Learning in an estimated medium-scale DSGE model

Sergey Slobodyan and Raf Wouters

DYNARE Conference
Paris,
September 11, 2007

Motivation for estimating a medium–scale model with learning

- Most empirical DSGE models retain the hypothesis of RE in the sense that expectations of agents are model consistent. How restrictive is this hypothesis for an estimated DSGE model?

Motivation for estimating a medium–scale model with learning

- Most empirical DSGE models retain the hypothesis of RE in the sense that expectations of agents are model consistent. How restrictive is this hypothesis for an estimated DSGE model?
- DSGE-VAR methodology of Del Negro et al (2007) measures and identifies potential mis-specification in the Smets & Wouters model. Is the RE assumption responsible for this observation? Can adaptive learning process contribute to a solution?

Motivation for estimating a medium–scale model with learning

- Most empirical DSGE models retain the hypothesis of RE in the sense that expectations of agents are model consistent. How restrictive is this hypothesis for an estimated DSGE model?
- DSGE-VAR methodology of Del Negro et al (2007) measures and identifies potential mis-specification in the Smets & Wouters model. Is the RE assumption responsible for this observation? Can adaptive learning process contribute to a solution?
- Milani (2004), Orphanides & Williams (2003) claim that learning can significantly influence the macroeconomic dynamics and increase the persistence especially in the inflation process. How robust are these claims in a medium-scale DSGE model that fits the data relatively well?

How do we introduce learning in the model?

- We follow Evans & Honkapohja (2001), Milani and Orphanides & Williams by assuming that economic agents do not have perfect knowledge of the reduced form parameters of the model when forming expectations about the future.

How do we introduce learning in the model?

- We follow Evans & Honkapohja (2001), Milani and Orphanides & Williams by assuming that economic agents do not have perfect knowledge of the reduced form parameters of the model when forming expectations about the future.
- Agents forecast future values of the lead variables with a linear function in the state and exogenous variables. These beliefs are updated with a constant–gain Recursive Least Squares (CG RLS) procedure.

Main results of the paper

- Relaxing the model–consistent expectations assumption can improve the marginal likelihood of the model, and approach the optimal fit obtained by the DSGE–VAR.

Main results of the paper

- Relaxing the model–consistent expectations assumption can improve the marginal likelihood of the model, and approach the optimal fit obtained by the DSGE–VAR.
- This result depends strongly on the specification of the initial beliefs and on the information set used in the forecasting equations. The best performing models are those where the initial beliefs are optimised to maximise the in–sample fit.

Main results of the paper (Cont)

- Limiting the information set in the forecasting equations can also improve the marginal likelihood.

Main results of the paper (Cont)

- Limiting the information set in the forecasting equations can also improve the marginal likelihood.
- The best performing learning models generate moderate time variation in the IRF.

Main results of the paper (Cont)

- Limiting the information set in the forecasting equations can also improve the marginal likelihood.
- The best performing learning models generate moderate time variation in the IRF.
    - The response of inflation to a monetary policy shock becomes more gradual and more persistent.

Main results of the paper (Cont)

- Limiting the information set in the forecasting equations can also improve the marginal likelihood.
- The best performing learning models generate moderate time variation in the IRF.
    - The response of inflation to a monetary policy shock becomes more gradual and more persistent.
    - The response of inflation to a productivity shock is immediate and very short–lived.

Main results of the paper (Cont)

- Limiting the information set in the forecasting equations can also improve the marginal likelihood.
- The best performing learning models generate moderate time variation in the IRF.
  - The response of inflation to a monetary policy shock becomes more gradual and more persistent.
  - The response of inflation to a productivity shock is immediate and very short–lived.
- These results resolve some of the shortcomings of the REE model and come closer to the IRFs of the DSGE–VAR approach.

The REE model of Smets and Wouters (2007)

- The medium–scale DSGE model based on CEE (2005) with many real and nominal frictions (habit persistence, capital adjustment costs, Calvo prices and wages, indexation to past price and wage inflation for non–optimizers, interest–rate smoothing) is linearized around the REE with a deterministic trend growth rate.

The REE model of Smets and Wouters (2007)

- The medium–scale DSGE model based on CEE (2005) with many real and nominal frictions (habit persistence, capital adjustment costs, Calvo prices and wages, indexation to past price and wage inflation for non–optimizers, interest–rate smoothing) is linearized around the REE with a deterministic trend growth rate.
- Augmented with seven stochastic shocks: five AR(1) processes (TFP productivity shock, risk premium shock, investment specific technology shock, public spending shock and monetary policy shock) and two ARMA(1,1) (price and wage mark–up).

The REE model of Smets and Wouters (2007) (Cont)

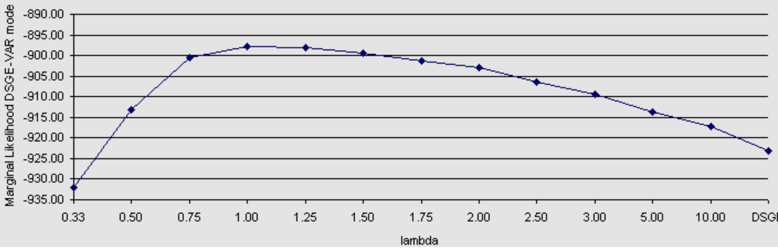
Estimated on US data over the period 1966:1 - 2005:4 using seven macro variables

$$\begin{bmatrix} dlGDP_t \\ dlCons_t \\ dlINV_t \\ dlWag_t \\ lHOURS_t \\ dlP_t \\ FEDFUNDS_t \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \end{bmatrix} + \begin{bmatrix} \hat{y}_t - \hat{y}_{t-1} \\ \hat{c}_t - \hat{c}_{t-1} \\ \hat{i}_t - \hat{i}_{t-1} \\ \hat{w}_t - \hat{w}_{t-1} \\ \hat{l}_t \\ \hat{\pi}_t \\ \hat{R}_t \end{bmatrix}$$

