Optimal Pooling and Finite Mixture Distribution Combinations of Probabilistic Forecasts

Giulia Mantoan
giulia.mantoan.16@mail.wbs.ac.uk

November 26th, 2018

Warwick Business School - Economic Modelling and Forecasting Group

Combination of density forecasts

- Two-step procedures
- One-step procedures

Combination of density forecasts

- Two-step procedures
  1. estimate forecast models (parameter uncertainty)
     $F_1(\boldsymbol{\theta_1})$; $F_2(\boldsymbol{\theta_2})$
  2. and then, estimate combination weights (weight uncertainty)
     $y_{t+1} = \eta_1 F_1 + (1 - \eta_1) F_2 + \varepsilon_{t+1}$
- One-step procedures

Combination of density forecasts

- Two-step procedures
  1. estimate forecast models (parameter uncertainty)
     $F_1(\boldsymbol{\theta_1})$; $F_2(\boldsymbol{\theta_2})$
  2. and then, estimate combination weights (weight uncertainty)
     $y_{t+1} = \eta_1 F_1 + (1 - \eta_1) F_2 + \varepsilon_{t+1}$
- One-step procedures
  jointly addresses both weight and parameter uncertainties estimating combination weights and forecasts' parameters simultaneously
  1. $y_{t+1} = \eta_1 F_1(\boldsymbol{\theta_1}) + (1 - \eta_1) F_2(\boldsymbol{\theta_2}) + \varepsilon_{t+1}$

Literature Review

- Two-step procedures
  - The optimal pooling: DeGroot and Mortera (1991), Stock and Watson (2004), Marcellino (2004), Wallis (2005), Genre et al. (2013); Kascha and Ravazzolo (2010), Gneiting et al. (2013).
  - Bayesian Model Averaging: Raftery et al. (1997), Raftery et al. (2005)Geweke and Whiteman (2006), Rossi and Sekhposyan (2014).
- One-step procedures
  - Finite mixture Models: LeSage and Magura (1992), Waggoner and Zha (2012), Billio et al. (2013), Del Negro et al. (2016), Mandel et al. (2016).
  - Infinite mixture models: Maheu and Yang (2016), Bassetti et al. (2018)

Motivation

Why a comparison between combination approaches?

- Are they the same?
- Which one is the most accurate?
- When one is more accurate than the other.

Motivation

Why a comparison between combination approaches?

- Are they the same? **No**
- Which one is the most accurate?
- When one is more accurate than the other.

Motivation

Why a comparison between combination approaches?

- Are they the same? **No**
- Which one is the most accurate? **It depends**
- When one is more accurate than the other.

Motivation

Why a comparison between combination approaches?

- Are they the same? **No**
- Which one is the most accurate? **It depends**
- When one is more accurate than the other. **Model specification, sample size, number of individual model, ect...**

Motivation

Why a comparison between combination approaches?

- Are they the same? **No**
- Which one is the most accurate? **It depends**
- When one is more accurate than the other. **Model specification, sample size, number of individual model, ect...**

Endow the DM with a tool to elicit not only the combination model but the family they belong to

In this paper

This paper proposes a comparison between

- Geweke and Amisano (2011) as two-step approach
- a simplified version of Waggoner and Zha (2012) as one-step approach

In this paper

This paper proposes a comparison between

- Geweke and Amisano (2011) as two-step approach
- a simplified version of Waggoner and Zha (2012) as one-step approach

The comparison has carried out through:

Simulated data:

1. Baseline
2. Single-break
3. (Multiple-break)

In this paper

This paper proposes a comparison between

- Geweke and Amisano (2011) as two-step approach
- a simplified version of Waggoner and Zha (2012) as one-step approach

The comparison has carried out through:

Simulated data:

1. Baseline
2. Single-break
3. (Multiple-break)

Macroeconomic applications:

1. Small Dataset
2. Medium-Large Dataset

Two-step and One-step procedures

Two-step Procedure: the Optimal Pooling

1. A set of predictive distributions f_k for a vector of time series y_t
   $f_k = p(y_{t+1}|y_t^o, \boldsymbol{\theta}_k)$ Estimate $\boldsymbol{\theta}_k$ through OLS.

Two-step Procedure: the Optimal Pooling

1. A set of predictive distributions f_k for a vector of time series y_t $f_k = p(y_{t+1}|y_t^o, \boldsymbol{\theta}_k)$ Estimate $\boldsymbol{\theta}_k$ through OLS.
2. are taken as given and combined according to:

$$g(y_t) = \sum_{k=1}^{K} \eta_k p(y_t; y_{t-1}^o, f_k)$$

where: $\sum_{k=1}^{K} \eta_k = 1$, $\eta_k \geq 0$, for $(k = 1, \ldots, K)$.

Two-step Procedure: the Optimal Pooling

1. A set of predictive distributions f_k for a vector of time series y_t
   $f_k = p(y_{t+1}|y_t^o, \boldsymbol{\theta}_k)$ Estimate $\boldsymbol{\theta}_k$ through OLS.
2. are taken as given and combined according to:

$$g(y_t) = \sum_{k=1}^{K} \eta_k p(y_t; y_{t-1}^o, f_k)$$

where: $\sum_{k=1}^{K} \eta_k = 1$, $\eta_k \geq 0$, for $(k = 1, \ldots, K)$.

Optimal weights η_k^*

are obtained by minimising the Kullback-Leiber Information Criterion (KLIC):

$$\eta_k^* = \underset{\eta}{\text{argmax}} \; \frac{1}{T} \sum_{t=1}^{T} \ln \; g(y_t)$$

Following Hall and Mitchell (2007) and Conflitti et al. (2015) details

One-step Procedure: Finite Mixture of Distributions Model

A standard Finite Mixture Distribution is used as limiting case of the Markov Switching Model (Waggoner and Zha, 2012):

$$p(y_t|y_{t-1}^0, \boldsymbol{\theta}_k) = \sum_{k=1}^{K} p(y_t|y_{t-1}^0, S_t = k, \boldsymbol{\theta}_k) Pr(S_t = k|y_{t-1}^0, \boldsymbol{\theta}_k)$$

$$p(y_t|y_{t-1}^0, \boldsymbol{\theta}_k) = \sum_{k=1}^{K} p(y_t|y_{t-1}^0, \boldsymbol{\theta}_k) \eta_k$$

One-step Procedure: Finite Mixture of Distributions Model

A standard Finite Mixture Distribution is used as limiting case of the Markov Switching Model (Waggoner and Zha, 2012):

$$p(y_t|y_{t-1}^0, \boldsymbol{\theta}_k) = \sum_{k=1}^{K} p(y_t|y_{t-1}^0, S_t = k, \boldsymbol{\theta}_k) Pr(S_t = k|y_{t-1}^0, \boldsymbol{\theta}_k)$$

$$p(y_t|y_{t-1}^0, \boldsymbol{\theta}_k) = \sum_{k=1}^{K} p(y_t|y_{t-1}^0, \boldsymbol{\theta}_k)\eta_k$$

