Dealing with misspecification in structural macroeconometric models

Fabio Canova, Norwegian Business School and CEPR

Christian Matthes, Richmond Fed

January 2018

Question

- Want to measure the **marginal propensity to consume** (MPC).

- Take a off-the-shelf permanent-income, life-cycle model, solve it, and derive implications for MPC.

- With quadratic preferences, constant interest rate, permanent and transitory exogenous labour income, the decision rules are

$$
\begin{aligned}
c_t &= \frac{r}{r+1} a_t + (y_t^P + \frac{r}{1-\rho+r} y_t^T) && (1) \\
a_{t+1} &= (1+r)[a_t - (y_t^T + y_t^P) - c_t] && (2) \\
y_t^T &= \rho y_{t-1}^T + e_t^T && (3) \\
y_t^P &= y_{t-1}^P + e_t^P && (4)
\end{aligned}
$$

where y_t^T is transitory income, y_t^P is permanent income, c_t consumption, a_t asset holdings, $\beta(1+r) = 1$, and $e_t^i \; iid \; (0, \sigma_i^2), \; i = T, P, y_t = y_t^P + y_t^T$.

Estimation of MPC_{y^T} I: neglecting model's restrictions

• Natural experiment: e.g. unexpected tax cut. In US $MPC_{y^T} \approx [0.5 - 0.6]$ (Johnson, et al., 2006; Parker et al., 2013).

• Identify a permanent and a transitory shock in a VAR with (y_t, a_t, c_t). Compute the effect of a transitory shock. $MPC_{y^T} \approx [0.4 - 0.6]$.

- Refinement: if a_t not observable, use a bivariate VAR(k), $k \to \infty$ with (y_t, c_t).

Estimation of MPC_{y^T} II: conditioning on model's restrictions

• Assume all agents face the same ex-post real rate; use moments to measure r (4% a year) and ρ ($\approx 0.6 - 0.7$). Then $MPC_{y^T} \approx [0.05 - 0.10]$.

- Refinement: group data according to consumer characteristics; estimate r, ρ and MPC_{y^T} for each group, take a (weighted) average. Then $MPC_{y^T} \approx [0.10 - 0.15]$ (see Caroll, et al., 2014).

• Write down the likelihood function for (c_t, a_t, y_t), using the model restrictions. Estimate r, ρ. Then $MPC_{y^T} \approx [0.10 - 0.15]$.

• **Why estimates obtained conditioning on the structural model are lower than those obtained using the model only a guidance for the analysis**?

Model is likely to be misspecified.

- The real interest rate is not constant over time.

- Labor income is not exogenous. (Income) uncertainty may matter.

- Preferences may not be quadratic in consumption; they may feature non-separable labor supply decisions. Home production, goods durability, etc. may matter.

- Disregard heterogeneities: some agents may have zero assets (ROT); others may be rich but liquidity constrained (HTM).

- Assets mismeasured.

• Moment-based and VAR-based estimates robust to some form of misspecification, e.g. lack of dynamics, model incompleteness (Cogley and Sbordone, 2010, Kim, 2002).

• Likelihood-based estimates invalid under misspecification.

• Current econometric misspecification literature (Cheng and Liao, 2015; Thryphonides, 2016; Giacomini et al., 2017) does not employ likelihood when a model is misspecified.

• Robustness (Hansen and Sargent, 2008) more concerned in fending off a malevolent nature than reducing estimation biases.

How do you guard yourself against misspecification if you insist in using likelihood methods?

Existing approaches

1) Estimate a general model with potentially missing features. Computationally demanding; identification issues; interpretation problems.

2) Capture misspecification with ad-hoc features. For example, with **habit in consumption** (h) we have

$$
\begin{aligned}
c_t &= \frac{h}{1+r}c_{t-1} + (1 - \frac{h}{1+r})w_t && (5)\\
w_t &= \frac{r}{1+r}[(1+r)a_{t-1} + \sum_{t=\tau}^{\infty}(1+r)^{t-\tau}E_t y_\tau] && (6)\\
y_t &= y_t^P + y_t^T && (7)\\
y_t^T &= \rho y_{t-1}^T + e_t^T && (8)\\
y_t^P &= y_{t-1}^P + e_t^P && (9)
\end{aligned}
$$

• Not all ad-hoc additions work. With **preference shocks**, we have

$$
\begin{aligned}
c_t &= (1 - \frac{1}{k_t} a_t + (y_t^P + \frac{r}{1-\rho+r} y_t^T) && (10)\\
a_{t+1} &= (1+r)(a_t - y_t - c_t) && (11)\\
y_t &= y_t^P + y_t^T && (12)\\
y_t^T &= \rho y_{t-1}^T + e_t^T && (13)\\
y_t^P &= y_{t-1}^P + e_{2t}^P && (14)
\end{aligned}
$$

where $k_t = E[\beta_t(1+r)^2]$. It mimics the presence of a time varying MPC_a. MPC_{y^T} unchanged.

3) Make the shock process more flexible; use AR(p) (Del Negro and Schorfheide, 2009); ARMA(1,1) (Smets and Wouters, 2007); correlated structural shocks (Curdia and Reis, 2010).

4) Add measurement errors to the decision rules (Hansen and Sargent, 1980, Ireland, 2004, etc.).

5) Add wedges to FOC (Chari et al, 2008), margins to the model (Inoue et al, 2016), or shocks to the decision rules (Den Haan and Drechsel 2017).

• Check the relevance of adds-on, via marginal likelihood (ML) comparison.

• Kocherlakota (2007): dangerous to use "fit" to select among misspecified models.

• All approaches condition on one model, but many potential model specifications on the table.

• All approaches neglect that different models may be more or less misspecified in different time periods (e.g. Del Negro et al., 2016).

• Interpretation problems with 3)- 5) when adds-on are serially correlated.

• Alternative: **Composite likelihood approach**, Canova and Matthes (2016).

• Take all relevant specifications, combine likelihoods geometrically, and jointly estimate the parameters for all specifications.

• Can design selection criteria for optimal selection.

• Posterior of model weights measure the extent of model misspecification (can be used as model selection criteria).

• Can be used to measure time varying misspecification.

