# Waldman triage

*DJM*

*June 6, 2016*

Outline: see waldman-suggestions

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.

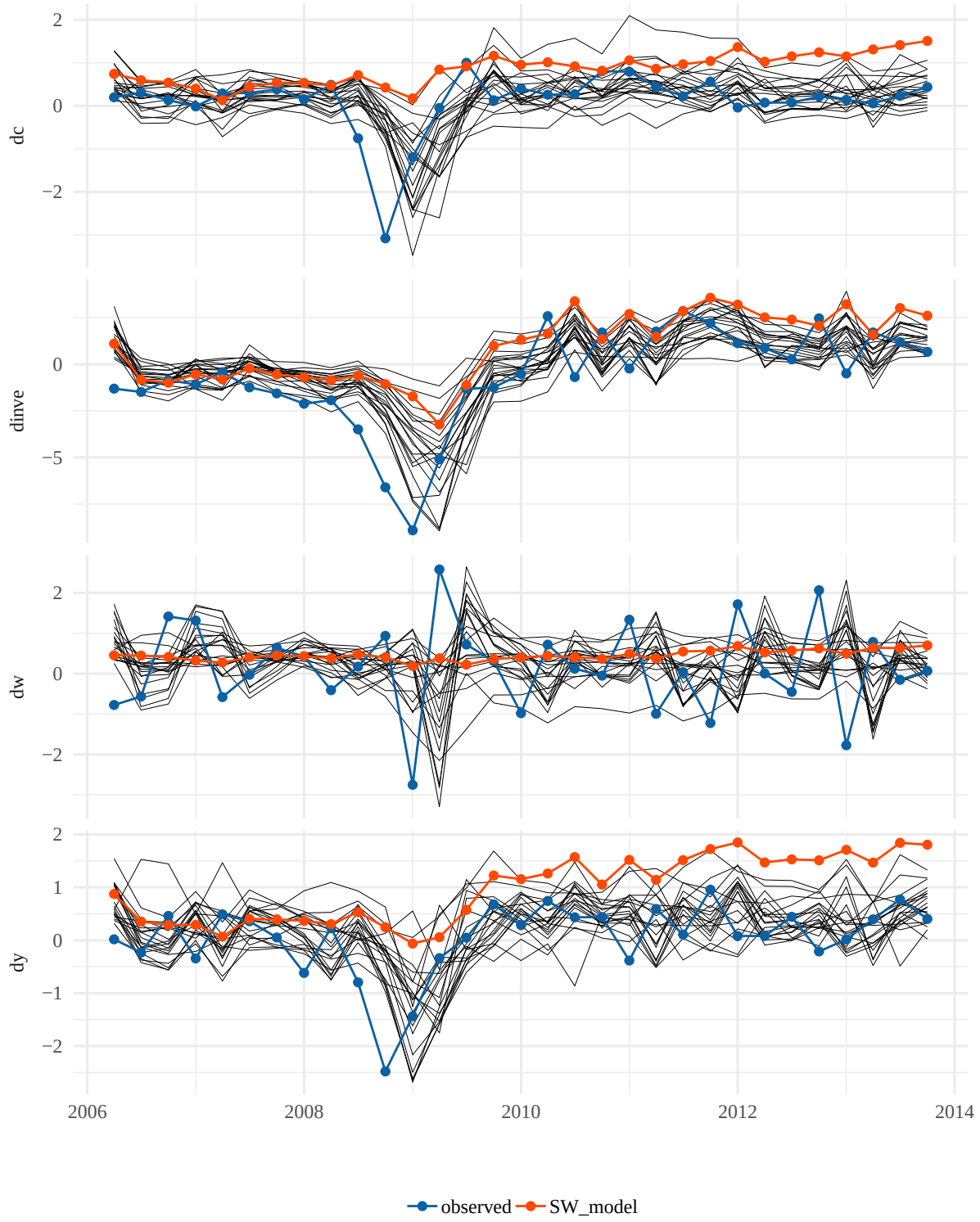
## 1 Does the best (permuted) model swap hours worked?