The weights are distributed according to

$$\eta_k \sim M\left(1, \left[\frac{p_1 f_N(y_t; \mu_{1,t}, \sigma_1^2)}{\sum_{k=1}^{K} p_k f_N(y_t; \mu_{k,t}, \sigma_k^2)}, \ldots, \frac{p_K f_N(y_t; \mu_{K,t}, \sigma_K^2)}{\sum_{k=1}^{K} p_k f_N(y_t; \mu_{k,t}, \sigma_k^2)}\right]\right)$$

where $p_k = (p_1, ..., p_K)$, $0 \leq p_k \leq 1$ and $\sum_{k=1}^{K} p_k = 1$.
Bayesian estimation with Gibbs sampling algorithm following Frühwirth-Schnatter (2006) details

Evaluation of Density Forecasts

- Logarithmic Scoring Rule: $LS(g(Y_t), A) = \log\,(g_i Y_t | A_i)$
- Continuous Ranked Probability Scores:

$$CRPS = \int_{-\infty}^{+\infty} (g(Y_{t+1}) - Y_{t+1})^2 dy$$

$$\text{TW-CRPS} = \int_{-\infty}^{+\infty} (g(z) - \mathbb{1}\{Y_{t+1} \leq z\})^2 w(z) dz$$

Evaluation of Density Forecasts

- Logarithmic Scoring Rule: $LS(g(Y_t), A) = \log\,(g_i Y_t | A_i)$
- Continuous Ranked Probability Scores:

$$CRPS = \int_{-\infty}^{+\infty} (g(Y_{t+1}) - Y_{t+1})^2 dy$$

$$\text{TW-CRPS} = \int_{-\infty}^{+\infty} (g(z) - \mathbb{1}\{Y_{t+1} \leq z\})^2 w(z) dz$$

- Probability Integral Transforms (PITs): the density forecast is called "calibrated" if z_t is uniform:

$$z_t = \int_{-\infty}^{Y_t} f_k(Y_t) d(Y_t)$$

Comparison with simulated data

No-break
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

- Data Generating Process (sample sizes $T = \{50, 200, 1000\}$)

$$y_t = 0.5 + 0.8y_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

No-break
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

- Data Generating Process (sample sizes $T = \{50, 200, 1000\}$)

$$y_t = 0.5 + 0.8y_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

- Predictive distributions f_k can be:

Nonnested case	Nested case
$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$	$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$
$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$	$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1} + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$

No-break
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

- Data Generating Process (sample sizes $T = \{50, 200, 1000\}$)

$$y_t = 0.5 + 0.8y_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

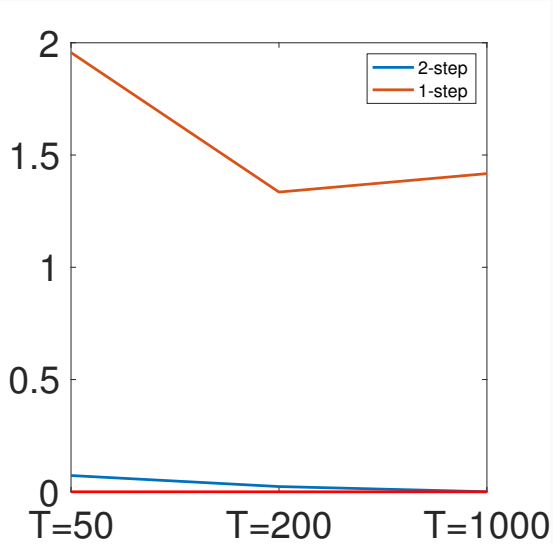
- Predictive distributions f_k can be:

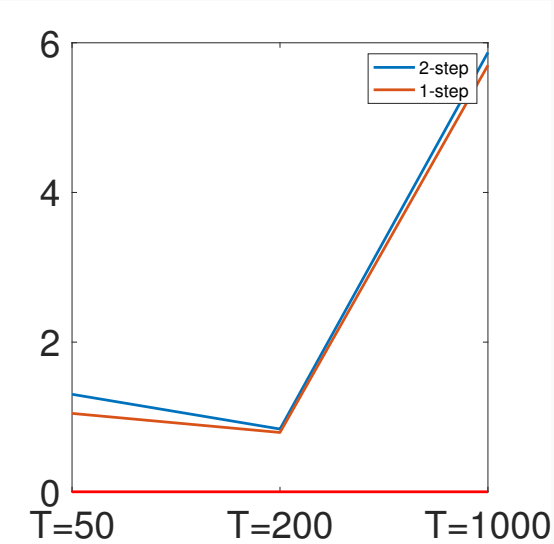
Nonnested case

$$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$$
$$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$$

Nested case

$$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$$
$$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1} + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$$





Single-Break Scenario: DGP Setups

- Data Generating Process

$$y_t = \Phi \boldsymbol{X}_t + \varepsilon_t \;\; \epsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

$$\boldsymbol{X}_t = \begin{bmatrix} 1 \\ y_{t-1} \\ x_{t-1} \end{bmatrix} \qquad \Phi = \begin{cases} [\phi_0 = 0.5 \quad \phi_1 = 0.8 \quad \phi_2 = 0] \;, & \text{if } t < T_b \\ [\phi_0 + d_0 \quad \phi_1 + d_1 \quad \phi_2 + d_2] \;, & \text{if } T_b \leq t \leq T. \end{cases}$$

$T = \{50, 200, 1000\}$ and $T_b = \{0.25, 0.50, 0.75, 0.95\}$ of T.

Single-Break Scenario: DGP Setups

- Data Generating Process

$$y_t = \Phi \boldsymbol{X}_t + \varepsilon_t \;\; \epsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

$$\boldsymbol{X}_t = \begin{bmatrix} 1 \\ y_{t-1} \\ x_{t-1} \end{bmatrix} \qquad \Phi = \begin{cases} [\phi_0 = 0.5 \quad \phi_1 = 0.8 \quad \phi_2 = 0] \;, & \text{if } t < T_b \\ [\phi_0 + d_0 \quad \phi_1 + d_1 \quad \phi_2 + d_2] \;, & \text{if } T_b \leq t \leq T. \end{cases}$$

$T = \{50, 200, 1000\}$ and $T_b = \{0.25, 0.50, 0.75, 0.95\}$ of T.

- Predictive distributions f_k can be:

Nonnested case:

$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$

$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$

Nested case:

$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$

$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1} + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$

Single-Break Scenario: DGP Setups

- Data Generating Process

$$y_t = \Phi \boldsymbol{X}_t + \varepsilon_t \quad \epsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

$$\boldsymbol{X}_t = \begin{bmatrix} 1 \\ y_{t-1} \\ x_{t-1} \end{bmatrix} \qquad \Phi = \begin{cases} [\phi_0 = 0.5 \quad \phi_1 = 0.8 \quad \phi_2 = 0] \ , & \text{if } t < T_b \\ [\phi_0 + d_0 \quad \phi_1 + d_1 \quad \phi_2 + d_2] \ , & \text{if } T_b \leq t \leq T. \end{cases}$$

$T = \{50, 200, 1000\}$ and $T_b = \{0.25, 0.50, 0.75, 0.95\}$ of T.