• Perform inference using geometric combination of models.

Advantages of CL approach

• May reduce misspecification and provide more reliable estimates of parameters common across models.

• Robustifies inference.

• Computationally as easy as Bayesian maximum likelihood (easier, if a two-step approach is used).

• It can be used when models feature different endogenous variables and concern data of different frequencies.

• It has a bunch of side benefits for estimation (see Canova and Matthes, 2016): it helps with identification, it can deal with singularity, large scale models, data of uneven quality, can be used with panel data, etc.

Logic

• When a model is misspecified, information in additional (misspecified) models restricts the range parameter estimates can take. This improves the quality of estimates (location and, possibly, magnitude of credible sets).

- DGP (ARMA(1,1)): $y_t = \rho y_{t-1} + \theta e_{t-1} + e_t,\ e_t \sim (0, \sigma^2)$.

- Estimated model 1 (AR1): $y_t = \rho_1 y_{t-1} + u_t,\ u_t \sim (0, \sigma_u^2)$

- Estimated model 2 (MA1): $y_t = u_t + \theta_1 u_{t-1},\ u_t \sim (0, \sigma_u^2)$.

- Focus on the relationship between $\hat{\sigma}_u^2$ and σ^2 (common parameter).

- Expect upward bias in $\hat{\sigma}_u^2$ because part of the serial correlation of the DGP is disregarded. Can CL reduce the bias?

• Simulate 150 data from DGP. Use T=[101,150] for estimation. Consider:

1) Fixed weights: ω (AR weight) $= 1 - \omega = 0.5$.

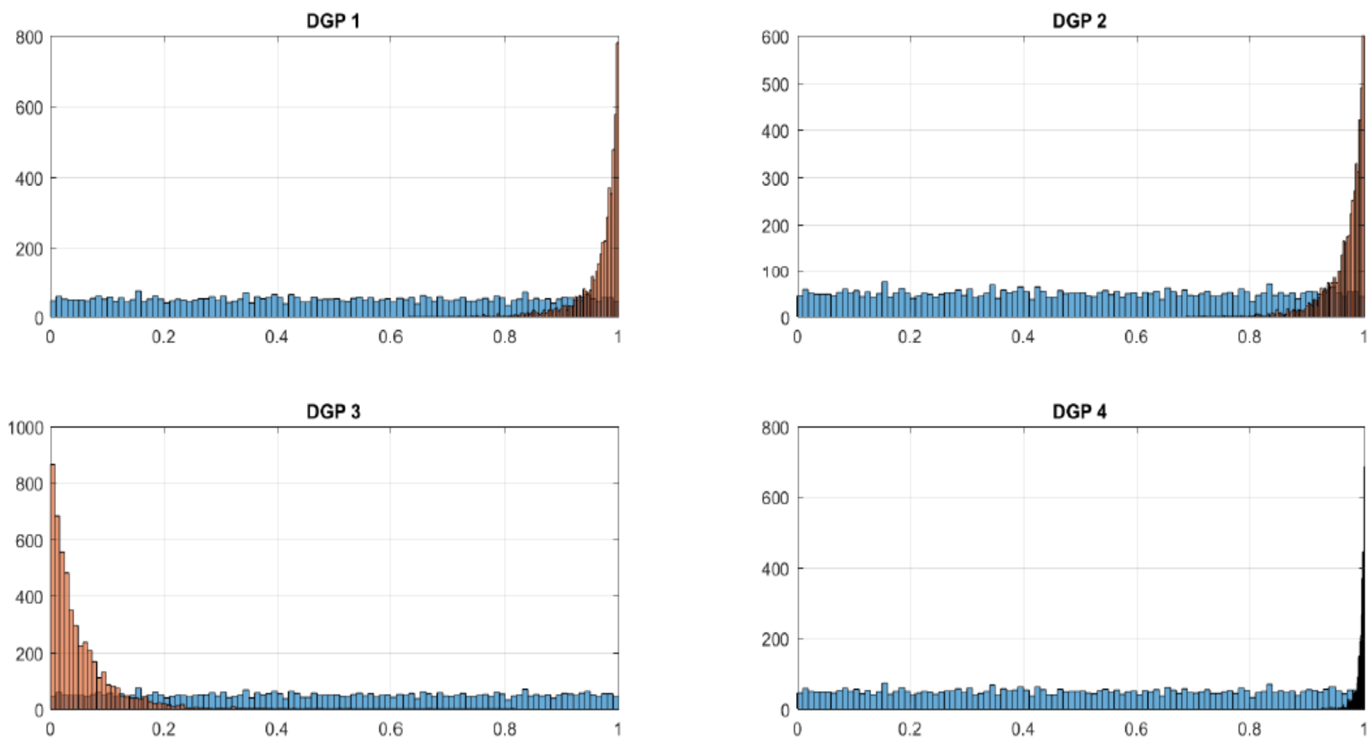
2) Fixed weights: based on relative MSEs in training sample T=[2,100]

3) Random weights. Prior on the weight is Beta with mean 0.5.

Table 1: Estimates of σ_u^2

$y_t = \rho y_{t-1} + \theta e_{t-1} + e_t,\ e_t \sim N(0, \sigma^2)$, T=50					
DGP	AR(1)	MA(1)	CL, Equal weights	CL, MSE weights	CL,Random weights
$\sigma^2 = 0.5, \rho = 0.6, \theta = 0.5$	0.75(0.06)	0.81 (0.07)	0.73 (0.05)	0.70 (0.06)	0.71 (0.05)
$\sigma^2 = 1.0, \rho = 0.6, \theta = 0.5$	1.08(0.07)	1.14 (0.08)	1.07 (0.07)	1.05 (0.07)	1.05 (0.07)
$\sigma^2 = 1.0, \rho = 0.3, \theta = 0.8$	1.14(0.08)	1.05 (0.08)	1.06 (0.07)	0.99 (0.07)	0.98 (0.07)
$\sigma^2 = 1.0, \rho = 0.9, \theta = 0.2$	1.06(0.07)	1.59 (0.10)	1.21 (0.08)	1.03 (0.07)	1.04 (0.07)

Posterior of ω (weight on AR(1))



- What if the DGP is one of the candidate models?