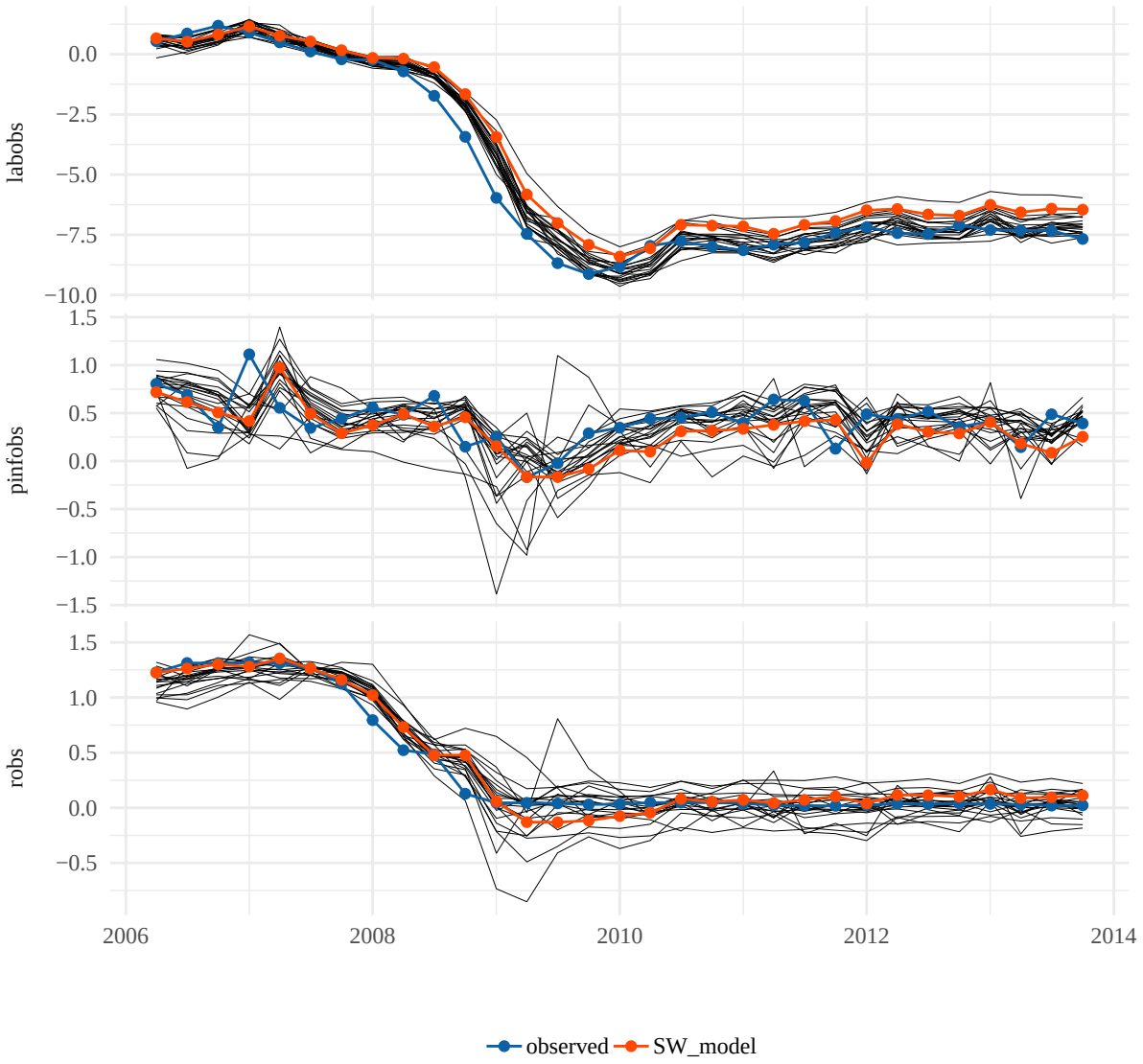
Table 1: Top 20 permutations based on (penalized) log-likelihood. The top row is the correct ordering. Hours worked should be in the first column.