The REE model of Smets and Wouters (2007) (Cont)

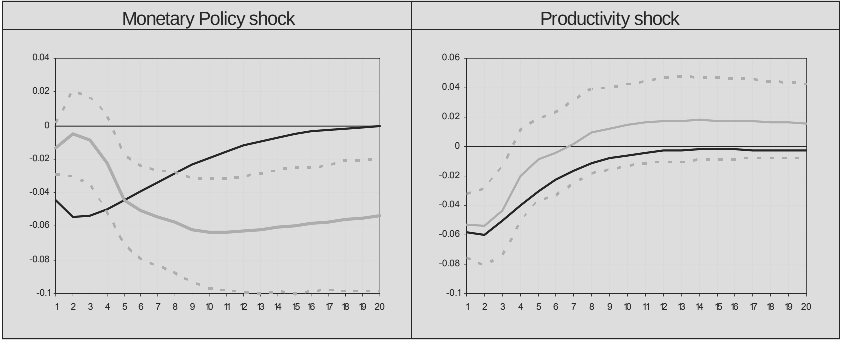
DSGE–VAR (model restrictions are imposed as a prior on the VAR(4)) with optimised weight ($\lambda = 1.25$) improves the marginal likelihood compared to the strict REE–DSGE model.

	Benchmark REE-DSGE model	DSGE-VAR model (λ=1.25)
Posterior Mode	-841.46	-834.17
Log Data Density (Laplace)	-923.05	-897.77
Log Data Density (Mod.Harm.Mean)	-922.15	-897.78



The REE model of Smets and Wouters (2007) (Cont)

Impulse Responses of a monetary policy and a productivity shock on inflation in the DSGE–VAR model and in the REE–DSGE model suggest misspecification in the restrictive REE–DSGE model.



The grey lines represent the benchmark DSGE-VAR IRF (mode in bold and 90% interval).

The black line is the response in the DSGE model.

Learning set-up

- Model representation in DYNARE:

$$A_0 \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + A_1 \begin{bmatrix} y_t \\ w_t \end{bmatrix} + A_2 E_t y_{t+1} + B\epsilon_t = 0,$$

where y_t is a vector of 31 endogenous variables, and w_t a vector of 9 exogenous processes including the moving average innovations.

Learning set-up

- Model representation in DYNARE:

$$A_0 \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + A_1 \begin{bmatrix} y_t \\ w_t \end{bmatrix} + A_2 E_t y_{t+1} + B\epsilon_t = 0,$$

where y_t is a vector of 31 endogenous variables, and w_t a vector of 9 exogenous processes including the moving average innovations.

- The RE solution of this system is:

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu + T \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R\epsilon_t.$$

Learning set-up (Cont)

- With learning, agents forecast future variables of the lead variables with a linear function of states and exogenous variables (under MSV learning):

$$y_t^f = \alpha_{t-1} + \beta_{t-1}^T \begin{bmatrix} y_{t-1}^s \\ w_t \end{bmatrix}.$$

Learning set-up (Cont)

- With learning, agents forecast future variables of the lead variables with a linear function of states and exogenous variables (under MSV learning):

$$y_t^f = \alpha_{t-1} + \beta_{t-1}^T \begin{bmatrix} y_{t-1}^s \\ w_t \end{bmatrix}.$$

- Agents' beliefs about the coefficients α (12x1) and β (12x20) are updated using the constant-gain RLS algorithm:

$$\phi_t = \phi_{t-1} + gR_t^{-1} Z_{t-1}(y_t^f - \phi_{t-1}^T Z_{t-1})^T$$
$$R_t = R_{t-1} + g(Z_{t-1} Z_{t-1}^T - R_{t-1}),$$

where $Z_t = (y_{t-1}^s, w_t^T)^T$, $\phi_t^T = (\alpha_t, \beta_t^T)$, and g is the constant gain parameter.

Learning set-up (Cont)

- We also consider other sets of variables used for forecasting:
  — agents are learning the steady state values of inflation, interest rate, hours, and the growth rate;
  — agents use a limited set of information in forecasting (VAR beliefs in model counterparts of observables; only levels are used).

Learning set-up (Cont)

- We also consider other sets of variables used for forecasting:
  — agents are learning the steady state values of inflation, interest rate, hours, and the growth rate;
  — agents use a limited set of information in forecasting (VAR beliefs in model counterparts of observables; only levels are used).

Learning set-up (Cont)

- We also consider other sets of variables used for forecasting:
  — agents are learning the steady state values of inflation, interest rate, hours, and the growth rate;
  — agents use a limited set of information in forecasting (VAR beliefs in model counterparts of observables; only levels are used).
- Thus obtained beliefs are used to solve the purely backward–looking DYNARE model and to obtain a representation

$$\begin{bmatrix} y_t \\ w_t \end{bmatrix} = \mu_t + T_t \begin{bmatrix} y_{t-1} \\ w_{t-1} \end{bmatrix} + R_t \epsilon_t .$$

Learning set-up (Cont)

- Time–dependent matrices replace DYNARE–produced μ, R, and T, and are then used in the Kalman filtering step.

Learning set-up (Cont)

- Time–dependent matrices replace DYNARE–produced μ, R, and T, and are then used in the Kalman filtering step.
- The updating of the belief coefficients at any time t depends on the data (best estimates of the state, the lead and the exogenous variables at respectively time $t-1$ and t) and on the initial beliefs.