- Predictive distributions f_k can be:

Nonnested case:

$$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$$
$$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$$

Nested case:

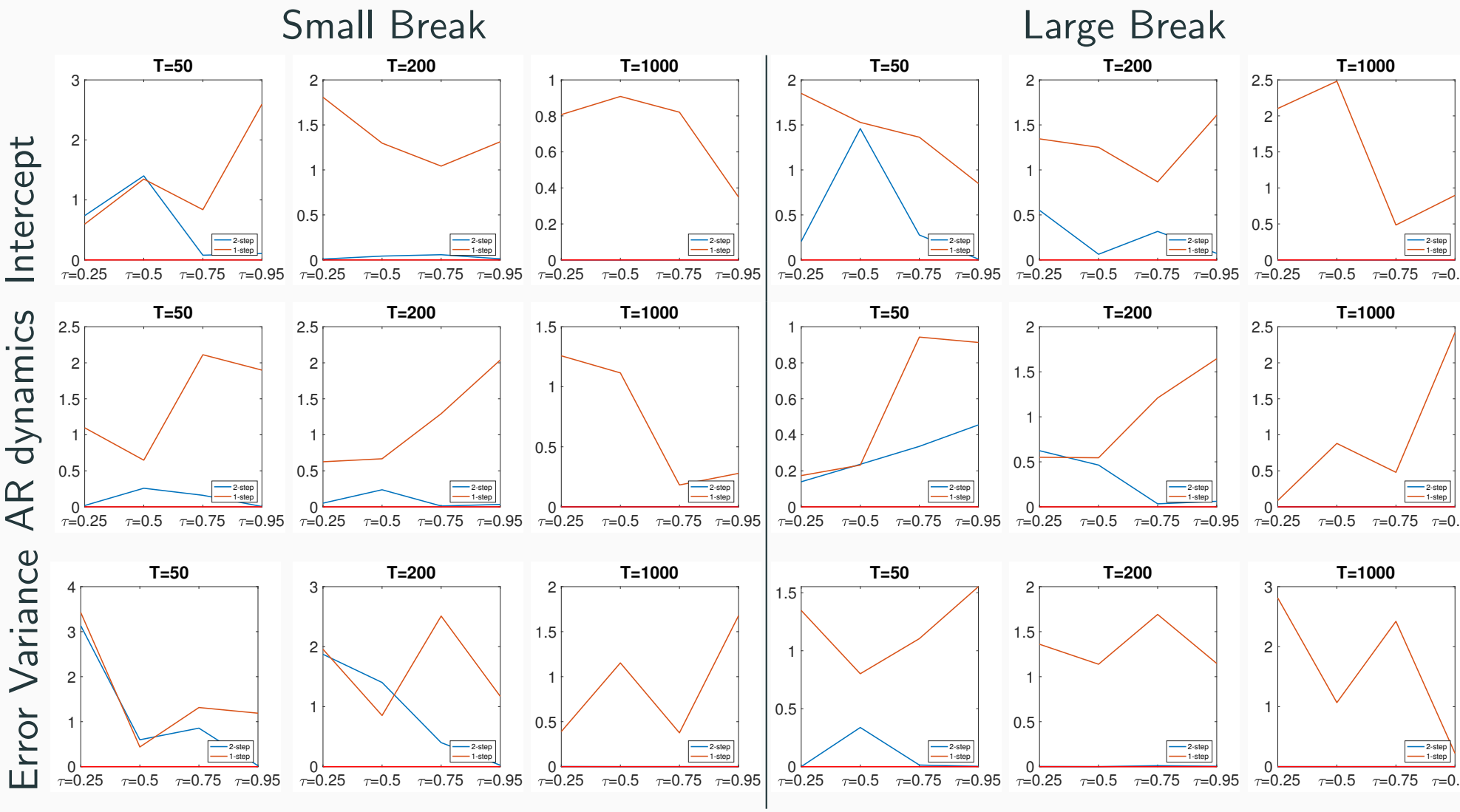
$$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$$
$$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1} + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$$

f_1 and f_2 are combined treating the break as unknown

Single-break scenario: Types of break

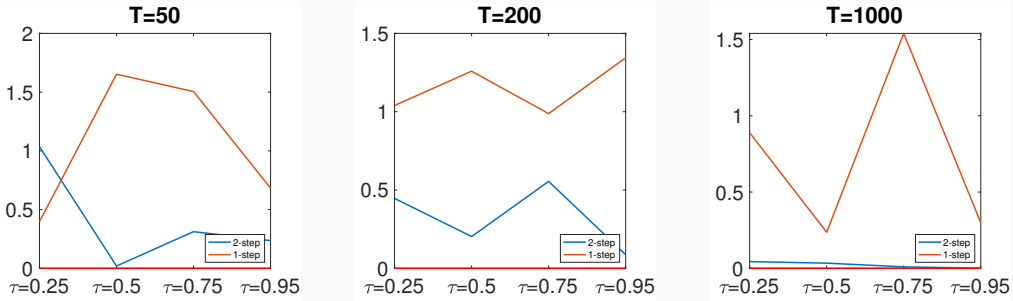
exp. #	d_0	d_1	d_2	σ_ϵ^2	Comments
1	0	0	0	0.6	no breaks
2	-0.4	0	0	0.6	small break in the intercept
3	-0.6	0	0	0.6	large break in the intercept
4	0	-0.2	0	0.6	small break in AR(1) dynamics
5	0	-0.4	0	0.6	large break in AR(1) dynamics
6	0	0	0.5	0.6	Small break in exo. var. coefficient
7	0	0	1	0.6	Large break in exo. var. coefficient
8	0	-0.2	0.5	0.6	Breaks in AR(1) and exo. var. coefficients
9	0	0	0	2	Increase in post-break variance
10	0	0	0	0.3	Decrease in post-break variance

Single-break Nonnested case
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

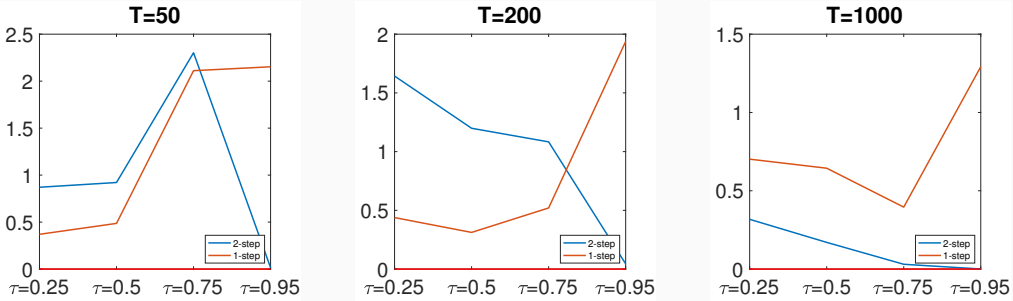


Single-break Nonnested case
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