| labobs | robs | pinfobs | dy | dc | dinve | dw |
|---|---|---|---|---|---|---|
| labobs | pinfobs | robs | dy | dw | dc | dinve |
| labobs | robs | pinfobs | dc | dw | dy | dinve |
| labobs | robs | pinfobs | dy | dw | dinve | dc |
| pinfobs | labobs | robs | dy | dc | dw | dinve |
| pinfobs | robs | labobs | dy | dw | dinve | dc |
| pinfobs | robs | labobs | dc | dw | dy | dinve |
| pinfobs | robs | labobs | dy | dc | dinve | dw |
| labobs | robs | pinfobs | dw | dy | dinve | dc |
| labobs | robs | pinfobs | dinve | dy | dw | dc |
| pinfobs | robs | labobs | dc | dinve | dy | dw |
| robs | pinfobs | labobs | dinve | dc | dy | dw |
| robs | pinfobs | labobs | dy | dc | dinve | dw |
| robs | pinfobs | labobs | dc | dw | dy | dinve |
| labobs | robs | pinfobs | dw | dc | dy | dinve |
| robs | labobs | pinfobs | dc | dy | dinve | dw |
| labobs | robs | pinfobs | dw | dinve | dy | dc |
| pinfobs | robs | labobs | dw | dy | dinve | dc |
| labobs | pinfobs | robs | dw | dy | dinve | dc |
| pinfobs | robs | labobs | dy | dw | dc | dinve |
| labobs | pinfobs | robs | dc | dy | dinve | dw |

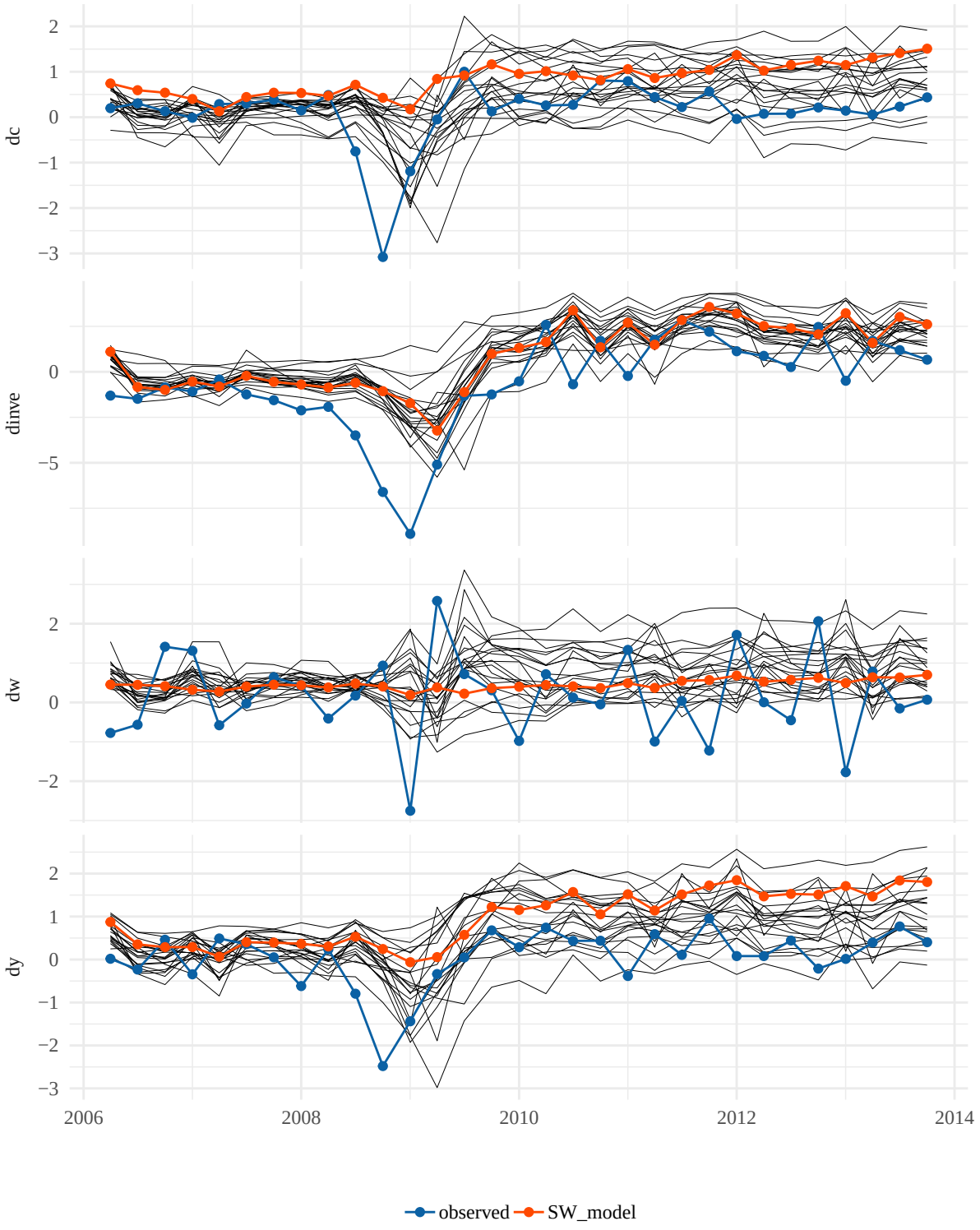
## 2 Plot some predictions along with the best model

