Estimating a medium-scale DSGE model with expectations based on small forecasting models

Sergey Slobodyan (CERGE–EI) and Raf Wouters (NBB)

March 30-31, 2009
"10th EABCN Workshop on Uncertainty over the Business Cycle", ECB, Frankfurt

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?
- Milani (2004), Orphanides & Williams (2003), Preston & Eusepi (2008) 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) 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 endogenous model variables. They believe that forecasting models' coefficients are autoregressive processes (with a constant), and use Kalman filter to estimate these coefficients.
- Slobodyan & Wouters (2007) used constant gain RLS learning, and a large set of right–hand–side variables in the forecasting equations. In this setting, learning hardly influences model dynamics. However, restricting the information available to the agents improves the model fit and produces IRFs that match better with those from the best–fitting DSGE–VAR model.

KF Learning: Motivation

- Sargent and Williams (2005) showed that even if Kalman filter and constant gain learning are asymptotically equivalent on average, their transitory behavior may differ a lot. In particular, Kalman filter learning tends to result in much faster adjustment of agents' beliefs.
- With faster adjustment of beliefs, we might be able to understand better whether the initial beliefs or time–varying coefficients matter more for the improved model fit.
- Expecting the agents to take too many variables into account is unrealistic. Therefore, we study what happens if forecasts are based on *small* models, much smaller than those implied by the REE solution.
- One can be never sure that a particular model is the best. Therefore, we allow the agents to run a set of forecasting models and create combined forecasts taking past performance into account, using Bayesian Model Averaging techniques.

Main results of the paper

- Relaxing the model–consistent expectations assumption improves the marginal likelihood of the model significantly.
- Relative to the DSGE model under rational expectations, models with learning are estimated to have lower persistence of shocks (price and wage markup shocks become almost *i.i.d.)*.
- Inflation persistence implied by learning agents' beliefs is consistent with results in the literature: approaching random walk in late 70es, falling persistence after mid–80es.
- Time–varying beliefs reproduce the Great Inflation and the Great Moderation in inflation, but not in output growth.
- The learning dynamics are also consistent with the observed flattening of the Phillips curve.

The REE model of Smets and Wouters (2007)

- Contains three types of agents: households (consume, work, set wages, invest), firms (hire labour and capital, produce goods, set prices), and the central bank (sets short–term interest rate).
- Contains a relatively large number of real and nominal frictions (CEE 2005): calvo price and wage setting with indexation, endogenous mark ups, habit, investment adjustment costs, variable capital utilisation, fixed costs
- Contains a relatively large set of structural shocks: supply, demand, cost-push and monetary policy shocks are all modelled as persistent AR(1) or ARMA(1,1) processes

Smets and Wouters (2007): Monetary Policy

- The monetary authorities follow a generalised Taylor rule: the interest rate adjusts gradually in response to inflation and the output gap, defined as the difference between actual and potential output implied by the underlying TFP process:

$$\begin{aligned}\widehat{R}_t &= \rho_R \widehat{R}_{t-1} + (1-\rho_R)(r_\pi \widehat{\pi}_t + r_y(\widehat{y}_t - \Phi \widehat{A}_t) \\ &\quad + r_{\Delta y}(\widehat{y}_t - \widehat{y}_{t-1} - \Phi(\widehat{A}_t - \widehat{A}_{t-1})) + r_t\end{aligned}$$

- The monetary policy equation is the only difference with Smets & Wouters (2007), where the output gap was defined as the difference between the sticky price and flexible economy output. As a result, there is no flexible economy block in this model.

Smets and Wouters (2007): Measurement Equations

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

$$\begin{bmatrix} \Delta \ln(GDP_t) \\ \Delta \ln(CONS_t) \\ \Delta \ln(INV_t) \\ \Delta \ln(W/P_t) \\ \ln(HOURS_t) \\ \Delta \ln(P_t) \\ FedFunds_t \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \end{bmatrix} + \begin{bmatrix} \widehat{y}_t - \widehat{y}_{t-1} \\ \widehat{c}_t - \widehat{c}_{t-1} \\ \widehat{i}_t - \widehat{i}_{t-1} \\ \widehat{w}_t - \widehat{w}_{t-1} \\ \widehat{l}_t \\ \widehat{\pi}_t \\ \widehat{R}_t \end{bmatrix}$$

Model Solution under RE

- 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 the vector of endogenous variables, and w_t the vector of exogenous processes including the moving average innovations.