Learning set-up (Cont)

- Time–dependent matrices replace DYNARE–produced μ, R, and T, and are then used in the Kalman filtering step.
- The updating of the belief coefficients at any time t depends on the data (best estimates of the state, the lead and the exogenous variables at respectively time $t-1$ and t) and on the initial beliefs.
- Best estimates are filtered values of the model variables taken from the Kalman filter that is used to construct the likelihood function of the model.

Learning set-up (Cont)

- Time–dependent matrices replace DYNARE–produced μ, R, and T, and are then used in the Kalman filtering step.
- The updating of the belief coefficients at any time t depends on the data (best estimates of the state, the lead and the exogenous variables at respectively time $t-1$ and t) and on the initial beliefs.
- Best estimates are filtered values of the model variables taken from the Kalman filter that is used to construct the likelihood function of the model.
- In principle, one could use smoothed rather than filtered estimates, re–smoothing every period and re-estimating past beliefs. This would represent a more consistent usage of available information, but is computationally very intensive and is not performed here.

Learning set-up: Initial beliefs

As it turns out that the results are very sensitive to the initial beliefs, we consider four alternative ways of selecting them.

- Initial beliefs are always consistent with the REE model evaluated at the estimated parameters:

$$\phi_0^T = E[ZZ^T]^{-1}E[y^f Z^T],$$
$$R_0 = E[ZZ^T].$$

In particular, this implies that for different parameter sets in the posterior, the initial beliefs are different.

Learning set-up: Initial beliefs

As it turns out that the results are very sensitive to the initial beliefs, we consider four alternative ways of selecting them.

- Initial beliefs are always consistent with the REE model evaluated at the estimated parameters:

$$\phi_0^T = E[ZZ^T]^{-1} E[y^f Z^T],$$
$$R_0 = E[ZZ^T].$$

  In particular, this implies that for different parameter sets in the posterior, the initial beliefs are different.

- Initial beliefs are consistent with a REE model estimated over pre–sample data.

Learning set-up: Initial beliefs (Cont)

- Initial beliefs are consistent with *some* REE and are selected to maximize the in–sample fit of the model. An extra model used to construct the belief is estimated simultaneously with the actual model; both models share the parameters of the exogenous processes.

Learning set-up: Initial beliefs (Cont)

- Initial beliefs are consistent with *some* REE and are selected to maximize the in–sample fit of the model. An extra model used to construct the belief is estimated simultaneously with the actual model; both models share the parameters of the exogenous processes.
- Initial beliefs are based on a regression with pre–sample data, using the filtered data from a model estimated under REE using pre–sample data.

Learning set-up: Additional Issues

- Agents are always assumed to have perfect information on the properties of the exogenous processes under MSV learning (under VAR learning, exogenous processes are excluded from the PLM).

Learning set-up: Additional Issues

- Agents are always assumed to have perfect information on the properties of the exogenous processes under MSV learning (under VAR learning, exogenous processes are excluded from the PLM).
- During updating, the transition matrix T_t is restricted to the stable domain by a version of a projection facility: if the largest eigenvalue of T_t is outside of the unit circle, we keep last period ϕ and R.

Learning set-up: Additional Issues

- Agents are always assumed to have perfect information on the properties of the exogenous processes under MSV learning (under VAR learning, exogenous processes are excluded from the PLM).
- During updating, the transition matrix T_t is restricted to the stable domain by a version of a projection facility: if the largest eigenvalue of T_t is outside of the unit circle, we keep last period ϕ and R.
- A standard projection facility (checking roots of the forecasting Vector Autoregression) cannot be implemented, as the relationship between lead (LHS of the PLM) and state (RHS of the PLM) variables depends on the solution of the model. T_t is the forecasting VAR for all model variables, including lead, state, and static.

Learning set-up: Additional Issues (Cont)

- The effective gain gR_t^{-1} is prevented from becoming too large with a ridge correction mechanism: if the smallest eigenvalue of R_t^{-1} is lower than some λ, we use $(R_t + \lambda I)^{-1}$ instead.

Learning set-up: Additional Issues (Cont)

- The effective gain gR_t^{-1} is prevented from becoming too large with a ridge correction mechanism: if the smallest eigenvalue of R_t^{-1} is lower than some λ, we use $(R_t + \lambda I)^{-1}$ instead.
- Optimization of the initial beliefs together with the model tends to generate R_0 with tiny smallest eigenvalue.

Learning set-up: Additional Issues (Cont)

- The effective gain gR_t^{-1} is prevented from becoming too large with a ridge correction mechanism: if the smallest eigenvalue of R_t^{-1} is lower than some λ, we use $(R_t + \lambda I)^{-1}$ instead.
- Optimization of the initial beliefs together with the model tends to generate R_0 with tiny smallest eigenvalue.
- Such initial beliefs exploit initial data points and adjust ϕ very fast. They are likely to perform poorly for slightly different data or parameter values.

Learning set-up: Additional Issues (Cont)

- The effective gain gR_t^{-1} is prevented from becoming too large with a ridge correction mechanism: if the smallest eigenvalue of R_t^{-1} is lower than some λ, we use $(R_t + \lambda I)^{-1}$ instead.
- Optimization of the initial beliefs together with the model tends to generate R_0 with tiny smallest eigenvalue.
- Such initial beliefs exploit initial data points and adjust ϕ very fast. They are likely to perform poorly for slightly different data or parameter values.
- As a result, optimization runs with R_0 with tiny smallest eigenvalue tend to converge to bad posterior mode, produce low marginal likelihood, or both.

Simulation results

- We generate random time series (1000 obs.) for REE model and different versions of the learning model: MSV learning (with and without constant) and VAR learning, and analyse the transitional (first 150 obs.) and permanent dynamics (last 150 obs.)