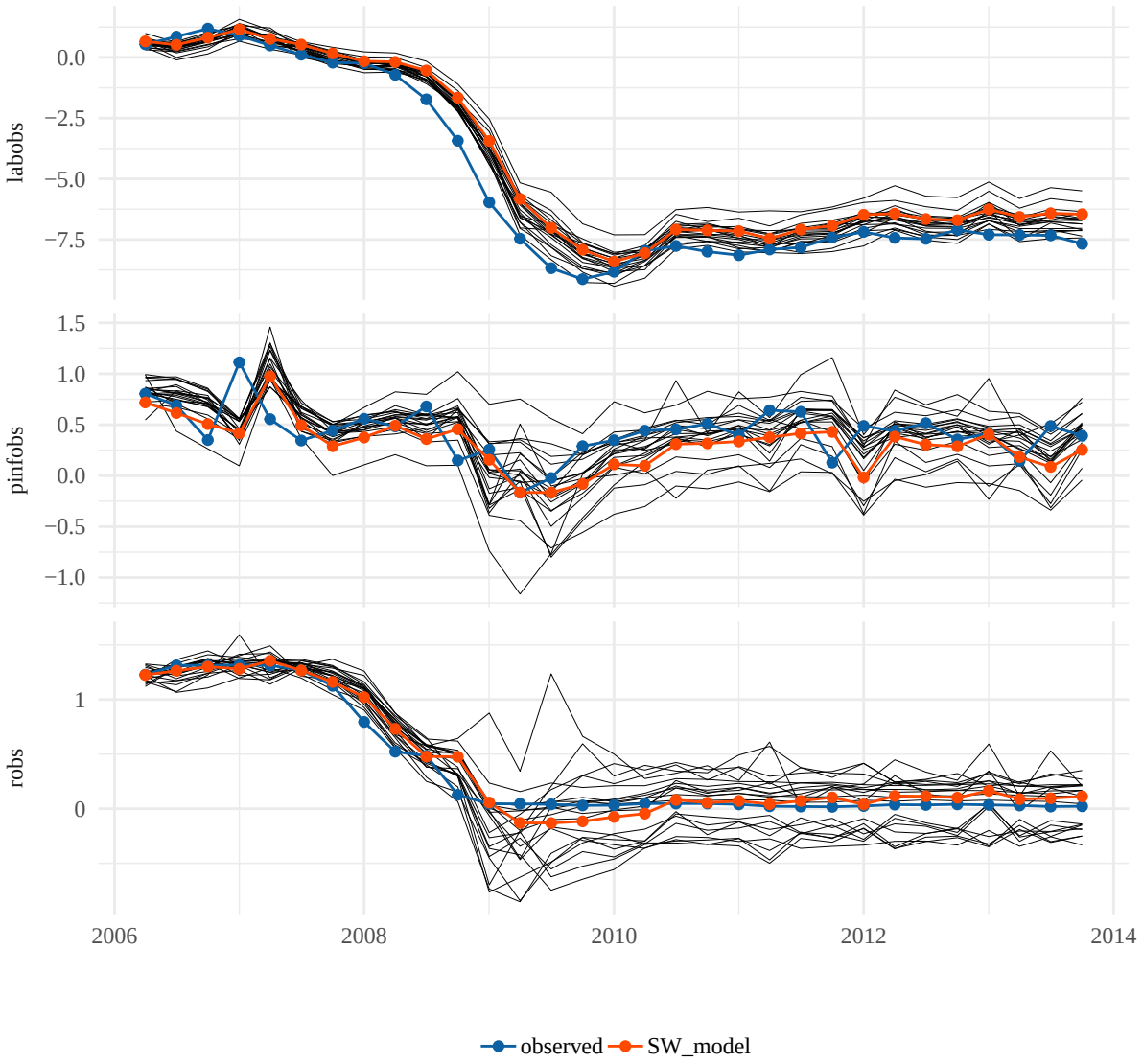
The following plots show the top 20 flips based on "average percent improvement" (this is a post-hoc measure). Blue-dotted is observed data while red is the SW model.