- The RE solution of this system is:

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

KF Learning: Set–up

- Under learning, the agents forecast the future value of the lead variables with a linear function of constants and a small set of endogenous variables:

$$y_t^f = \beta_t^T X_{t-1} + \eta_t.$$

- Five alternative small forecasting models are considered: AR(1), AR(2), AR(1)+π, AR(1)+π+R, AR(1)+π+R+y. These five forecasting models are combine with Equal Weights or with BIC Weights.
- The agents believe that the coefficients of these forecasting models follow a vector autoregressive process:

$$vec\left(\beta_t - \overline{\beta}\right) = F \cdot vec\left(\beta_{t-1} - \overline{\beta}\right) + v_t,$$

where F is a diagonal matrix with $\rho \leq 1$ on the main diagonal. $\overline{\beta}$ corresponds with the initial belief.

KF Learning: Set–up (Cont)

- Each model can be written in SURE–like form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}^T \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix},$$

$$y^f = \beta^T X + U,$$

$$\Sigma = E\left[UU^T\right].$$

- The efficient estimator is given by:

$$\widehat{\beta}_{GLS} = \left(X^T\Sigma^{-1}X\right)^{-1} X^T\Sigma^{-1}Y,$$

$$E\left[\widehat{\beta}_{GLS}\widehat{\beta}_{GLS}^T\right] = \left(X^T\Sigma^{-1}X\right)^{-1}.$$

KF Learning: Set–up (Cont)

- With this notation, the Kalman filter equations are given as

$$\beta_{t|t} = \beta_{t|t-1} + P_{t|t-1} X_{t-1} \left[\Sigma + X_{t-1}^{T} P_{t|t-1} X_{t-1}\right]^{-1} \left(y_t^f - \beta_{t|t-1}^{T} X_{t-1}\right)$$

$$\textit{with } (\beta_{t+1|t} - \overline{\beta}) = F \cdot (\beta_{t|t} - \overline{\beta}).$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} X_{t-1} \left[\Sigma + X_{t-1}^{T} P_{t|t-1} X_{t-1}\right]^{-1} X_{t-1}^{T} P_{t|t-1},$$

$$\textit{with } P_{t+1|t} = F \cdot P_{t|t} \cdot F^{T} + V.$$

- Thus obtained beliefs are aggregated over the five models and used to substitute $E_t y_{t+1}$ *by* $\beta_{t+1|t}^{T,aggr} X_t$.This results in a time-varying and purely backward–looking model:

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

KF Learning: initialisation

- β_0 and Σ are derived from the REE at the estimated parameters:

$$\beta_0 = E[XX^T]^{-1}E[y^f X^T].$$

$$\Sigma = E\left[UU^T\right] = E\left[\left(y^f - \beta^T X\right)\left(y^f - \beta^T X\right)^T\right],$$

- Following Sargent and Williams (2005), we assume:

$$\begin{aligned} P_0 &= \gamma \cdot \left(X^T \Sigma^{-1} X\right)^{-1}, \\ V &= \sigma \cdot \left(X^T \Sigma^{-1} X\right)^{-1}, \end{aligned}$$

- Learning is determined by ρ, γ and σ.
- The robustness to presample data is also evaluated.

Results: Model comparison in terms of Marginal Likelihood

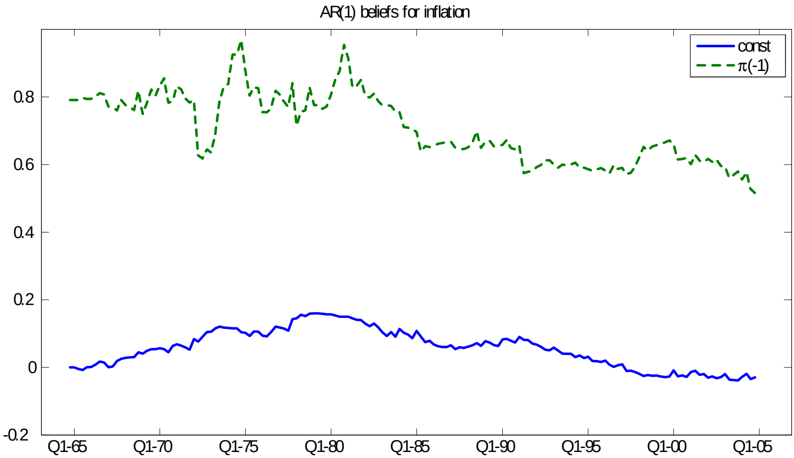
REE model	-926
KF Learning with small models: same sample for beliefs and models estimation	
5 models, BIC selection (γ, σ estimated, ρ fixed)	-917
5 models, equal weights (γ, σ estimated, ρ fixed)	-910
5 models, BIC selection (only ρ estimated)	-911
5 models, equal weights (only ρ estimated)	-909
KF Learning with small models: **longer sample for beliefs** estimation than for model estimation	
5 models, BIC selection (γ, σ estimated, ρ fixed)	-916
5 models, equal weights (γ, σ estimated, ρ fixed)	-910
5 models, BIC selection (only ρ estimated)	-910
5 models, equal weights (only ρ estimated)	-909
No learning, constant beliefs	
5 models, constant beliefs from BIC selection	-920
5 models, constant beliefs from EW combination	-916