Simulation results

- We generate random time series (1000 obs.) for REE model and different versions of the learning model: MSV learning (with and without constant) and VAR learning, and analyse the transitional (first 150 obs.) and permanent dynamics (last 150 obs.)
- For each model, we draw 1000 random initial beliefs from the posterior distribution of the REE model: initial beliefs are consistent with the REE implied moment matrixes for ϕ_0 and R_0.

Simulation results

- We generate random time series (1000 obs.) for REE model and different versions of the learning model: MSV learning (with and without constant) and VAR learning, and analyse the transitional (first 150 obs.) and permanent dynamics (last 150 obs.)
- For each model, we draw 1000 random initial beliefs from the posterior distribution of the REE model: initial beliefs are consistent with the REE implied moment matrixes for ϕ_0 and R_0.
- We repeat the exercise for different values of the learning gain $g = [0.01, 0.02, 0.05]$.

Simulation results

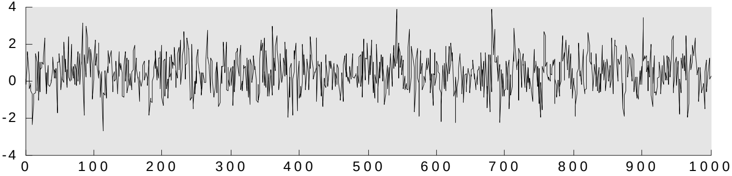
- We generate random time series (1000 obs.) for REE model and different versions of the learning model: MSV learning (with and without constant) and VAR learning, and analyse the transitional (first 150 obs.) and permanent dynamics (last 150 obs.)
- For each model, we draw 1000 random initial beliefs from the posterior distribution of the REE model: initial beliefs are consistent with the REE implied moment matrixes for ϕ_0 and R_0.
- We repeat the exercise for different values of the learning gain $g = [0.01, 0.02, 0.05]$.
- We impose the projection facility during the updating process.

	REE model		Learning with MSV beliefs			
			$\gamma = 0.01$		$\gamma = 0.05$	
	transition	permanent	transition	permanent	transition	permanent
St. Dev						
Output growth	0.94	0.94	0.94	0.94	0.98	1.89
Hours	2.33	2.43	2.20	2.44	2.43	5.89
Inflation	0.50	0.51	0.48	0.51	0.60	2.40
Interest rate	0.54	0.55	0.54	0.55	0.59	2.50
Autocorrelation						
Output growth	0.27	0.28	0.26	0.28	0.30	0.46
hours	0.97	0.97	0.96	0.97	0.97	0.97
Inflation	0.82	0.82	0.80	0.83	0.83	0.81
Interest rate	0.89	0.89	0.89	0.90	0.91	0.91

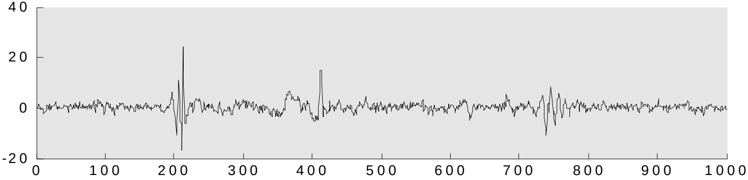
	Learning with MSV beliefs				Learning with MSV beliefs + cte			
	γ = 0.01		γ = 0.05		γ = 0.01		γ = 0.05	
	transition	permanent	transition	permanent	transition	permanent	transition	permanent
St. Dev								
Output growth	0.94	0.94	0.98	1.89	0.94	0.95	0.99	2.56
Hours	2.20	2.44	2.43	5.89	2.20	2.44	2.47	9.89
Inflation	0.48	0.51	0.60	2.40	0.48	0.52	0.61	4.25
Interest rate	0.54	0.55	0.59	2.50	0.54	0.56	0.60	5.16
Autocorrelation								
Output growth	0.26	0.28	0.30	0.46	0.26	0.28	0.31	0.52
hours	0.96	0.97	0.97	0.97	0.96	0.97	0.97	0.97
Inflation	0.80	0.83	0.83	0.81	0.80	0.83	0.86	0.84
Interest rate	0.89	0.90	0.91	0.91	0.89	0.90	0.93	0.93

	Learning with MSV beliefs + cte				Learning with VAR beliefs			
	γ = 0.01		γ = 0.05		γ = 0.01		γ = 0.05	
	transition	permanent	transition	permanent	transition	permanent	transition	permanent
St. Dev								
Output growth	0.94	0.95	0.99	2.56	0.98	0.88	1.66	2.93
Hours	2.20	2.44	2.47	9.89	3.00	2.43	4.11	10.87
Inflation	0.48	0.52	0.61	4.25	0.76	0.56	1.86	8.68
Interest rate	0.54	0.56	0.60	5.16	0.69	0.60	1.45	9.39
Autocorrelation								
Output growth	0.26	0.28	0.31	0.52	0.39	0.30	0.59	0.55
hours	0.96	0.97	0.97	0.97	0.98	0.96	0.97	0.97
Inflation	0.80	0.83	0.86	0.84	0.85	0.84	0.89	0.93
Interest rate	0.89	0.90	0.93	0.93	0.92	0.90	0.96	0.97

