very long time, and even if it was feasible, "permutations fit better by about as much as you'd expect if the model was true" is not exactly easy to interpret, or a ringing endorsement of SW.*

2) or the other way, what if the permuted model with the highest penalized likelihood were The Truth ? How badly would the SW model fit and forecast ? Here I actually guess it would do OK.

> *This seems easy, and gets at the but-are-the-permutations-really-different question.*

3) I don't know if I have said this already, but I would be very interested in an up close and personal look at the permuted model with the highest penalized likelihood. I trust that it fits better out of sample than the SW model. I also trust that it will look silly to theorists (who are therefore the silly ones).

> *Seems simple to do, if we haven't.*

Now 4 points
a) I can't resist trying to guess the results of estimation
b) I can't resist trying to understand complicated theoretical models and guess

their implications using heuristic verbal reasoning.

c) My guesses are always wrong

d) but see points a&b above.

So here goes. I guess that a big problem for SW is the key role of hours worked [heuristic non rigorous theoretical arguments typed and deleted]. In the data the variable isn't so key [1960s style regressions estimate and not reported]. I guess that the permuted models which work best replace hours worked with another flow variable (GDP consumption or investment).

Easy to check, with and acknowledgment to R.W. if he's right.

4   Recommendations/triage

In order of decreasing priority (balancing ease of implementation against impact):

1. Suggestion 3.4: check whether the best permuted models swap hours worked for "another flow variable".

2. Suggestion 2.2: pick some of the best permutations and plot their predictions along with those of the baseline, un-permuted model.

3. Suggestion 3.3: a detailed examination of the best permuted model.

4. Suggestion 1.4: do the "deep" parameters co-vary with the policy parameters?

5. Suggestion 1.3: look at out-of-sample forecasts under a different policy rule.

6. Suggestion 3.1: p-value for how much the SW model is beaten by its permutations.