Table 2: Posterior of ω, different sample sizes

	Mode	Mean	Median	Standard deviation
Prior	NA	0.5	0.5	0.288
$y_t = 0.8y_{t-1} + e_t,\ e_t \sim N(0, \sigma^2)$, T=50				
T=50	0.994	0.978	0.985	0.023
T=100	0.997	0.983	0.986	0.018
T=250	0.998	0.990	0.993	0.010
T=500	0.999	0.993	0.995	0.006
$y_t = 0.7e_{t-1} + e_t,\ e_t \sim N(0, \sigma^2)$, T=50				
T=50	0.356	0.468	0.432	0.187
T=100	0.007	0.220	0.147	0.177
T=250	0.003	0.048	0.030	0.050
T=500	0.002	0.034	0.021	0.030

Results

• When the DGP is among the estimated models, the posterior distribution of ω clusters around 1 for that model, as $T \to \infty$.

• When the DGP is NOT among the estimated models, the posterior distribution of ω clusters around the value that minimize the Kullback-Leibner distance between the composite model and the DGP, as $T \to \infty$.

Intuition about CL estimation in misspecified models

• Two misspecified models: A, B; with implications for y_{At} and y_{Bt}, $y_{At} \neq y_{Bt}$.

• Decision rules are:

$$y_{At} = \rho_A y_{At-1} + \sigma_A e_t \tag{15}$$

$$y_{Bt} = \rho_B y_{Bt-1} + \sigma_B u_t \tag{16}$$

e_t, u_t are iid N(0,I);y_{At} and y_{Bt} scalars; samples:T_A and $T_B, T_B \leq T_A$.

• **Suppose** $\rho_B = \delta\rho_A, \sigma_B = \gamma\sigma_A$

- The (normal) log-likelihood functions are

$$\log L_A \propto -T_A \log \sigma_A - \frac{1}{2\sigma_A^2} \sum_{t=1}^{T_A} (y_{At} - \rho_A y_{At-1})^2 \tag{17}$$

$$\log L_B \propto -T_B \log \sigma_B - \frac{1}{2\sigma_B^2} \sum_{t=1}^{T_B} (y_{Bt} - \rho_B y_{Bt-1})^2 \tag{18}$$

- Let weights be $(\omega, 1 - \omega)$, fixed. The composite log-likelihood is:

$$\log CL = \omega \log L_A + (1 - \omega) \log L_B \tag{19}$$

- Suppose we care about $\theta = (\rho_A, \sigma_A)$.

- Maximization of the composite likelihood leads to:

$$\rho_A = (\sum_{t=1}^{T_A} y_{At-1}^2 + \zeta_2 \sum_{t=1}^{T_B} y_{Bt-1}^2)^{-1} (\sum_{t=1}^{T_A} y_{At} y_{At-1} + \zeta_1 \sum_{t=1}^{T_B} y_{Bt} y_{Bt-1}) \tag{20}$$

$$\sigma_A^2 = \frac{1}{\xi} (\sum_{t=1}^{T_A} (y_{At} - \rho_A y_{At-1})^2 + \frac{1-\omega}{\omega\gamma^2} \sum_{t=1}^{T_B} (y_{Bt} - \delta \rho_A y_{Bt-1})^2) \tag{21}$$

where $\zeta_1 = \frac{1-\omega}{\omega} \frac{\delta}{\gamma^2}$, $\zeta_2 = \zeta_1 \delta$; $\xi = (T_A + T_B \frac{1-\omega}{\omega\gamma^2})$ is "effective" sample size.

• Shrinkage estimators for θ. Formulas are same as in i) Least Square problem with uncertain linear restrictions, ii) prior-likelihood approach, iii) DSGE-VAR.

• For θ, model B plays the role of a prior for model A.

• Informational content of model B data for θ measured by $(\gamma, \delta, 1-\omega)$. The larger is γ and the smaller is δ, the lower is model B information.

• **More weight given to data assumed to be generated by a model with higher persistence and lower standard deviation**.

• When constant, ω is the (a-priori) trust in model A information.

- For multiple models, equation (20) is

$$\rho = (\sum_{t=1}^{T_1} y_{1t-1}^2 + \sum_{i=2}^{K} \zeta_{i2} \sum_{t=1}^{T_i} y_{it-1}^2)^{-1} (\sum_{t=1}^{T_1} y_{1t} y_{1t-1} + \sum_{i=2}^{K} \zeta_{i1} \sum_{t=1}^{T_i} y_{it} y_{it-1}) \quad (22)$$

where $\zeta_{i1} = \frac{\omega_i}{\omega_1} \frac{\delta_i}{\gamma_i^2}$, $\zeta_{i2} = \zeta_{i1}\delta_i$.

- **Robustification**: estimates of (ρ, σ^2) forced to be consistent with the restrictions present in all models.

- y_{At} and y_{Bt} may be

- different variables. Can use models with different observables.

- the same variables with different level of aggregation (say, aggregate vs. individual consumption) or in different subsamples (pre and post financial crisis)

- T_A and T_B may

- have different length. Can combine models relevant at different frequencies (e.g. a quarterly and an annual model).

- be two samples for the same variables coming from different cross sectional units.

Difference from what you may know

- Different from **BMA** (e.g. Giacomini, et al., 2017): averaging done using estimates obtained using the restrictions present in each model; $y_{At} \neq y_{Bt}$.

- Different from **ex-post averaging**: common parameters θ are jointly estimated using the restrictions present in each model.

- Different from **finite mixture** (Waggoner and Zha, 2012): y_{At} may be different from y_{Bt} and of different length.

Model selection and model misspecification

- Posterior of ω informs us about model misspecification.

- Can be used for model selection, but bad idea to pick a model if there are data instabilities. Use prediction pools.

Choosing the composite likelihood combination

• How to choose the optimal combination of models entering (both the dimensionality of the pool and the models in the pools)?

• Models not independent. Trade-off between the number of models and composite likelihood gains.

• Let $S = \sum_{k=2}^{K-2} \frac{k!}{r!(k-r)!}$ be an index for the composite combination, allow at least two models in the composite pool, and let $y = y_1 = \ldots = y_S$.