MSV learning - gain = 0.01

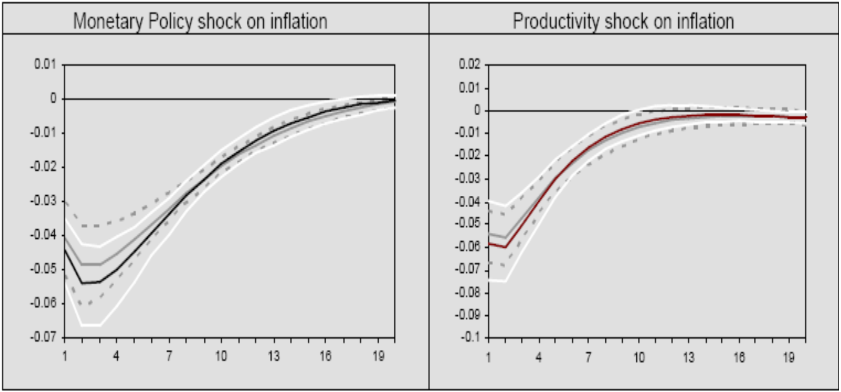


MSV learning - gain = 0.05

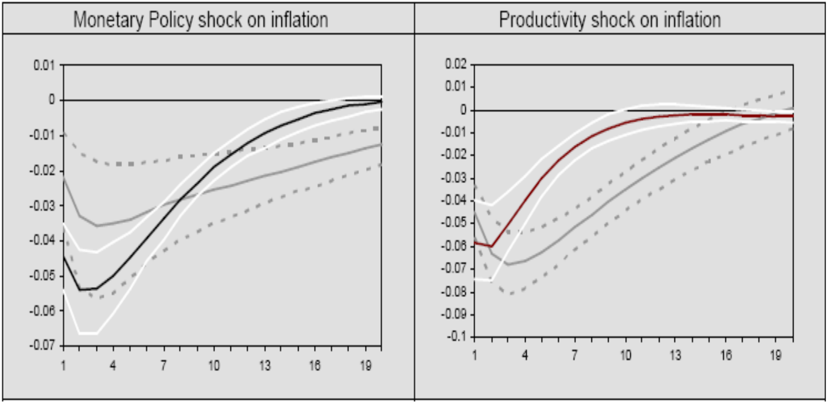


	REE model		Learning with MSV beliefs + cte (gain = 0.02)				Learning with VAR beliefs (gain = 0.02)			
			including escapes		excluding escapes		including escapes		excluding escapes	
[redacted]										
St. Dev										
Output growth	0.94	0.94	0.94	0.96	0.94	0.95	1.24	1.12	1.07	0.88
Hours	2.33	2.43	2.26	2.50	2.25	2.45	3.44	3.41	3.16	2.38
Inflation	0.50	0.51	0.50	0.57	0.49	0.52	1.20	1.35	0.91	0.66
Interest rate	0.54	0.55	0.54	0.59	0.54	0.56	0.89	1.51	0.77	0.79
Autocorrelation										
Output growth	0.27	0.28	0.26	0.29	0.26	0.28	0.51	0.39	0.48	0.30
hours	0.97	0.97	0.96	0.97	0.96	0.97	0.97	0.97	0.98	0.96
Inflation	0.82	0.82	0.81	0.83	0.81	0.83	0.86	0.88	0.87	0.87
Interest rate	0.89	0.89	0.89	0.90	0.89	0.90	0.93	0.94	0.93	0.93

Impact of modest changes in the beliefs on the IRF of monetary policy and productivity shocks on inflation under MSV learning with learning gain = 0.02.



Impact of modest changes in the beliefs on the IRF of monetary policy and productivity shocks on inflation under VAR learning with learning gain = 0.02.



Model comparison in terms of Marginal Likelihood.

REE-DSGE model	-922
DSGE-VAR(λ=1)	-898
Learning with MSV beliefs	
Consistent initial beliefs	-922
Consistent initial beliefs + cte	-922
Optimised initial beliefs	-911
Presample model based initial beliefs	-944
Presample regression based initial beliefs	tba
Learning with VAR beliefs	
Consistent initial beliefs	-922
Optimised initial beliefs	-904
Presample model based initial beliefs	-938
Presample regression based initial beliefs	tba

Model comparison in terms of estimated parameters.

	φ	λ	ξ_w	ι_w	ξ_p	ι_p	ρ_r	gain
REE-DSGE model	5.49	0.71	0.74	0.59	0.66	0.23	0.82	-
DSGE-VAR(λ=1)	3.84	0.66	0.69	0.51	0.65	0.39	0.78	-
Learning with MSV beliefs								
Consistent initial beliefs	5.30	0.71	0.74	0.62	0.62	0.20	0.82	0.012
Consistent initial beliefs + cte	5.30	0.71	0.74	0.62	0.62	0.20	0.82	0.012
Optimised initial beliefs	2.83	0.79	0.64	0.59	0.60	0.18	0.84	0.017
Presample model based initial beliefs	3.11	0.62	0.58	0.53	0.49	0.53	0.52	0.024
Presample regression based beliefs								
Learning with VAR beliefs								
Consistent initial beliefs	3.89	0.72	0.74	0.61	0.68	0.07	0.82	0.001
Optimised initial beliefs	3.39	0.74	0.71	0.57	0.62	0.07	0.85	0.001
Presample model based initial beliefs	4.25	0.79	0.75	0.47	0.61	0.57	0.87	0.002
presample regression based beliefs								

Estimation results

- MSV learning with model consistent initial beliefs generates IRFs that remain constant over time, although the gain is relatively high. These IRFs are very similar to the REE-DSGE model.