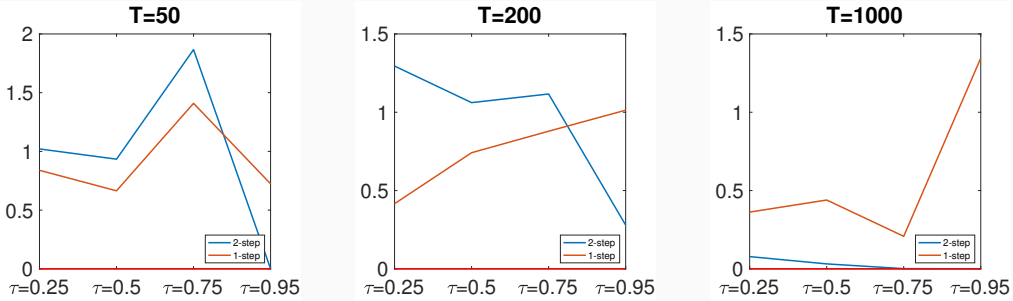
Small Break Exogenous Variable



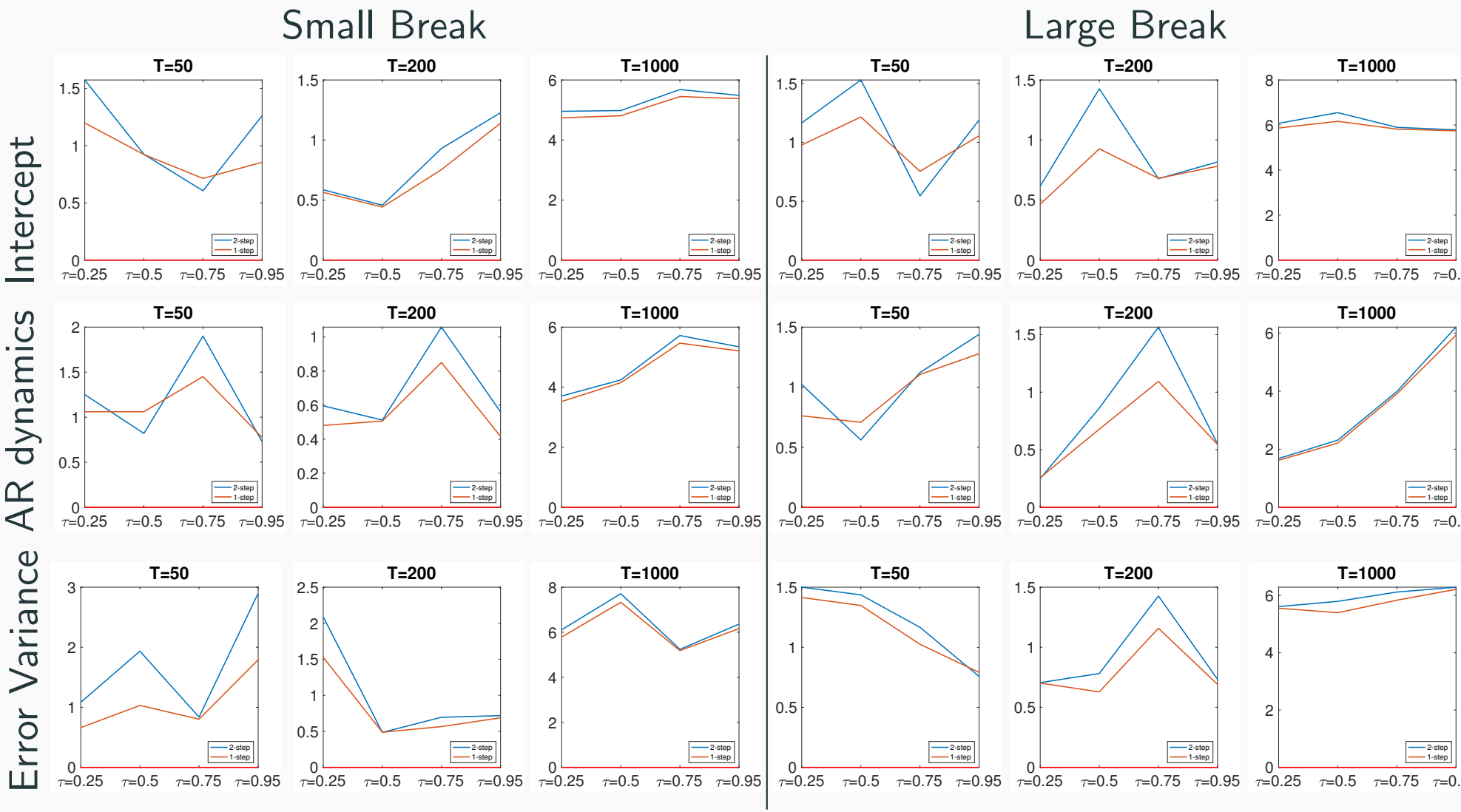
Large Break Exogenous variable



Break in both AR and Exogenous var.

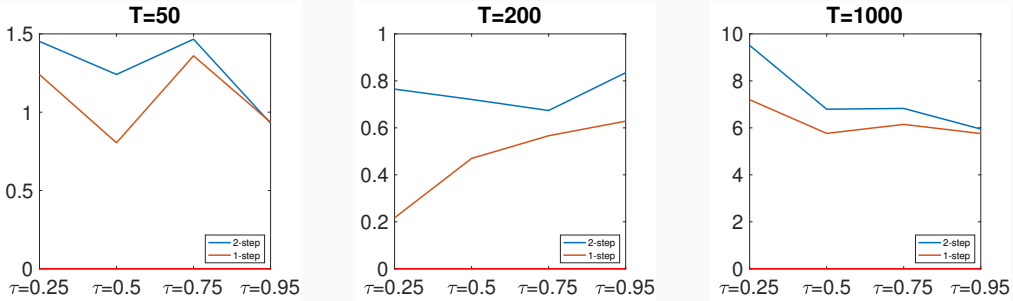


Single-break Nested case
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

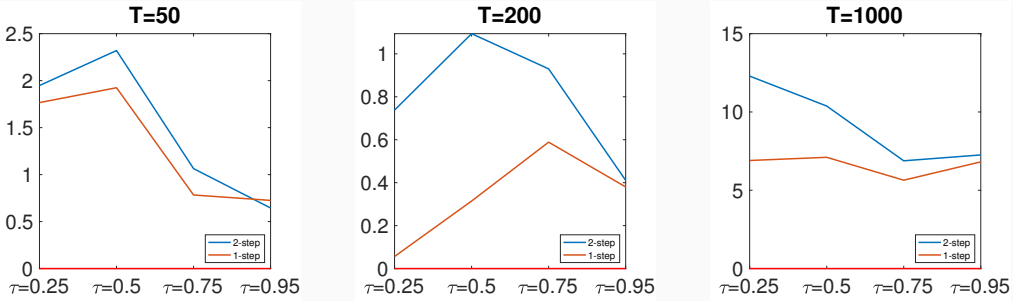


Single-break Nested case
Accuracy Loss: $(CRPS_{comb} - CRPS_{DGP})/CRPS_{DGP}$

Small Break Exogenous Variable

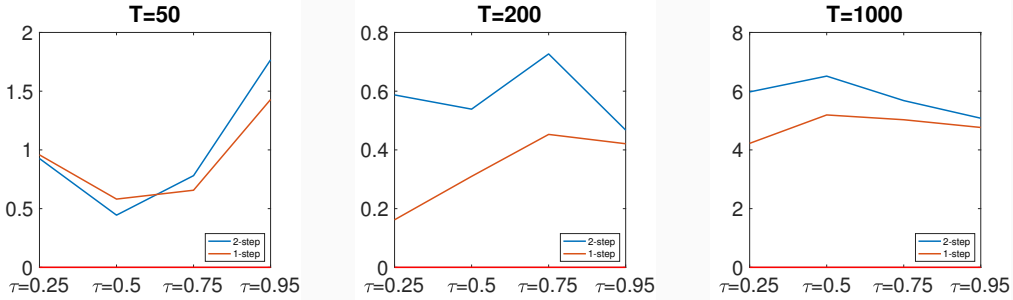


Large Break Exogenous variable



Multiple Breaks

Break in both AR and Exogenous var.