Results: Model comparison in terms of estimated parameters

	φ	λ	ξ_w	ι_w	ξ_p	ι_p	ρ_r	ρ_p	μ_p	ρ_w	μ_w
REE model: no flex gap	5.63	0.77	0.71	0.59	0.70	0.22	0.84	0.87	0.73	0.97	0.88
KF: same sample beliefs											
5 models, BIC selection	4.56	0.75	0.76	0.36	0.60	0.23	0.90	0.69	0.57	0.59	0.48
5 models, equal weights	3.17	0.79	0.74	0.38	0.60	0.29	0.90	0.57	0.53	0.65	0.45
5 models, BIC, ρ est	4.63	0.76	0.79	0.34	0.64	0.25	0.88	0.48	0.48	0.56	0.46
5 models, EW, ρ est	3.73	0.79	0.77	0.38	0.64	0.24	0.90	0.48	0.49	0.62	0.46
KF: long sample beliefs											
5 models, BIC selection	4.04	0.78	0.75	0.37	0.60	0.20	0.90	0.70	0.60	0.68	0.55
5 models, equal weights	2.91	0.82	0.73	0.39	0.59	0.25	0.91	0.59	0.52	0.69	0.48
5 models, BIC, ρ est	4.37	0.75	0.77	0.37	0.64	0.22	0.89	0.55	0.54	0.68	0.58
5 models, EW, ρ est	3.85	0.78	0.77	0.40	0.65	0.19	0.89	0.51	0.53	0.69	0.51
Constant beliefs											
BIC initial beliefs	4.84	0.66	0.67	0.51	0.53	0.12	0.88	0.96	0.60	0.96	0.78
EW initial beliefs	4.93	0.77	0.69	0.38	0.60	0.70	0.89	0.25	0.42	0.96	0.72

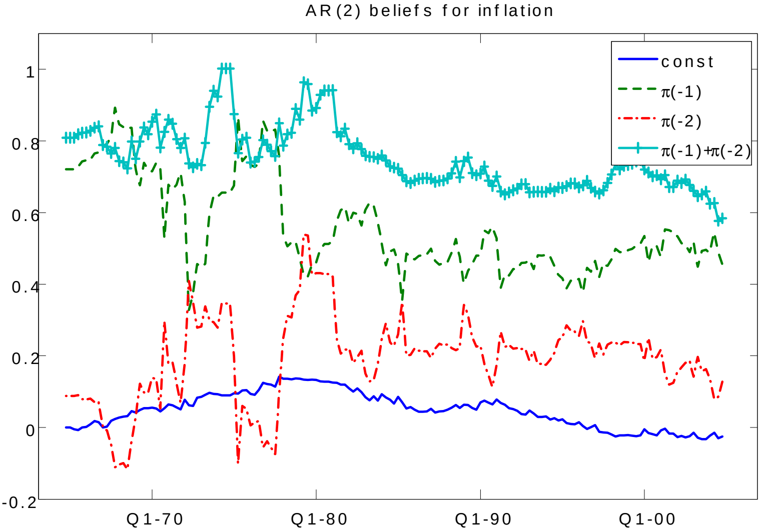
Results: Beliefs

Typical behavior of AR(1) beliefs about inflation.



Results: Beliefs

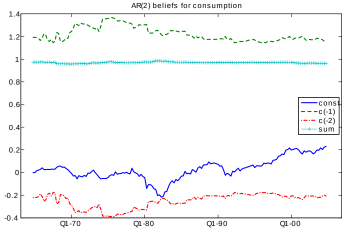
Typical behavior of AR(2) beliefs about inflation.



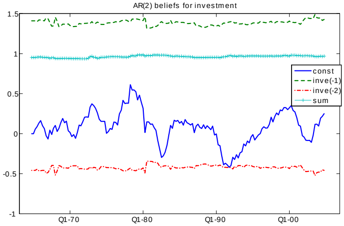
Results: Beliefs

Typical behavior of AR(2) beliefs for other variables.