• Under regularity conditions on the prior, (Lv and Liu, 2014):

$$GBIC_{s,CL} \propto -2CL(\theta_{CL}, \eta_{s,CL}, y) + 2dim(\theta_{CL}, \eta_{s,CL}) \log T_s + 2I(H_s, J_s) \tag{23}$$

$$I(H_s, J_s) = \frac{1}{2}(tr(Q_s) - \ln |Q_s| - dim(\psi_s)) \ , \ Q_s = J_s^{-1} H_s$$

• $I(H_s, J_s)$ is the log of the KL divergence between two $dim(\psi_s)$ vectors of normal variables, one with zero mean and covariance J_s (variability matrix) and the other with zero mean and covariance H_s (the sensitivity matrix).

• GBIC: fit, dimensionality, misspecification.

• If composite model $\bar{s}$ is the DGP, $J_{\bar{s}} \approx H_{\bar{s}}$, $I(J_{\bar{s}}, H_{\bar{s}}) \approx 0$, GBIC$=$ BIC.

• When models share the same observables, $I(H_s, J_s)$ measures the misspecification in composite model s.

• Different from ω (it informs us about the relative support of a model in the estimated composite pool).

Prediction pools

- $\tilde{y}_{t+l}$: future values of variables appearing in all models, $l = 1, 2, ...$.

- θ Common parameters, η_i model specific parameters.

- $f(\tilde{y}_{t+l}|y_{it}, \theta, \eta_i)$ = prediction of $\tilde{y}_{t+k}$ made with model i. Let

$$f^{cl}(\tilde{y}_{t+l}|y_{1t}, \dots, y_{Kt}, \theta, \eta_1, \dots, \eta_K, \omega_1, \dots \omega_K) = \prod_{i=1}^{K} f(\tilde{y}_{t+l}|y_{it}, \theta, \eta_i)^{\omega_i} \quad (24)$$

The composite predictive distribution of $\tilde{y}_{t+l}$, given the weights is

$$p(\tilde{y}_{t+l}|y_{1t}, \dots, y_{Kt}, \omega_1, \dots \omega_K) \propto \int f^{cl}(\tilde{y}_{t+l}|y_{1t}, \dots, y_{Kt}, \theta, \eta_1, \dots, \eta_K, \omega_1, \dots, \omega_K) \\ p(\theta, \eta_1, \dots, \eta_K|\omega_1, \dots, \omega_K, y_{1t}, \dots, y_{Kt}) d\theta d\eta_1 \dots d\eta_K \quad (25)$$

Comparison with other pooling devices

• **Linear pooling** (finite mixtures predictive densities, BMA , static pools) (Amisano and Geweke, 2011; Waggoner and Zha, 2012; del Negro et al. 2016).

• **Logarithmic pooling** (CL). Predictive densities generally unimodal and less dispersed than linear pooling; invariant to the arrival of new information (updating the components of the composite likelihood commutes with the pooling operator).

• **Exponential tilting** (ET) Under certain conditions CL produces ET results (see Cover and Thomas, 2006).

Composite impulse responses and counterfactuals

- Same logic.

- Compute responses/ counterfactuals for each model, compute a geometric pool, integrate with respect to the composite posterior of the parameters.

Measuring MPC_y^T (preliminary!)

• **BASIC**:Quadratic preferences, constant real rate, $\beta(1+r)=1$, exogenous permanent (RW) and AR(1) transitory income.

• **PRECAUTIONARY**: Exponential preferences, constant real rate, $\beta(1+r)=1$, exogenous permanent (RW) and AR(1) transitory income, time varying income risk (AR(1)).

• **RBC**: non-separable CRRA preferences, labor supply, endogenous real rate, permanent (RW) and AR(1) transitory TFP shocks.

• **ROT**: Two agents, CRRA preferences, exogenous permanent (RW) and AR(1) transitory income, constant interest rate $\beta(1+r)/G^{\gamma-1}=1$, G growth rate of permanent income, zero saving for agents 2 (share 0.25).

• Sample 1980:1-2016:4; use real per-capita detrended (C_t, y_t, a_t).

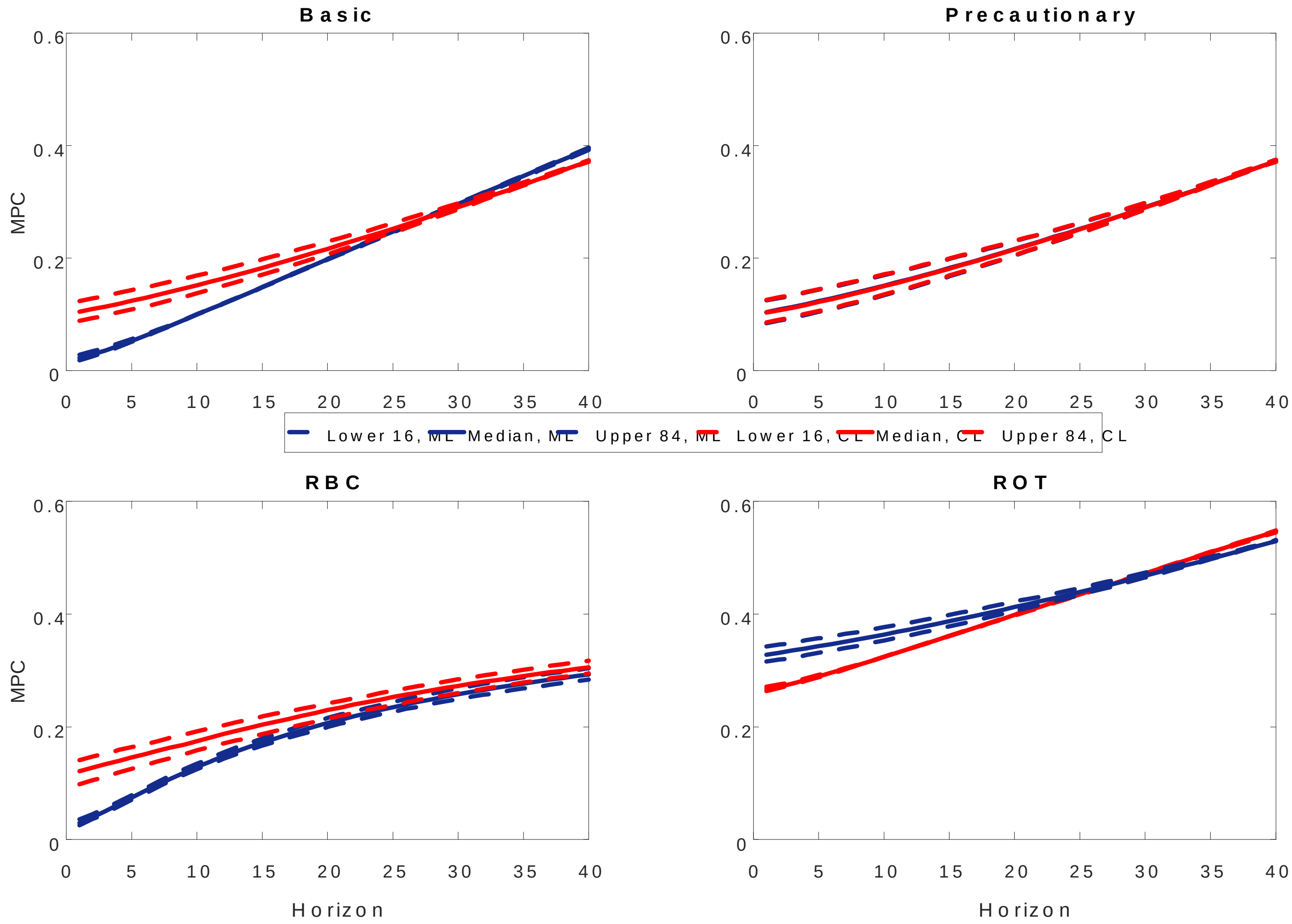
• Prior on ω Dirichlet mean:[0.25, 0.25, 0.25, 0.25].