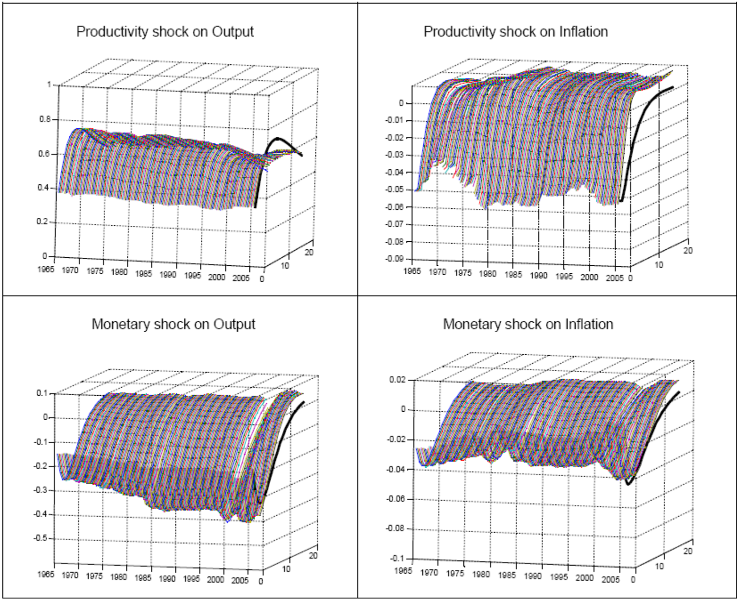
Estimation results

- MSV learning with model consistent initial beliefs generates IRFs that remain constant over time, although the gain is relatively high. These IRFs are very similar to the REE-DSGE model.
- MSV learning with optimised initial beliefs generates IRFs with modest time variation. The IRFs for inflation are closer to the DSGE-VAR IRF: the misspecification is reduced considerably.

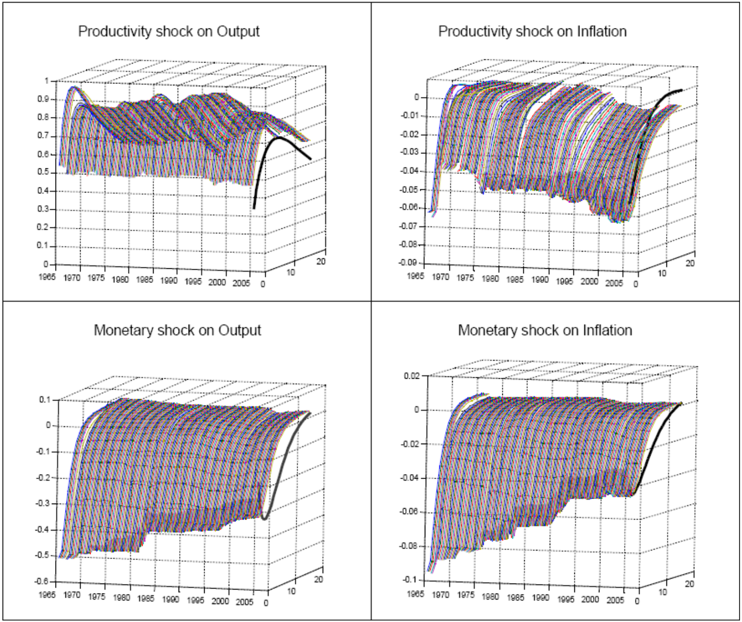
Estimation results

- MSV learning with model consistent initial beliefs generates IRFs that remain constant over time, although the gain is relatively high. These IRFs are very similar to the REE-DSGE model.
- MSV learning with optimised initial beliefs generates IRFs with modest time variation. The IRFs for inflation are closer to the DSGE-VAR IRF: the misspecification is reduced considerably.
- MSV learning with pre-sample based initial beliefs generates mixed results.

IRFs for the MSV learning with optimized initial beliefs.



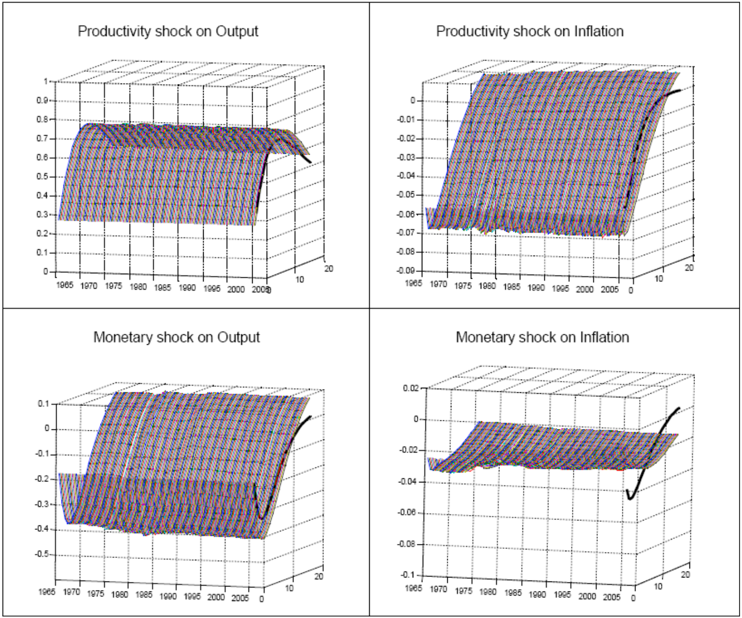
IRFs for the MSV learning with pre–sample based initial beliefs.



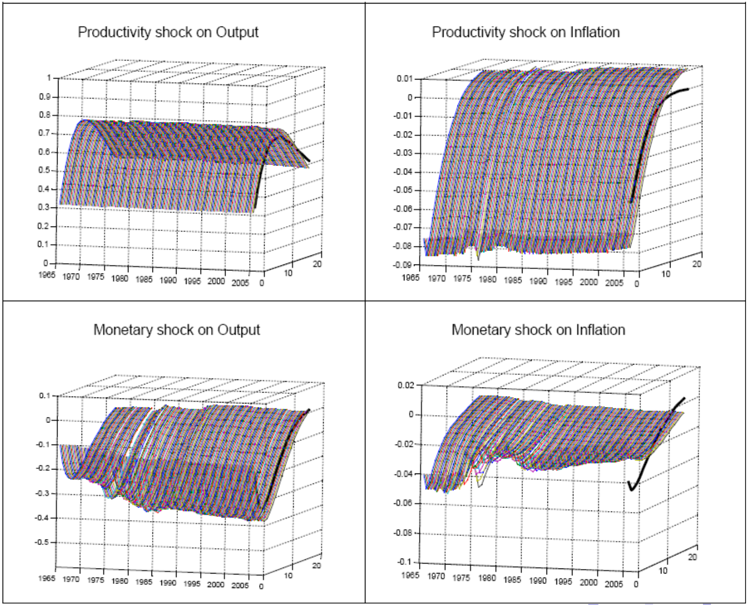
Estimation results (Cont)

VAR learning with any initial belief specification generates a more gradual and persistent response of inflation to monetary policy, in line with the DSGE-VAR.