Empirical Applications

Forecasting GDP with industrial Production Index and Employment rate

Forecasting Real Output growth rate for US based on

- Industrial Production Index and Non-farm payroll employment
- In-sample forecast
- from 1964:Q1 to 2018:Q1
- One-step ahead rolling-window forecasts
- Using two datasets:
  1. Ex-post revised data (FRED database)
  2. Real-time data (Philadelphia Fed database)

Forecasting GDP with industrial Production Index and Employment rate

- Two predictive regressions:
  - $y_{t+1}^{IP} = \hat{\alpha}_{1,t} + \sum_{i=0}^{p} \hat{\beta}_{i,t} y_{t-i}^{o} + \sum_{j=0}^{Q} \hat{\delta}_{j,t} IP_{t-j} + \varepsilon_{1,t}$ and
  - $y_{t+1}^{EMP} = \hat{\alpha}_{2,t} + \sum_{i=0}^{p} \hat{\beta}_{i,t} y_{t-i}^{o} + \sum_{n=0}^{N} \hat{\gamma}_{n,t} EMP_{t-n} + \varepsilon_{2,t}$

  where $\varepsilon_{k,t} \sim N(0, \sigma_{k,t}^2)$.

Forecasting GDP with industrial Production Index and Employment rate

- Two predictive regressions:
  - $y_{t+1}^{IP} = \hat{\alpha}_{1,t} + \sum_{i=0}^{p} \hat{\beta}_{i,t} y_{t-i}^{o} + \sum_{j=0}^{Q} \hat{\delta}_{j,t} IP_{t-j} + \varepsilon_{1,t}$ and
  - $y_{t+1}^{EMP} = \hat{\alpha}_{2,t} + \sum_{i=0}^{p} \hat{\beta}_{i,t} y_{t-i}^{o} + \sum_{n=0}^{N} \hat{\gamma}_{n,t} EMP_{t-n} + \varepsilon_{2,t}$

  where $\varepsilon_{k,t} \sim N(0, \sigma_{k,t}^2)$.
- Forecasts combined according to

$$p^c(y_{t+1}|y_t^o, \boldsymbol{\theta}_{k,t}) = \sum_{k=1}^{K=2} \hat{\eta}_{k,t} p(y_{t+1}|y_t^o, \boldsymbol{\theta}_{k,t})$$

Where $\Phi_{k,t} = \{\alpha_{k,t}, \delta_{j,t}, \beta_{i,t}, \gamma_{n,t}, \sigma_{k,t}^2, \eta_{k,t}\}$ are estimated according to the one-step and two-step procedures

Forecasting GDP with industrial Production Index and Employment rate

- Two predictive regressions:
  - $y_{t+1}^{IP} = \hat{\alpha}_{1,t} + \sum_{i=0}^{p} \hat{\beta}_{i,t} y_{t-i}^{o} + \sum_{j=0}^{Q} \hat{\delta}_{j,t} IP_{t-j} + \varepsilon_{1,t}$ and
  - $y_{t+1}^{EMP} = \hat{\alpha}_{2,t} + \sum_{i=0}^{p} \hat{\beta}_{i,t} y_{t-i}^{o} + \sum_{n=0}^{N} \hat{\gamma}_{n,t} EMP_{t-n} + \varepsilon_{2,t}$

  where $\varepsilon_{k,t} \sim N(0, \sigma_{k,t}^2)$.
- Forecasts combined according to

$$p^c(y_{t+1}|y_t^o, \boldsymbol{\theta}_{k,t}) = \sum_{k=1}^{K=2} \hat{\eta}_{k,t} p(y_{t+1}|y_t^o, \boldsymbol{\theta}_{k,t})$$

Where $\Phi_{k,t} = \{\alpha_{k,t}, \delta_{j,t}, \beta_{i,t}, \gamma_{n,t}, \sigma_{k,t}^2, \eta_{k,t}\}$ are estimated according to the one-step and two-step procedures

One-step and Two-step procedures compared with respect to:

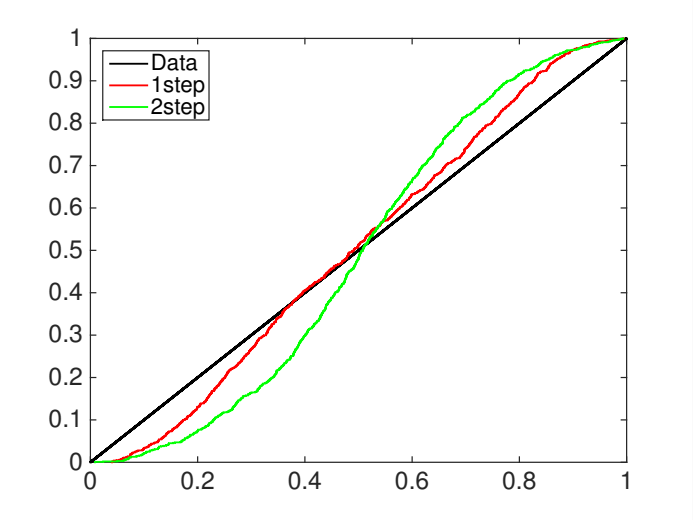
- a benchmark AR(1) model
- both density and point forecasts

Macroeconomic Application - Results

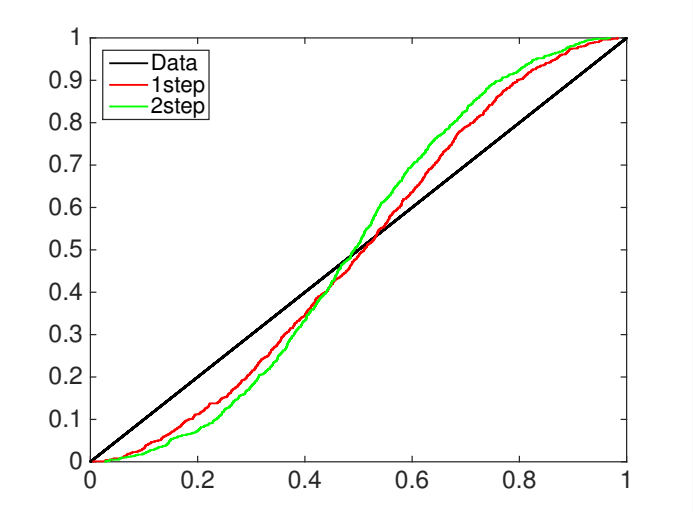
	Ex-Post Revised data		
	AR(1)	2-step	1-step
Log Score	-0.9818	-1.1765	-0.9134
		(0.000)	(0.000)
			(0.000)
CRPS	0.2289	0.4879	0.1331
		(0.000)	(0.000)
			(0.000)
TW-CRPS	0.2289	0.4880	0.1330
		(0.000)	(0.000)
			(0.000)
MSPE	0.2209	0.2215	0.2220
		(0.8236)	(0.3923)
			(0.8236)