• Estimate each model by ML. Estimate persistence of transitory income (TFP) and model weights (ω) by Bayesian CL.

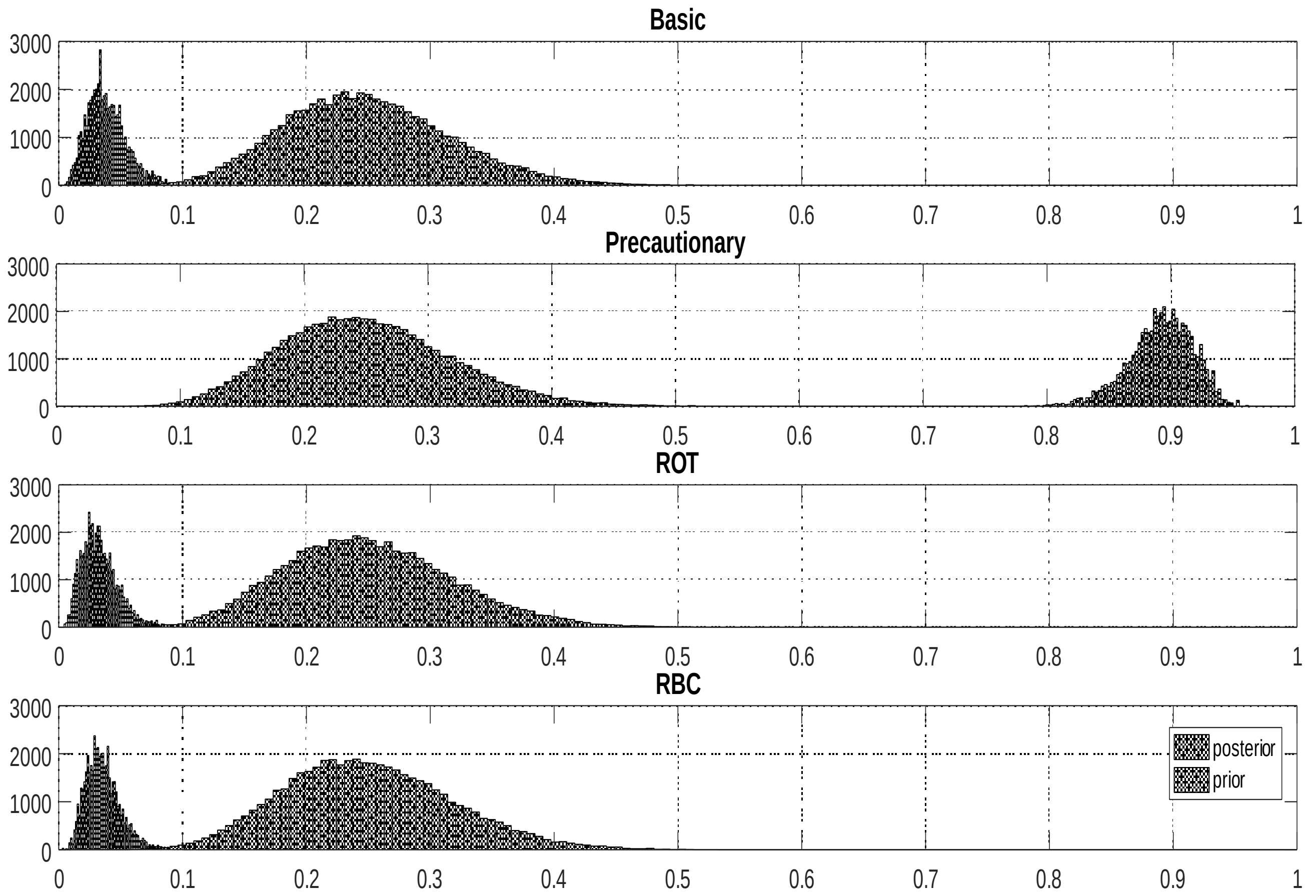
- Dynamic $MPC_y^T(l)$: $\frac{\sum_{j=1}^{l} c_{t+j}|e_t^T}{\sum_{j=1}^{l} y_{t+j}|e_t^T}$, $l = 1, 2, ...40$.