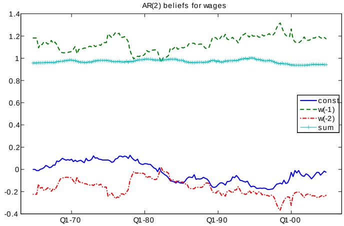
Consumption



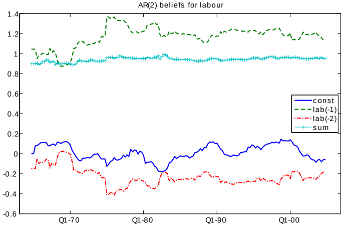
Investment



Real wage

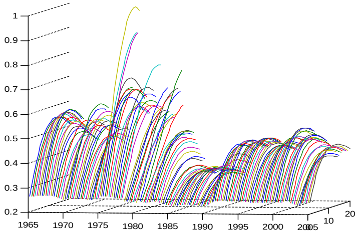


Labour

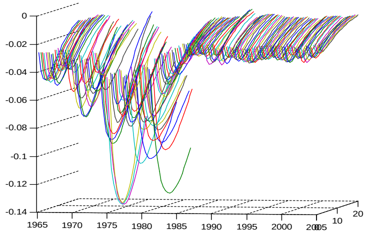


Results: IRF (EW learning model)

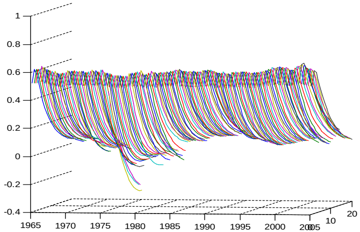
Productivity shock on output



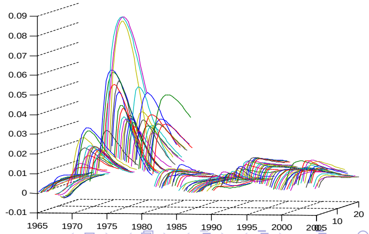
Productivity shock on inflation



Risk premium shock on output

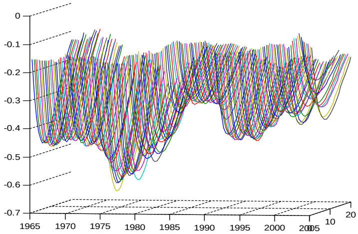


Risk premium shock on inflation

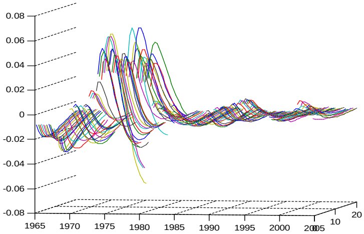


Results: IRF (EW learning model)

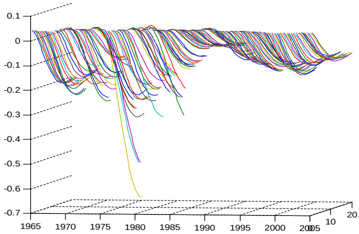
Monetary policy shock on output



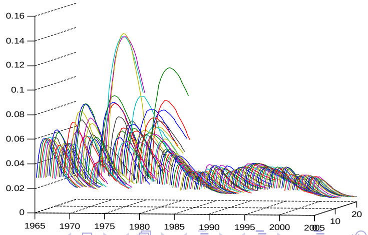
Monetary policy shock on inflation



Wage mark up shock on output



Wage mark up shock on inflation



Results: Great Moderation

- Did we replicate the Great Moderation? Yes for inflation, No for output growth.

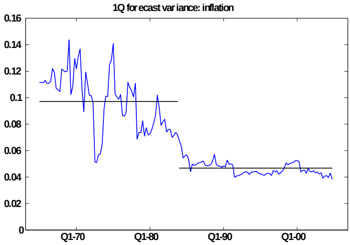
Mean and volatility of inflation and output growth				
	Data		EW Model	
	Before and after 1984		Before and after 1984	
Inflation				
mean	1.43	0.59	1.12	0.63
stdev	0.60	0.24	0.59	0.30
Growth				
mean	0.38	0.50	0.36	0.52
stdev	1.12	0.54	0.95	0.93

These results are averages over 500 draws from the posterior distribution. For each draw, we calculate the corresponding μ_t, T_t and R_t over the observation period. Given these time-varying matrixes, we simulate artificial data for inflation and output growth and calculate the mean and st.dev over the two sub-periods.