	Real-Time data		
	AR(1)	2-step	1-step
Log Score	-0.9298	-0.9974	-0.9126
		(0.000)	(0.000)
			(0.000)
CRPS	0.1008	0.2916	0.0756
		(0.000)	(0.000)
			(0.000)
TW-CRPS	0.1008	0.2916	0.0760
		(0.000)	(0.000)
			(0.000)
MSPE	0.3629	0.3573	0.3914
		(0.3964)	(0.0362)
			(0.0521)

Ex-post revised data



Real-Time data



Forecasting US GDP and Inflation in a Large Macroeconomic Dataset

Forecasting Real Output growth and Inflation rate for US based on

- Replicate Rossi and Sekhposyan (2014)
- Subsample of Stock and Watson (2003) dataset details
- In-sample forecast
- from 1984:Q1 to 2010:Q3
- One-step ahead rolling-window forecasts of 40 observations

Forecasting US GDP and Inflation in a Large Macroeconomic Dataset

- K=32 forecasting models:

$$y_{t+1,k} = \beta_0 + \beta_1(L)X_{t,k} + \beta_{2,k}(L)y_t + \varepsilon_{t+1,k}$$

where $\varepsilon_{k,t} \sim N(0, \sigma^2_{k,t})$, $\beta_i(L) = \sum_{p=0}^{P} \beta_i L^i$

Forecasting US GDP and Inflation in a Large Macroeconomic Dataset

- K=32 forecasting models:

$$y_{t+1,k} = \beta_0 + \beta_1(L)X_{t,k} + \beta_{2,k}(L)y_t + \varepsilon_{t+1,k}$$

where $\varepsilon_{k,t} \sim N(0, \sigma^2_{k,t})$, $\beta_i(L) = \sum_{p=0}^{P} \beta_i L^i$

- Forecasts combined according to

$$p(y_{t+1}|y_t^o, \boldsymbol{\theta}_{k,t}) = \sum_{k=1}^{K} \eta_k p(y_{t+1,k}|y_t^o, \boldsymbol{\theta}_{k,t})$$

Where $\Phi_{k,t} = \{\beta_{i,t}, \sigma^2_{k,t}, \eta_{k,t}\}$ are estimated according to the one-step and two-step procedures algorithm

Forecasting US GDP and Inflation in a Large Macroeconomic Dataset

- K=32 forecasting models:

$$y_{t+1,k} = \beta_0 + \beta_1(L)X_{t,k} + \beta_{2,k}(L)y_t + \varepsilon_{t+1,k}$$

  where $\varepsilon_{k,t} \sim N(0, \sigma^2_{k,t})$, $\beta_i(L) = \sum_{p=0}^{P} \beta_i L^i$
- Forecasts combined according to

$$p(y_{t+1}|y_t^o, \boldsymbol{\theta}_{k,t}) = \sum_{k=1}^{K} \eta_k p(y_{t+1,k}|y_t^o, \boldsymbol{\theta}_{k,t})$$

Where $\Phi_{k,t} = \{\beta_{i,t}, \sigma^2_{k,t}, \eta_{k,t}\}$ are estimated according to the one-step and two-step procedures algorithm

One-step and Two-step procedures compared with respect to:

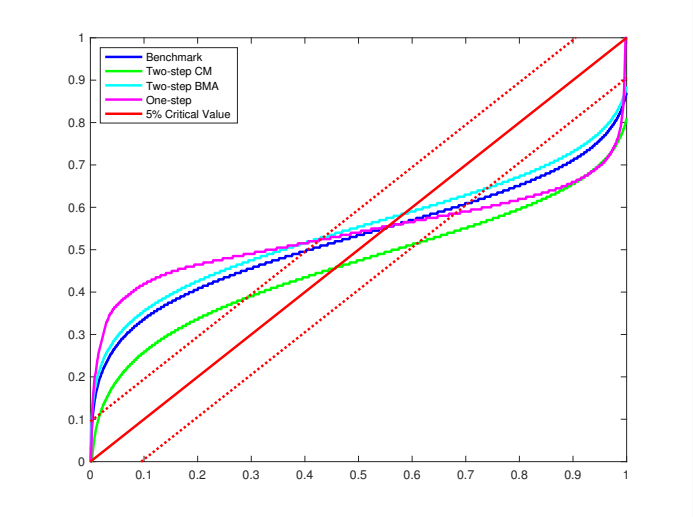
- a benchmark AR(1) model
- BMA pooling model as in Rossi and Sekhposyan (2014)

Large Macroeconomic Dataset Application - Results

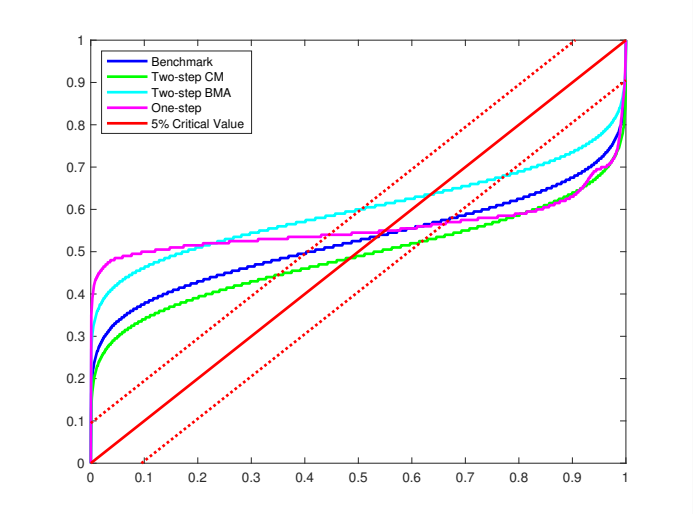
	AR(1)	2-stepCM	2-stepBMA	1-step
LogScore	-0.9258	-0.9376	-0.9247	-0.9300
		(0)	(0.3003)	(0)
			(0)	(0.0008)
				(0)
CRPS	0.2312	0.2406	0.2310	0.2374
		(0)	(0.6521)	(0)
			(0.0006)	(0.0966)
				(0)
TW-CRPS	0.2313	0.2408	0.2311	0.2372
		(0)	(0.6239)	(0)
			(0.0006)	(0.0737)
				(0)

	AR(1)	2-step CM	2-step BMA	1-step
Log Score	-0.9220	-0.9278	-0.9212	-1.1458
		(0.0062)	(0.0008)	(0)
			(0.0025)	(0)
				(0)
CRPS	0.2427	0.2467	0.2426	0.5211
		(0.0118)	(0.9014)	(0)
			(0.0113)	(0)
				(0)
TW-CRPS	0.2429	0.2466	0.2425	0.5211
		(0.0178)	(0.0015)	(0)
			(0.0103)	(0)
				(0)

GDP



Inflation



Conclusions

To elicit the combination approaches, the DM has to account for:

- Individual models specification:
  - Nested in favour of one-step approach
  - Nonnested in favour of two-step approach
- Number of models to combine:
  - Two-step is not affected by the number of individual models
  - Carefully accounting for large dataset estimating the one-step approach.