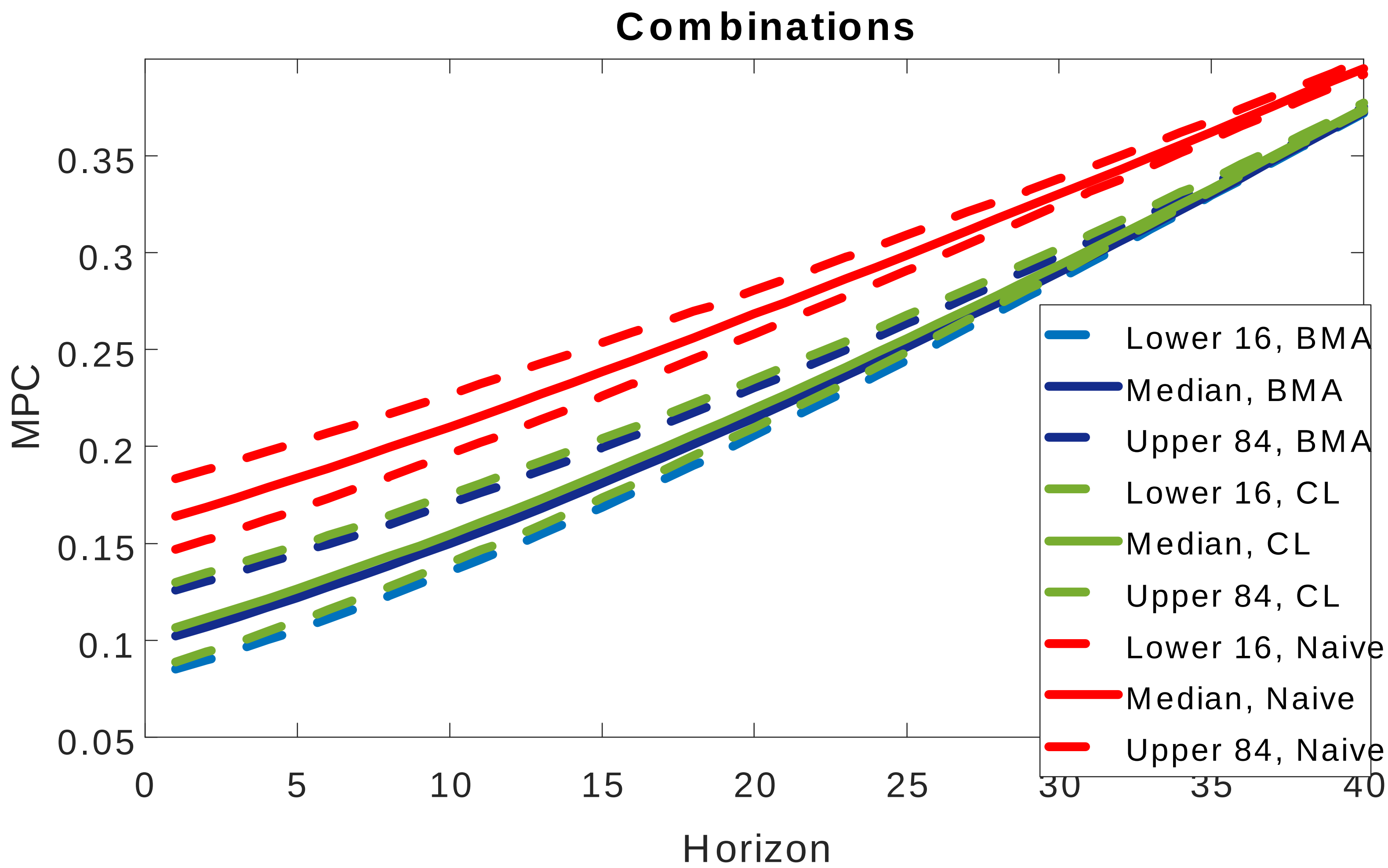
Table 3: Posterior of ρ, ML and CL

Model	16th	50th	84th
Basic	0.44	0.57	0.66
Precautionary	0.90	0.91	0.91
RBC	0.41	0.52	0.63
ROT	0.46	0.56	0.65
CL	0.85	0.90	0.96



Basic
Precautionary
RBC
ROT
MPC
Horizon
0
0.2
0.4
0.6
0
5
10
15
20
25
30
35
40
Lower 16, ML
Median, ML
Upper 84, ML
Lower 16, CL
Median, CL
Upper 84, CL


Basic
Precautionary
ROT
RBC
posterior
prior
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



Combinations
MPC
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0
5
10
15
20
25
30
35
40
Horizon
Lower 16, BMA
Median, BMA
Upper 84, BMA
Lower 16, CL
Median, CL
Upper 84, CL
Lower 16, Naive
Median, Naive
Upper 84, Naive

Measuring the slope of the Phillips curve

- Conventional wisdom (SW, 2007, ACEL, 2011): slope small $\simeq$ 0.012.

- Schorfheide (2008): Estimates depend on model specification.

- Employ CL to estimate the slope of the Phillips curve using:

i) Small scale NK model with sticky prices, non-observable marginal costs are (use: detrended Y, $\pi - \bar{\pi}, R - \bar{R}$). (Rubio-Rabanal, JME, 2005)

ii) Small scale NK model with sticky prices and wages, observable marginal costs (use: detrended Y, $\pi - \bar{\pi}, R - \bar{R}$, detrended w) (Rubio and Rabanal, JME, 2005)

iii) Medium scale NK model with capital adjustment costs (Justiniano et al., JME, 2010) (use: detrended Y, $\pi - \bar{\pi}, R - \bar{R}$, detrended C, detrended I, detrended w,detrended N).

iv) Search and matching NK model (Christoffel and Kuester, JME,2008) (use: detrended Y, $\pi - \bar{\pi}, R - \bar{R}$, detrended w/p)