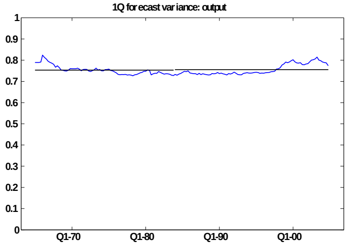
Time varying variance implied by the learning model (EW)

1Q forecast variance

Inflation

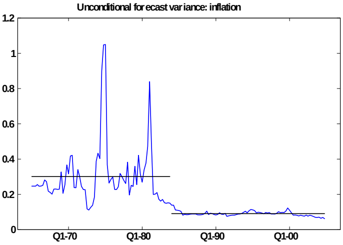


Output level

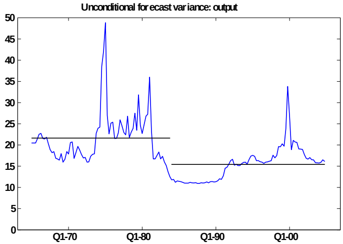


Unconditional forecast variance

Inflation



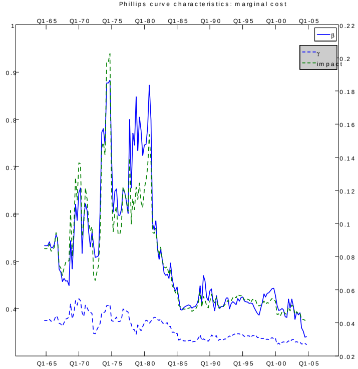
Output level



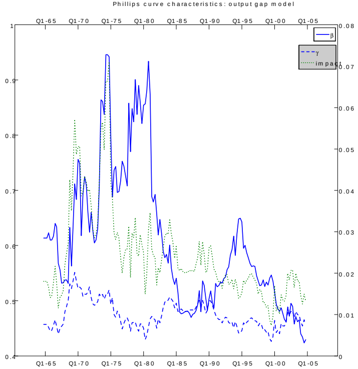
Results: Phillips Curve Flattening?

- What about the slope of the Phillips curve implied by learning dynamics?
  Some evidence on flattening towards 90–es and later.
  One–year impact on inflation of a sustained change in gap measure is driven mostly by inflation persistence.

$$\pi_t = \beta\pi_{t-1} + \gamma mc_t$$



$$\pi_t = \beta\pi_{t-1} + \gamma ygap_t$$



Conclusion

- The model, estimated under Kalman filter learning with multiple small forecasting models, produces significantly better marginal likelihood than under RE.
- Restricted information set and time variability matter for the better likelihood to an approximately equal degree.
- Equal weighting of forecasting models works better than BIC selection.
- Estimated price and wage mark–up shocks are essentially *i.i.d.*
- Only a few structural rigidity parameters change systematically.
- Beliefs about inflation replicate earlier results in the literature.

Conclusion (Cont)

- Responses of inflation and output are stronger and more persistent when inflation itself is perceived to be persistent.
- IRFs depend on the set of models used. Behavior similar to the "price puzzle" for MP shock is possible.
- We can explain almost all decline in standard deviation of inflation (Great Moderation) but not in output growth.
- Learning agents would have perceived a flattening Phillips curve towards 90es and later.
- One–year ahead inflation expectations are better under adaptive learning than under RE after 1985.

KF Learning: Set–up (Cont)

- Time–dependent matrices replace DYNARE–produced μ, R, and T, and are then used in the Kalman filtering step used to produce the model likelihood.
- Updating of the beliefs at any t depends on the data (best estimates of the state and the lead vector at respectively time $t-1$ and t) and on the initial beliefs.
- Need to specify β_0, P_0, Σ, and V.
  - V is Var–Covar matrix of the state equations errors.
  - Σ is Var–Covar matrix of the measurement errors.

KF Learning: Estimation

- Following Sargent and Williams (2005), we assume

$$
\begin{aligned}
P_0 &= \gamma \cdot \left(X^T \Sigma^{-1} X\right)^{-1}, \\
V &= \sigma \cdot \left(X^T \Sigma^{-1} X\right)^{-1},
\end{aligned}
$$

  and estimate γ and σ separately.
  - Consistently with Sargent, Williams, and Zha (2006), we find that the estimated σ is much larger than γ^2. If $\sigma = \gamma^2$, Kalman filter learning is asymptotically equivalent to constant gain learning.
- We started our estimation runs with random walk beliefs ($\rho = 1$). However, it turns out that fixing ρ at a value such as 0.995 improves the estimated posterior mode (Should we use diffuse priors for the beliefs Kalman filter?)
- Most successful approach is to fix γ and σ at "reasonable" values from random walk beliefs MCMC, and estimate only ρ.

KF Learning: Model Selection

- We allow the agents to use several models at the same time, and consider alternative model averaging techniques:
  - EW: fixed model weights equal to $\frac{1}{N}$, where N is the number of models
  - BIC: bayesian model averaging with agents tracking the forecasting performance of the individual models

$$B_{i,t} = t \cdot \ln \det \left(\frac{1}{t} \sum_{i=1}^{t} u_i u_i^T \right) + \kappa_i \cdot \ln t,$$

  where κ_i is number of degrees of freedom in forecasting model M_i, and u_i the i–th model forecasting errors. This is a generalization of the sum of squared errors adjusted for degrees of freedom using Bayesian Information Criterion (BIC).