labobs
pinfobs
robs
0.0
−2.5
−5.0
−7.5
−10.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
2006
2008
2010
2012
2014
observed
SW_model

The next set of plots is the Top 20 permutations by negative (penalized) log-likelihood.




labobs
0.0
−2.5
−5.0
−7.5
pinfobs
1.5
1.0
0.5
0.0
−0.5
−1.0
robs
1
0
2006
2008
2010
2012
2014
observed
SW_model

# 3 Best model by penalized log-likelihood

The true model is

labobs, robs, pinfobs, dy, dc, dinve, dw

while the best one is

labobs, pinfobs, robs, dy, dw, dc, dinve.

This model is pretty clearly scary to a macroeconomist. First of all, the standard model says that the monetary authority sets the interest rate based on inflation and deviation of output from trend. This model reverses cause and effect: inflation is the control variable while the interest rate is the input to the Taylor rule. I think, actually, the most interesting exercise would be to copy and paste the first section of the SW paper, but change out all the notation.

I think it's also important to note here, that the penalized negative log-likelihood is very flat relative to permutations. In terms of that metric, 6.2% of permutations are within 10% of the true permutation.

# 4 Do the 'deep' parameters co-vary with the policy parameters?

Table 2: Correlations between 'deep' parameters and Taylor rule parameters. From the 'Simulate and estimate' exercise.

| | r[pi] | rho | r[y] | r[Delta][y] |
|---|---|---|---|---|
| phi1 | 0.197 | 0.097 | -0.034 | 0.113 |
| sigma[c] | 0.061 | 0.257 | 0.169 | 0.017 |
| h | 0.326 | 0.076 | 0.016 | -0.343 |
| xi[w] | -0.055 | 0.427 | 0.020 | 0.082 |
| sigma[l] | 0.006 | 0.190 | 0.033 | -0.373 |
| xi[p] | -0.261 | 0.360 | 0.125 | 0.186 |
| iota[w] | 0.096 | 0.036 | 0.052 | 0.007 |
| iota[p] | 0.027 | 0.168 | -0.027 | 0.038 |
| Psi | 0.340 | 0.005 | 0.104 | 0.331 |
| Phi | 0.019 | 0.031 | 0.030 | 0.021 |
| r[pi] | 1.000 | 0.315 | 0.235 | -0.004 |
| rho | 0.315 | 1.000 | 0.324 | 0.122 |
| r[y] | 0.235 | 0.324 | 1.000 | -0.036 |
| r[Delta][y] | -0.004 | 0.122 | -0.036 | 1.000 |
| bar(pi) | -0.037 | -0.048 | -0.077 | 0.008 |
| 100(beta^{-1} -1) | -0.037 | -0.184 | -0.143 | -0.043 |
| bar(l) | 0.066 | 0.104 | 0.077 | -0.050 |
| bar(gamma) | 0.017 | 0.035 | 0.050 | -0.015 |
| rho[ga] | -0.234 | -0.146 | -0.005 | 0.007 |
| alpha | 0.179 | 0.108 | 0.162 | -0.039 |

# 5 Simulate from the model, change Taylor parameters, examine forecasts

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?*

I'm not sure I see the relevance of this one. He seems to suggest simulating from the baseline many times, estimating on part of the time series, then changing the parameters and forecasting the rest of the time series (which was actually generated with different conditions). Is the goal to determine if conditional forecasts are accurate? It seems like we really want to do some sort of hypothesis test as follows:

1. Generate data up to some change point $T_1$. At the change point, alter only the policy parameters and generate the future from $T_1$ to $T_2$.
2. Estimate a model with a change point at $T_1$. See whether allowing only the policy parameters to vary is sufficient, or if we need to let all the 'deep' parameters vary.
3. Test the null that the policy parameters are enough (LRT).

Considering that we're generating data out of, and then using, the correct model, I find it hard to imagine we would reject the null (true) model with a change point.

# 6 P-value

This is awful for the reasons previously discussed.

# 7 Another one (3.2)

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.*

Procedure:

1. generate data out of the best permutation.
2. get forecasts using it and the SW model (say 100 each). Which is better?

The average negative log-likelihood of the true model is -1220 compared to 409 for the SW model. The MSEs are 0.00389 compared to 0.00442 respectively. The next set of figures shows the data (black), predictions from the true model (blue), and SW model (red).