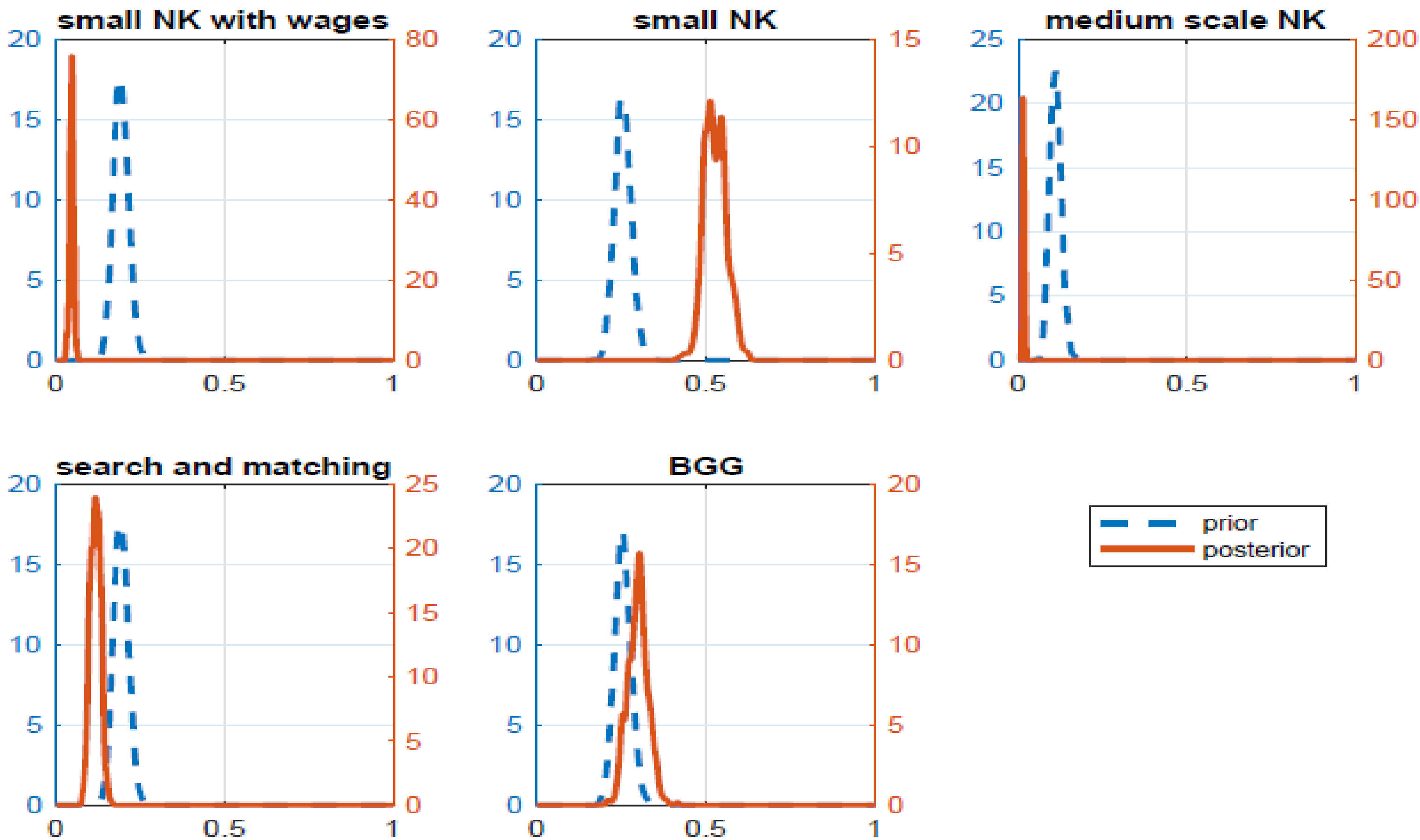
v) A financial friction NK model (NK version of Bernanke, et al., AER, 1999)(use: detrended Y, $\pi - \bar{\pi}, R - \bar{R}$)

- Sample 1960:1-2005:4; quadratic detrended data.

- Prior mean for $\omega = (0.20, 0.20, 0.20, 0.20, 0.20)$.

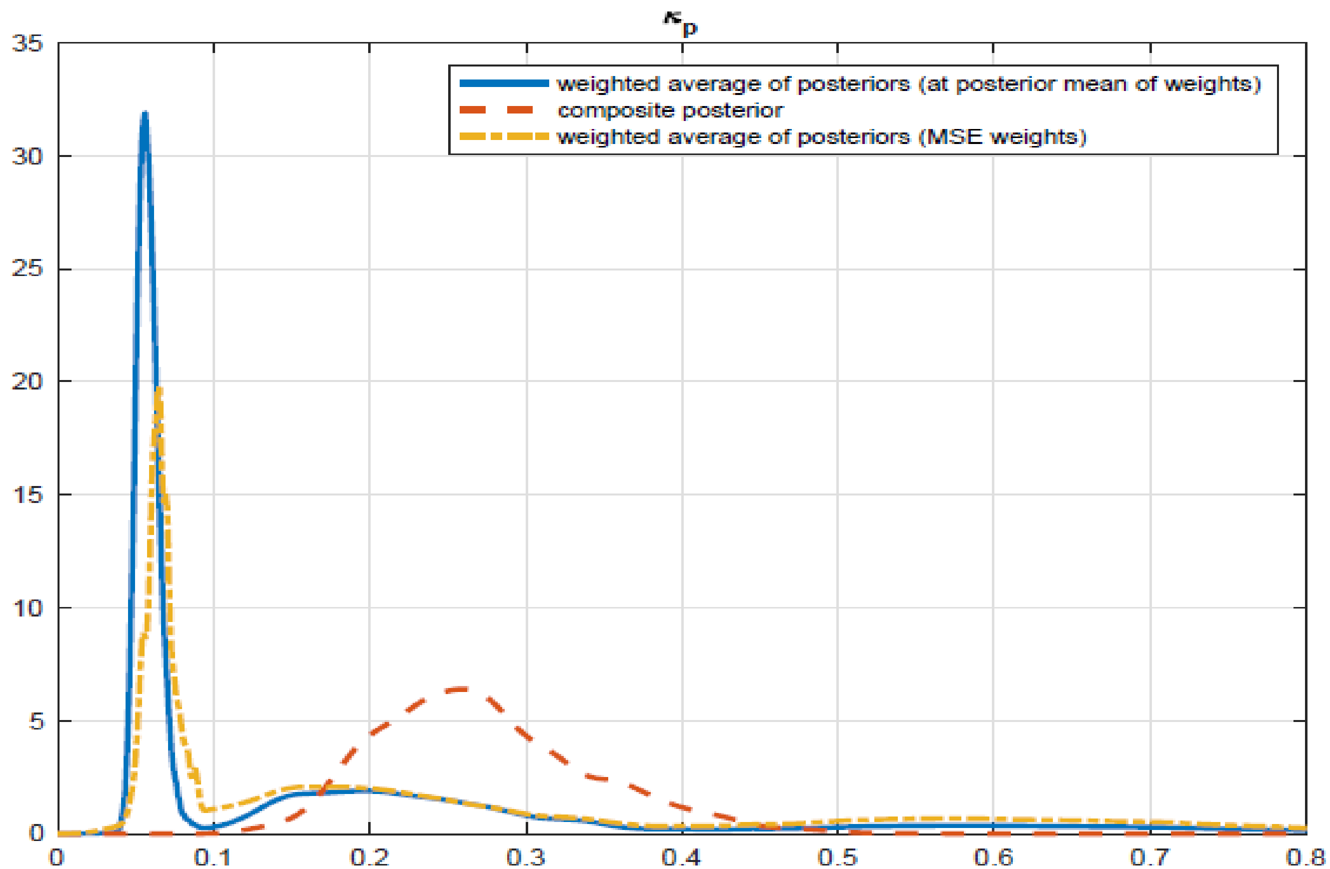
Percentiles of the posterior of the slope of the Philips curve