- These weights are used to form an aggregate forecast (in our setup, just the aggregate vector β_t^{aggr}).

KF Learning: Forecasting Models

- So far, we used 5 models (all the models include a constant):
  1. AR(1): every forward–looking variable is predicted based on its own lagged value and a constant;
  2. AR(2): every forward–looking variable is predicted based on two own lags and a constant;
  3. AR(1) + 1: in addition to own lag and a constant, inflation is added to the RHS of every forecating equation;
  4. AR(1) + 2: in addition to own lag and a constant, interest rate and inflation are added to the RHS of every forecating equation;
  5. AR(1) + 3: in addition to own lag and a constant, interest rate, inflation, and output are added to the RHS of every forecasting equation.

KF Learning: Projection Facility

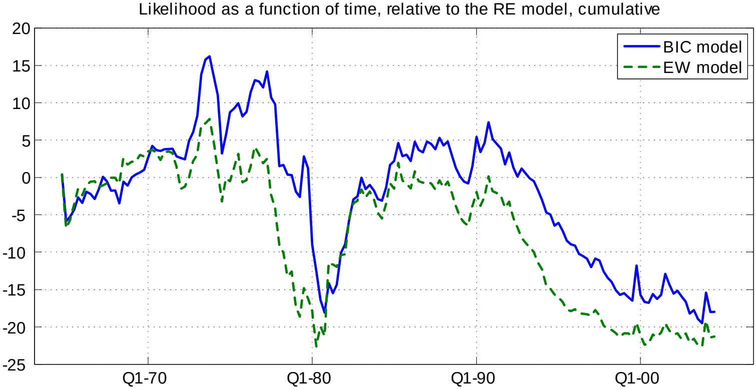
- During updating, the transition matrix T_t (derived as if a single forecasting model at a time were used) is restricted to the stable domain by a version of a projection facility: if the largest eigenvalue of T_t is outside of the unit circle, we keep last period β.
  - A standard projection facility (checking roots of the forecasting Vector Autoregression) cannot be implemented generally, 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.
  - Discontinuous adjustment of the beliefs with many small models is found to work better than a continuous adjustment of the T matrix.

Results: Likelihood (Cont)

- Modified Smets & Wouters (2007) model, estimated under Kalman filter learning with multiple small forecasting models, produces significantly better marginal likelihood than RE estimation.
  - Learning models tend to work better than RE in 2^{nd} half of 70es and after 1990.
- Both the restricted information set, available to the learning agents, and the time variability introduced by learning, contribute to the better likelihood to an approximately equal degree (as measured by the improvement in the marginal likelihood).
- Disregarding past forecasting performance and weighting the forecasting models equally works better than the BIC selection. This might be related to the very high correlation between expectational errors of the different models.

Better likelihood (negative in the graph) for the learning models is mostly accumulated in the post–1990 period.

Figure 1. Cumulative likelihood for the best performing BIC and EW models over time, relative to the RE model.



Results: Parameters

- Estimated parameters point to significantly lower persistence of price and wage mark–up shocks (become *i.i.d.* shocks).
  - DSGE–VAR estimation in Slobodyan & Wouters (2007) also showed significant decline of these autocorrelations and MA terms, especially for price mark–up shock that was *i.i.d.*
- There are some systematic changes in the structural rigidity parameters: decline in wage indexation and investment adjustment costs. Wage stickiness and gradualism of the monetary policy rule decrease.
- There are no significant changes in other model parameters.

Results: *i.i.d.* shocks

- Guided by the previous results, we performed estimation of the model assuming that both price and wage mark–up shocks are *i.i.d.* (Same sample for beliefs and model estimation, equal weights, only ρ estimated).
- Marginal likelihood is the same as for the estimation of this type: -909.
- Structural rigidity parameters are essentially the same.
- The same outcome if, in addition, monetary policy and risk premium shocks are also forced to be *i.i.d.*

Results: Beliefs (Cont)