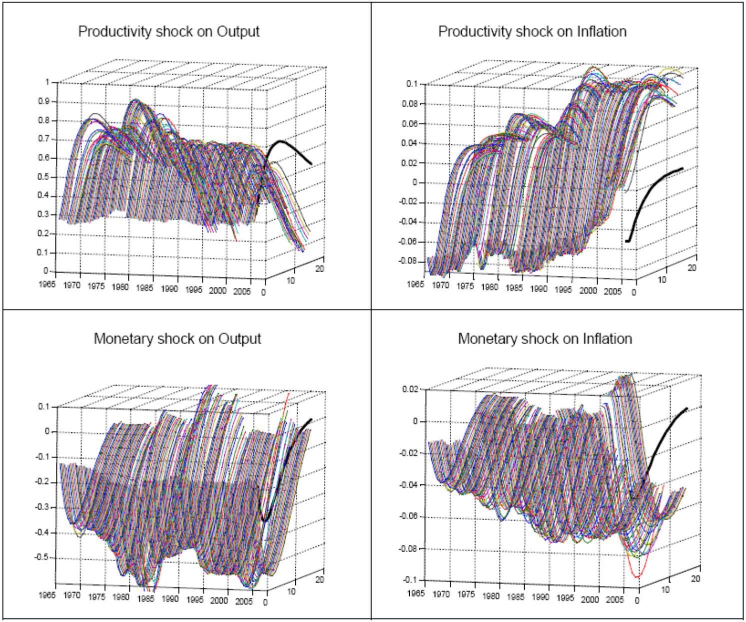
IRFs for VAR learning with model consistent initial beliefs.



IRFs for VAR learning with optimised initial beliefs.



IRFs for VAR learning with pre–sample based initial beliefs.



Estimation results: GSG Learning

- We also tried estimation with the second matrix moments fixed at its initial value R_0, which amounts to a Generalised Stochasic Gradient estimation.

Estimation results: GSG Learning

- We also tried estimation with the second matrix moments fixed at its initial value R_0, which amounts to a Generalised Stochasic Gradient estimation.
- For MSV learning with model–consistent beliefs, posterior mode and the estimated parameters are almost the same as for CG RLS.

Estimation results: GSG Learning

- We also tried estimation with the second matrix moments fixed at its initial value R_0, which amounts to a Generalised Stochastic Gradient estimation.
- For MSV learning with model–consistent beliefs, posterior mode and the estimated parameters are almost the same as for CG RLS.
- For other specifications: MSV learning with pre–sample REE–consistent beliefs, pre–sample regression–based beliefs, or any VAR learning, even finding a posterior mode did not succeed.

Estimation results: GSG Learning

- We also tried estimation with the second matrix moments fixed at its initial value R_0, which amounts to a Generalised Stochastic Gradient estimation.
- For MSV learning with model–consistent beliefs, posterior mode and the estimated parameters are almost the same as for CG RLS.
- For other specifications: MSV learning with pre–sample REE–consistent beliefs, pre–sample regression–based beliefs, or any VAR learning, even finding a posterior mode did not succeed.
- We did not perform estimation with optimised initial beliefs for either MSV or VAR learning yet.

Conclusions

- Is the RE hypothesis restrictive? Yes, relaxing the model–consistent expectations assumption can improve the marginal likelihood of the model, and approach the optimal fit obtained by the DSGE–VAR.

Conclusions

- Is the RE hypothesis restrictive? Yes, relaxing the model–consistent expectations assumption can improve the marginal likelihood of the model, and approach the optimal fit obtained by the DSGE–VAR.
- Can learning solve the misspecification in the DSGE model as detected by the DSGE–VAR approach? The answer is: yes, the best performing models (with optimised initial beliefs) generate IRFs that resemble the results of the DSGE–VAR approach.

Conclusions

- Is the RE hypothesis restrictive? Yes, relaxing the model–consistent expectations assumption can improve the marginal likelihood of the model, and approach the optimal fit obtained by the DSGE–VAR.
- Can learning solve the misspecification in the DSGE model as detected by the DSGE–VAR approach? The answer is: yes, the best performing models (with optimised initial beliefs) generate IRFs that resemble the results of the DSGE–VAR approach.
- The response of inflation to a monetary policy shock becomes more gradual and more persistent. The response of inflation to a productivity shock is immediate and very short–lived.

Conclusions

- Is the RE hypothesis restrictive? Yes, relaxing the model–consistent expectations assumption can improve the marginal likelihood of the model, and approach the optimal fit obtained by the DSGE–VAR.
- Can learning solve the misspecification in the DSGE model as detected by the DSGE–VAR approach? The answer is: yes, the best performing models (with optimised initial beliefs) generate IRFs that resemble the results of the DSGE–VAR approach.
- The response of inflation to a monetary policy shock becomes more gradual and more persistent. The response of inflation to a productivity shock is immediate and very short–lived.
- Does learning change the dynamics? Yes, but the estimated structural frictions remain quite robust.

Remaining issues

- Estimating models with higher gains is computationally difficult. We need to develop more robust estimation approaches.

Remaining issues

- Estimating models with higher gains is computationally difficult. We need to develop more robust estimation approaches.
- What is the relative contribution of the specific initial beliefs versus the learning updating process to the improved fit?