Appendix

Conflitti et al. (2015)

$$\Phi(\omega_j^*) = \frac{1}{T}\sum_{t=1}^{T} \ln\, g(Y_t) \tag{1}$$

where $\omega_{\hat{OPT}}$ maximises $\Phi(\omega_j^*)$ subject to the constraints $\omega_j \geq 0$ and $\sum_{j=1}^{m} \omega_j = 1$. Let us define the $T \times J$ matrix $\hat{G}$ composed by nonnegative elements $\hat{G}_{tj} = \hat{g}_t(Y_t)$. Then equation 1 can be rewritten as:

$$\Phi(\omega_j^*) = \frac{1}{T}\sum_{t=1}^{T} \ln\, (\hat{G}\omega_j). \tag{2}$$

Let as introduce the following Lagrange multiplier λ to take into account the constraints of the weights:

$$\Phi(\omega_j^*) = \frac{1}{T}\sum_{t=1}^{T} \ln\, (\hat{G}\omega_j) - \lambda \sum_{j=1}^{m} \omega_j. \tag{3}$$

Following Conflitti et al. (2015), we introduce a "surrogate" cost function depending on a vector of arbitrary weights a_j, such that:

$$\Psi_\lambda(\omega_j, a_j) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{m}\frac{\hat{G}_{jt}a_j}{\sum_{j=1}^{m}\hat{G}_{jt}a_j}\ln\left(\frac{\omega_j}{a_j}\sum_{j=1}^{m}\hat{G}_{jt}a_j\right) - \lambda\sum_{j=1}^{m}\omega_j. \tag{4}$$

Let us define the following algorithm for k numbers of iterations:

$$\omega_{j,\lambda}^{k+1} = \underset{\omega}{\operatorname{argmax}}\ \Psi_\lambda(\omega_j, \omega_{j,\lambda}^{k}) \tag{5}$$

Rewriting last equation in terms of ω_j^k, the iterative algorithm becomes:

$$\omega_j^{k+1} = \omega_j^k\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{G}_{jt}}{\sum_{j=1}^{m}\hat{G}_{jt}\omega_j^k}. \tag{6}$$

The nonnegative constrain is satisfied imposing positive weights that sum to one as initial values (i.e. $\omega_j^0 = 1/m$). The iterates are expected to converge to the maximiser $\omega_{\hat{OPT}}$ due to the monotonicity of the cost function in (4) and the constraints. The algorithm has also a stop criterion based on negligible difference between two successive iterates. back

Description of Data Series

Label	Trans	Period	Name	Description	Source
			Asset Prices		
rovnght@us	level	59:M1-10:M9	FEDFUNDS	Int. Rate: Fed Funds (Effective)	F
rtbill@us	level	59:M1-10:M9	TB3MS	Int. Rate: 3-Mn Tr. Bill, Sec Mkt Rate	F
rbnds@us	level	59:M1-10:M9	GS1	Int. Rate: US Tr. Const. Mat., 1-Yr	F
rbndm@us	level	59:M1-10:M9	GS5	Int. Rate: US Tr. Const. Mat., 5-Yr	F
rbndl@us	level	59:M1-10:M9	GS10	Int. Rate: US Tr. Const. Mat., 10-Yr	F
stockp@us	Δln	59:M1-10:M9	SP500	US Share Prices: S&P 500	F
exrate@us	Δln	73:M1-10:M9	111	NEER	I
			Real Activity		
rgdp@us	Δln	59:Q1-10:Q3	GDPC12	Real GDP, sa	F
ip@us	Δln	59:M1-10:M9	INDPRO	Industrial Production Index, sa	F
capu@us	level	59:M1-10:M9	CUMFNS	Capacity Utilization Rate: Man., sa	F
emp@	Δln	59:M1-10:M9	CE16OV	Civilian Employment: thsnds,sa	F
unemp@us	level	59:M1-10:M9	UNRATE	Civilian Unemployment,sa	F
			Wages and Prices		
pgdp@us	Δln	59:Q1-10:Q3	GDPDEF	GDP Deflator, sa	F
cpi@us	Δln	59:M1-10:M9	CPIAUCSL	CPI: Urban, All items, sa	F
ppi@us	Δln	59:M1-10:M9	PPIACO	Producer Price Index, nsa	F
earn@us	Δln	59:M1-10:M9	AHEMAN	Hourly Earnings: Man., nsa	F
			Money		
mon0@us	Δln	59:M1-10:M9	AMBSL	Monetary Base, sa	I
mon1@us	Δln	59:M1-10:M9	M1SL	Money: M1, sa	I
mon2@us	Δln	59:M1-10:M9	M2SL	Money: M2, sa	I
mon3@us	Δln	59:M1-06:M2	M3SL	Money: M3, sa	I

back

One-step procedure - Bayesian Inference

Bayesian inference technique of MCMC estimation using data augmentation and Gibbs sampling.
Posterior Distributions are:

- Combination Weights η_k: $p(\eta_k|\boldsymbol{S}) \sim \mathcal{D}(e_1(\boldsymbol{S}), \ldots, e_K(\boldsymbol{S}))$
- Autoregressive parameters ϕ_k: $p(\phi_k|\boldsymbol{\sigma}_k^2, y_t^o, \boldsymbol{S}) \sim \mathcal{N}(a_k, A_k)$
  where $A_k = (A_0^{-1} + \frac{1}{\sigma_k^2}\boldsymbol{z}_k'\boldsymbol{z}_k)^{-1} \quad ; \quad a_k = A_k(A_0^{-1}a_0 + \frac{1}{\sigma_k^2}\boldsymbol{z}_k'y_k)$
- Variances σ_k^2: $p(\sigma_k|\boldsymbol{\theta}_k, y_t^o, \boldsymbol{S}^{m-1}) \sim \mathcal{G}^{-1}(c_k, C_k)$
  where $c_k = c_0 + \frac{N_k}{2}, \qquad C_k = C_0 + \frac{1}{2}\epsilon_k'\epsilon_k$ and $\epsilon_k = y_t^o - Z_k\phi_k$.

A common choice of prior takes the form:

$$p(\boldsymbol{\theta}_k) = \mathcal{D}(e_0, e_0|\boldsymbol{S}) \prod_{k=1}^{2} \mathcal{N}(\phi_k|a_0, A_0)\mathcal{G}^{-1}(\sigma_k^2|c_0, C_0).$$

One-step procedure - Bayesian Inference

Dealing with Label Switching

$$p(y_t|\boldsymbol{\theta}_k, \eta_k) = \sum_{k=1}^{K} \eta_k f_N(y_t|\boldsymbol{\theta}_k) = \sum_{k=1}^{K} \eta_{\rho(k)} f_N(y_t|\boldsymbol{\theta}_{\rho(k)})$$

where ρ is an arbitrary permutation of $\{1, \ldots, K\}$.

back.