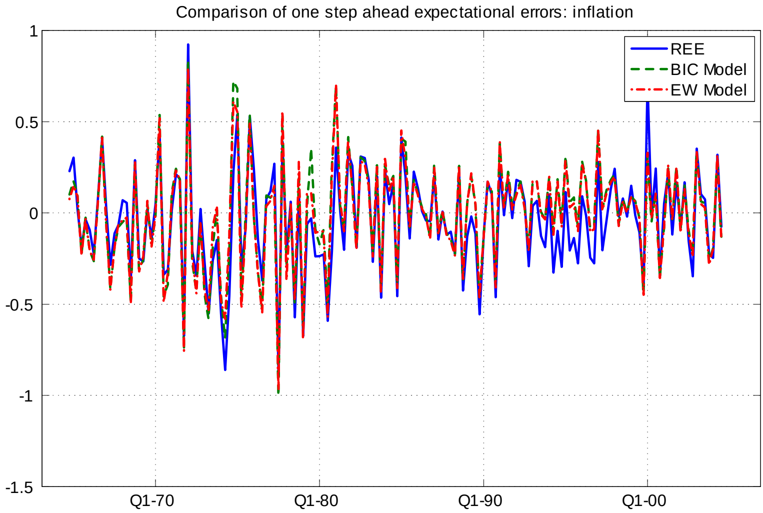
- Beliefs about all forward–looking variables vary a lot. Learning about constants is important as there is large variability.
- Implied inflation persistence exhibits pattern similar to that of Milani (2004), Cogley, Primicery & Sargent (2007) or Beechey & Osterholm (2007): a double peak in 70es and a gradual decline later.
- Pattern of beliefs about consumption, investment, real wage, and labour suggest that the best model for them could be AR(1) in first differences.

Results: IRFs

- Response of inflation depends a lot on the perceived inflation persistence. As a result, the IRFs exhibit stronger and more persistent responses in 70es.
- Similar but less pronounced effect exists for output.
- The response of inflation to monetary policy shock was positive on impact in 70es and turned negative only after several quarters. This might be related to correlation between interest rates and expected future inflation during this time period (agents could interpret unexpectedly higher interest rate as a signal of increased future inflation).
    - IRFs depend somewhat on the set of forecasting model that is used. The "price puzzle" above is not observed if only AR(2) model is used to form expectations.
    - In AR(2) case inflation responds less to monetary policy shock on impact, but the response is more persistent, relative to the RE. This is similar to Slobodyan & Wouters (2007).

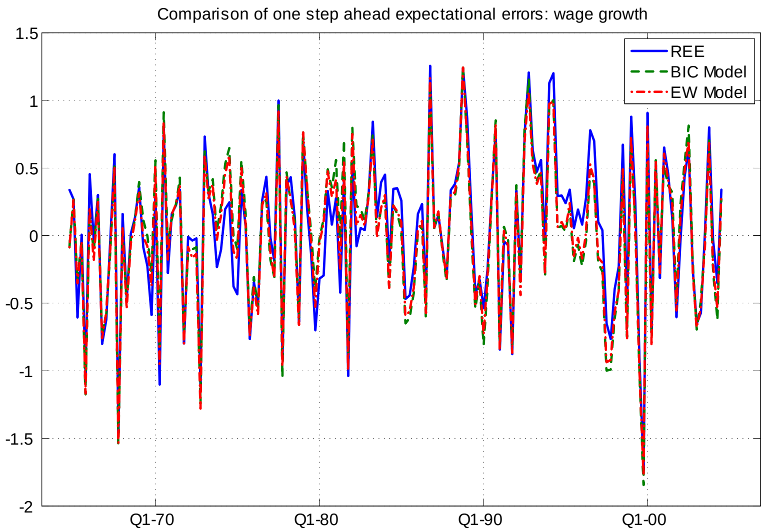
One–step ahead predictions for inflation are mostly similar, but in mid–90es RE model underpredicts inflation, while learning models are doing better.

Figure 2: One step ahead prediction performance for inflation.



In mid-90es, RE model overpredicts wage growth relative to the learning models.

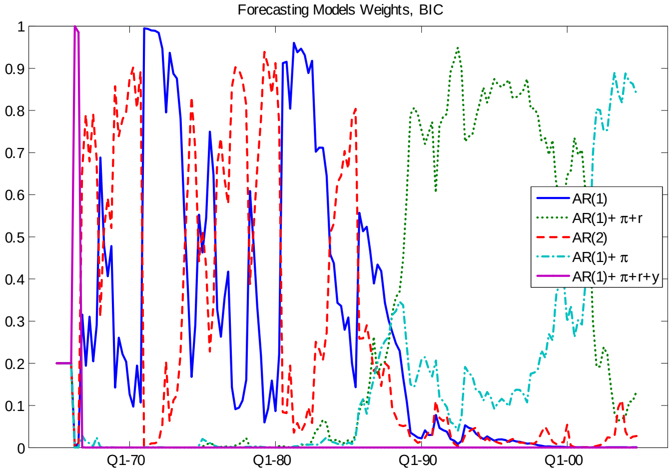
Figure 3: One period ahead prediction performance for wage growth.