	5%	50%	95%
Prior	0.01	0.80	1.40
Basic NK	0.06	0.18	0.49
Basic NK with nominal wages	0.05	0.06	0.07
SW with capital and adj.costs	0.04	0.05	0.07
Search	0.44	0.62	0.86
BGG	0.13	0.21	0.35
CL	0.18	0.26	0.40
CL (corrected)	0.18	0.28	0.44

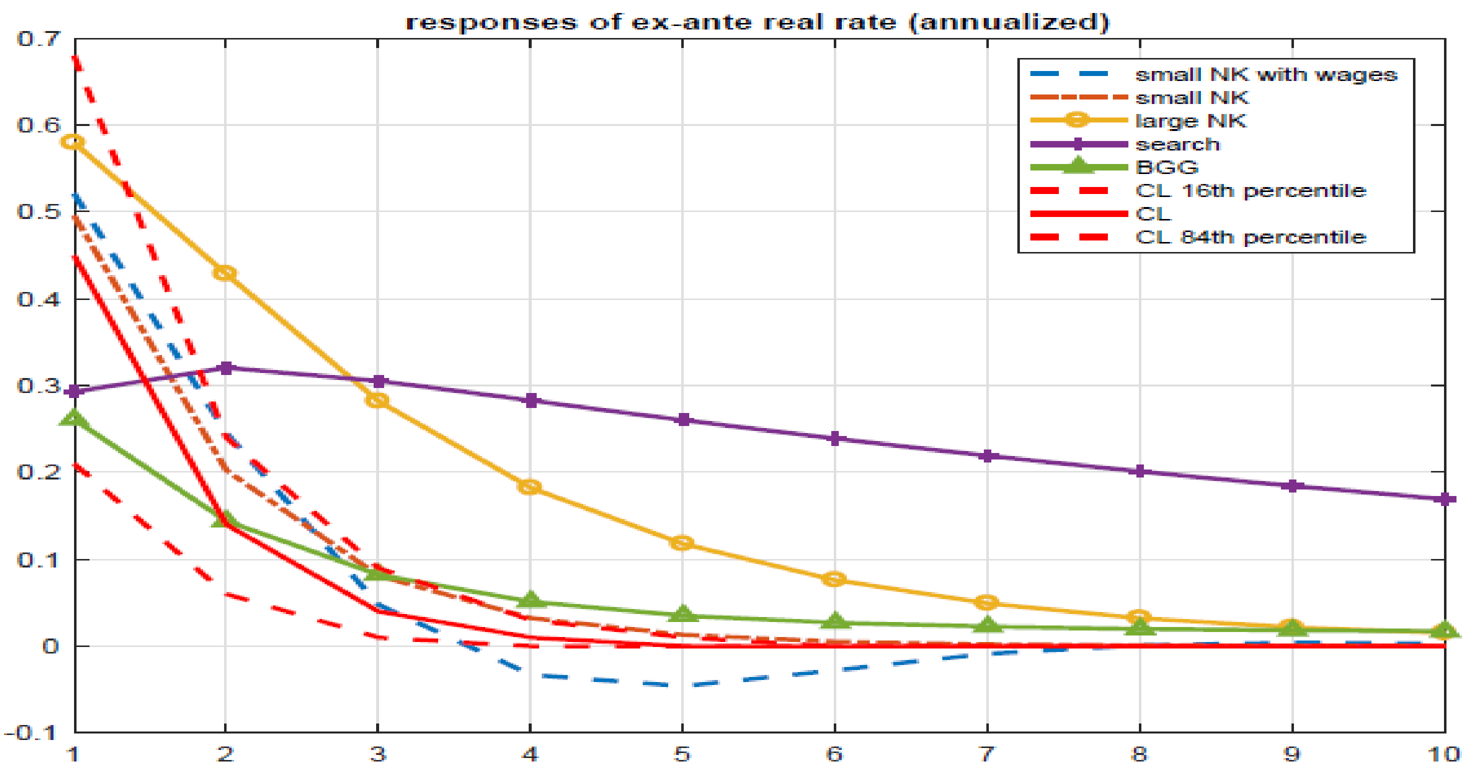

small NK with wages
small NK
medium scale NK
search and matching
BGG
prior
posterior

White distance

Model	Distance
Basic NK	4700
Basic NK with nominal wages	57300
SW with capital and adj.costs	43500
Search	415
BGG	2070
CL (loose prior)	1433
CL (tight prior)	744



κ_p
weighted average of posteriors (at posterior mean of weights)
composite posterior
weighted average of posteriors (MSE weights)
35
30
25
20
15
10
5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8



responses of ex-ante real rate (annualized)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
1
2
3
4
5
6
7
8
9
10
small NK with wages
small NK
large NK
search
BGG
CL 16th percentile
CL
CL 84th percentile