Remaining issues

- Estimating models with higher gains is computationally difficult. We need to develop more robust estimation approaches.
- What is the relative contribution of the specific initial beliefs versus the learning updating process to the improved fit?
- We need to test alternative information sets in the belief regression.

Remaining issues

- Estimating models with higher gains is computationally difficult. We need to develop more robust estimation approaches.
- What is the relative contribution of the specific initial beliefs versus the learning updating process to the improved fit?
- We need to test alternative information sets in the belief regression.
- We need to evaluate alternative, and more efficient learning mechanism like the Kalman filter based approach.

Why is VAR learning different?

- The model can be represented as

$$y_t = \beta E_t y_{t+1} + \delta y_{t-1} + \kappa w_t,$$
$$w_t = \rho w_{t-1} + \epsilon_t.$$

Why is VAR learning different?

- The model can be represented as

$$y_t = \beta E_t y_{t+1} + \delta y_{t-1} + \kappa w_t,$$
$$w_t = \rho w_{t-1} + \epsilon_t.$$

- Instead of the Perceived Law of Motion (PLM) given by

$$y_t = b y_{t-1} + c w_t$$

(MSV solution consistent beliefs), the agents use

$$y_t = b y^o_{t-1} = b H y_{t-1}.$$

Why is VAR learning different? (Cont)

The Actual Law of Motion (ALM), instead of

$$y_t = (I - \beta b)^{-1} \delta y_{t-1} + (I - \beta b)^{-1} (\beta c \rho + \kappa) w_t,$$

for MSV solution consistent beliefs, becomes

$$y_t = (I - \beta H b)^{-1} \delta y_{t-1} + (I - \beta H b)^{-1} \kappa w_t = T_y y_{t-1} + T_w w_t.$$

Why is VAR learning different? (Cont)

- Value of b that equates T_y and bH, even if it exists, does not set T_w to zero.

Why is VAR learning different? (Cont)

- Value of b that equates T_y and bH, even if it exists, does not set T_w to zero.
- The updating equations become

$$
\begin{aligned}
b_t &= b_{t-1} + gR_t^{-1} z_{t-1} (y_t - b_{t-1}^T z_{t-1})^T \\
&= b_{t-1} + gR_t^{-1} z_{t-1} (T_y y_{t-1} + T_w w_t - b_{t-1}^T z_{t-1})^T \\
&= b_{t-1} + gR_t^{-1} z_{t-1} (y_{t-1}^T, w_t^T)(T_y^T, T_w^T)^T - gR_t^{-1} z_{t-1} z_{t-1}^T b_{t-1}.
\end{aligned}
$$

Why is VAR learning different? (Cont)

- After taking time limits and expectations, the E-stability ODE becomes

$$\frac{db}{dt} = R^{-1}E[z_{t-1}y_{t-1}^{T}]T_y + R^{-1}E[z_{t-1}w_t^{T}]T_w - b,$$
$$\frac{dR}{dt} = E[z_{t-1}z_{t-1}^{T}] - R.$$

Why is VAR learning different? (Cont)

- After taking time limits and expectations, the E-stability ODE becomes

$$\frac{db}{dt} = R^{-1}E[z_{t-1}y_{t-1}^T]T_y + R^{-1}E[z_{t-1}w_t^T]T_w - b,$$
$$\frac{dR}{dt} = E[z_{t-1}z_{t-1}^T] - R.$$

- The equilibrium is a fixed point of this system of equations.

Why is VAR learning different? (Cont)

- After taking time limits and expectations, the E-stability ODE becomes

$$\frac{db}{dt} = R^{-1}E[z_{t-1}y_{t-1}^T]T_y + R^{-1}E[z_{t-1}w_t^T]T_w - b,$$
$$\frac{dR}{dt} = E[z_{t-1}z_{t-1}^T] - R.$$

- The equilibrium is a fixed point of this system of equations.
- However, expectations are themselves complicated functions of b: for example,

$$E[z_{t-1}w_t^T] = HE[y_{t-1}w_{t-1}^T]\rho^T = HM_{yw}\rho^T,$$
$$M_{yw} = T_y M_{yw}\rho^T + T_w E[w_{t-1}w_{t-1}^T],$$

Why is VAR learning different? (Cont)

- Thus, VAR learning probably changes the variance-covariance matrix of the endogenous model variables. The equilibrium is a complicated restrictive perceptions equilibrium.

Why is VAR learning different? (Cont)

- Thus, VAR learning probably changes the variance-covariance matrix of the endogenous model variables. The equilibrium is a complicated restrictive perceptions equilibrium.
- There is no guarantee that the 'REE model consistent' VAR beliefs are close to the equilibrium under VAR learning.

Why is VAR learning different? (Cont)

- Thus, VAR learning probably changes the variance-covariance matrix of the endogenous model variables. The equilibrium is a complicated restrictive perceptions equilibrium.
- There is no guarantee that the 'REE model consistent' VAR beliefs are close to the equilibrium under VAR learning.
- This might explain the simulation puzzle, when for $g = 0.01$ transitional dynamics exhibit more volatility than the permanent one. With MSV, volatility changes in the opposite direction, as we are guaranteed to start from the exactly correct beliefs.

Why is VAR learning different? (Cont)

- Thus, VAR learning probably changes the variance-covariance matrix of the endogenous model variables. The equilibrium is a complicated restrictive perceptions equilibrium.
- There is no guarantee that the 'REE model consistent' VAR beliefs are close to the equilibrium under VAR learning.
- This might explain the simulation puzzle, when for $g = 0.01$ transitional dynamics exhibit more volatility than the permanent one. With MSV, volatility changes in the opposite direction, as we are guaranteed to start from the exactly correct beliefs.
- How to find equilibrium – simulate E–stability ODE? Is it E–stable?