(Preliminary) Results: Model weights

Typical behavior of model weights.

Figure 5: Weights of the individual model, BIC



Results: Model weights (Cont)

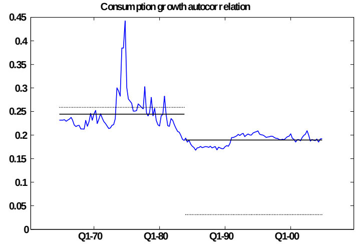
- In the first half of the estimation period (until mid–80es) simple AR(1) and AR(2) forecasting models dominate, while later more complicated models, that use additional information such as inflation and the interest rate, become prevalent.
- Forecasting models with more than 3 variables on the RHS are almost never selected by the BIC criterion.
- With the equal weighting scheme, the aggregate model is competitive with the best single forecasting model, but it is much worse under BIC selection.

No Great Moderation in Output Growth

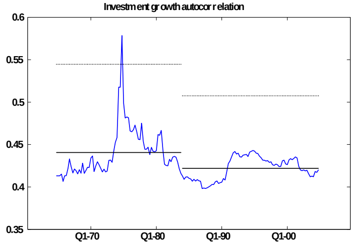
- Do we fail to model the transmission mechanism properly?
- Unconditional autocorrelations in real variables' growth rates are implied to be much more stable (across subperiods) in the model than in the data.
- The model cannot replicate relatively high pre–84 output growth autocorrelation, which makes its chances to generate Great Moderation in output growth slim.

Time varying autocorrelations implied by the learning model (EW)

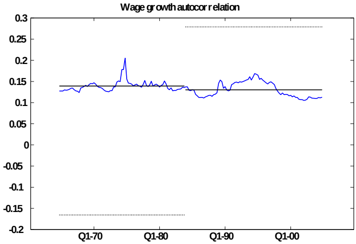
Consumption growth



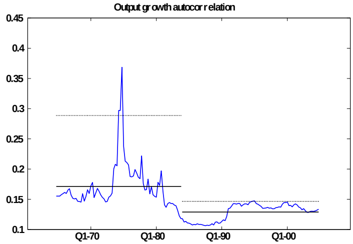
Investment growth



Wage growth



Output growth

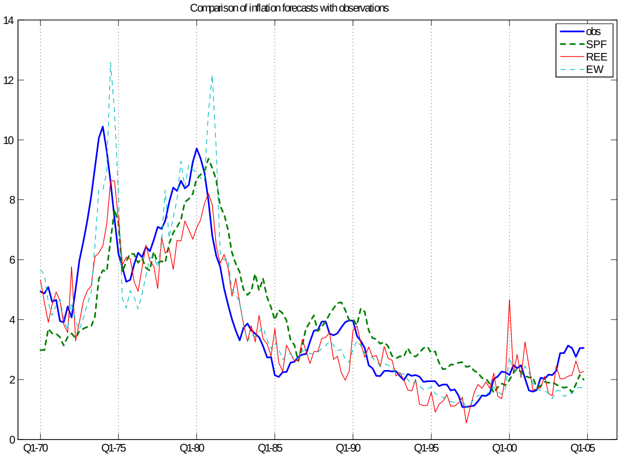


KF Learning: Model Selection (Cont)

- The weight of a model i at time t is then proportional to $\exp\left\{-\frac{1}{2}B_{i,t}\right\}$.
- These weights are used to form an aggregate forecast (in our setup, just the aggregate vector β_t^{aggr}), and proceed to calculating new matrices T and R as before in the constant gain learning method.
- For linear models with normal priors and normal errors, this procedure is asymptotically equivalent to weighting the models using posterior odds ratio.
- We also tried Akaike penalty for degrees of freedom (AIK), as well as fixing the model weights at $\frac{1}{N}$, where N is the number of models used.

Results: 1Y Forecasts

What about longer forecasts? We compare model predictions with SPF forecasts of inflation in the next year.



Results: 1Y Forecasts (Cont)

- Adaptive expectations are usually very close to the RE expectations, except for inflation peak periods when they were higher than RE and SPF but much closer to the actual inflation.
- Both disinflations of 1975 and 80es are missed by the adaptive expectations, but again RE and adaptive expectations are tracking closer to the observed inflation that to the SPF.
- Adaptive expectations are less volatille than RE after 1985.
- Model-based expectations predict ex-post realized inflation better than the SPF: standard deviation of the difference is 1.47, 1.08, and 1.16 for SPF, RE, and EW, respectively, over the whole sample. After 1985Q1, adaptive expectations are better: 0.77, 0.70, and 0.57, respectively.