3. Multi-break Scenario

$$y_t = \Phi \boldsymbol{X}_t + \varepsilon_t \qquad \epsilon_t \overset{\text{iid}}{\sim} (0, \sigma_\varepsilon^2)$$

$$\boldsymbol{X}_t = \begin{bmatrix} 1 \\ y_{t-1} \\ x_{t-1} \end{bmatrix} \qquad \Phi = \begin{cases} [\phi_0 = 0.5 \quad \phi_1 = 0.8 \quad \phi_2 = 0] \ , & \text{if } t < T_b \\ [\phi_0 + d_0 \quad \phi_1 + d_1 \quad \phi_2 + d_2] \ , & \text{if } T_b \leq t \leq T_c \\ [\phi_0 + d_0^* \quad \phi_1 + d_1^* \quad \phi_2 + d_2^*] \ , & \text{if } T_c \leq t \leq T \end{cases} \tag{7}$$

$T = \{50, 200, 1000\}$, $T_b = 0.25$, $T_c = 0.75$ of T.

exp #	d_0	d_1	d_2	d_0^*	d_1^*	d_2^*	σ_ϵ^2	Comments
11	-0.2	0	0	0	0	0	0.6	mean reversion in intercept
12	-0.2	0	0	-0.4	0	0	0.6	decreasing trend in the intercept
13	0	-0.2	0	0	0	0	0.6	mean reversion in AR dynamics
14	0	-0.2	0	0	-0.4	0	0.6	decreasing trend in AR dynamics
15	0	0.2	0	0	0.4	0	0.6	increasing trend in AR dynamics
16	0	0	1	0	0	0	0.6	mean reversion in predictor coefficient
17	0	0	1	0	0	2	0.6	trending break in predictor coefficient
18	0	0	0	0	0	0	2	increase in post-break variance
19	0	0	0	0	0	0	0.3	decrease in post-break variance

2. Combination

- Individual Models

Nonnested case:	Nested case:
$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$	$f_1 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1}, \hat{\sigma}^2_{\varepsilon_1}\big)$
$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$	$f_2 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1} + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_2}\big)$
$f_3 : y_{T+h} \sim \mathcal{N}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_3}\big)$	$f_3 : y_{T+h} \sim \mathcal{N}\big(\hat{\theta}_0 + \hat{\theta}_1 y_{T+h-1} + \hat{\beta}_1 x_{T+h-1}, \hat{\sigma}^2_{\varepsilon_3}\big)$

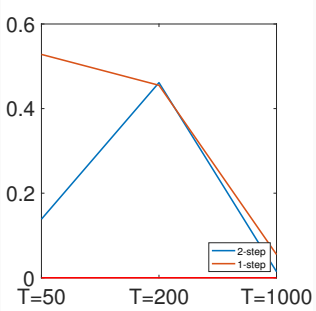
- $\hat{\theta}_0, \hat{\beta}_0, \hat{\theta}_1, \hat{\beta}_1, \hat{\sigma}^2_{\varepsilon_1}, \hat{\sigma}^2_{\varepsilon_2}$ are estimated by Two-step and One-step approach,
- f_1 and f_2 are combined treating the break as unknown.

back

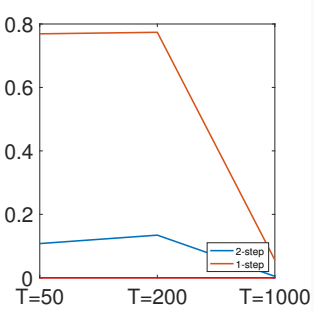
3. Multi-break Results

Nonnested Case

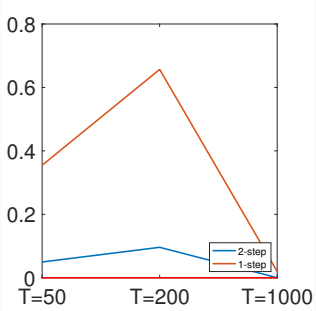
Exp.11



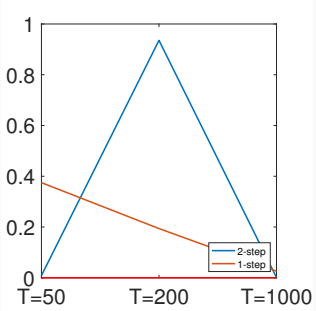
Exp.12



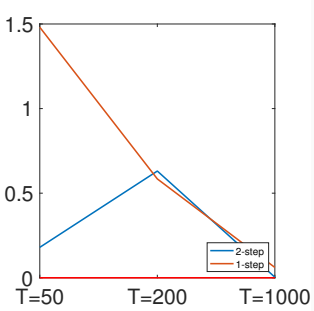
Exp.13



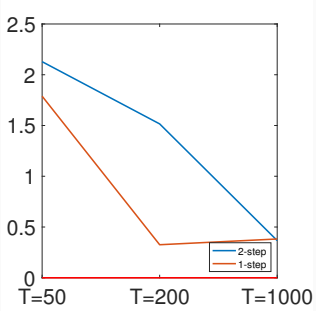
Exp.14



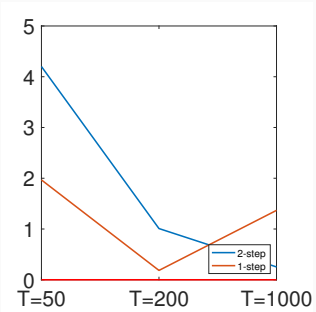
Exp.15



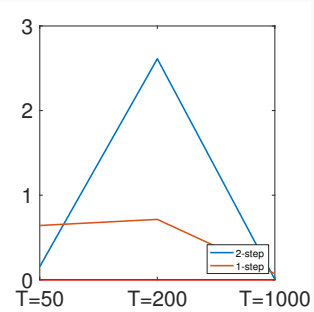
Exp.16



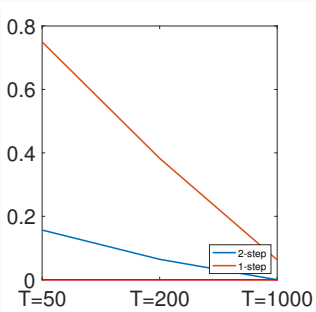
Exp.17



Exp.18

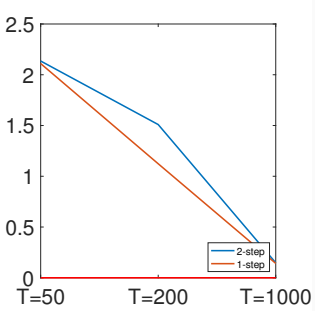


Exp.19

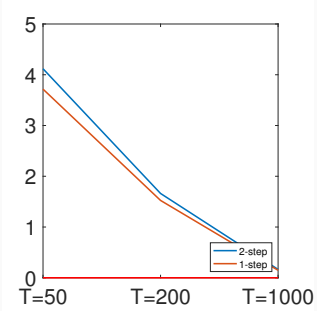


Nested case

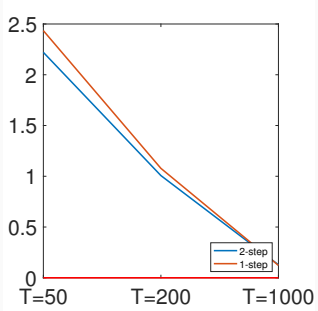
Exp.11



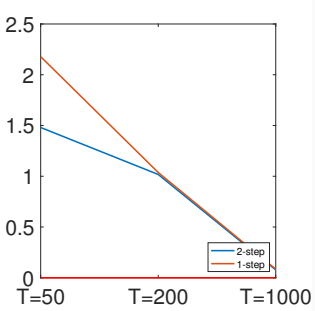
Exp.12



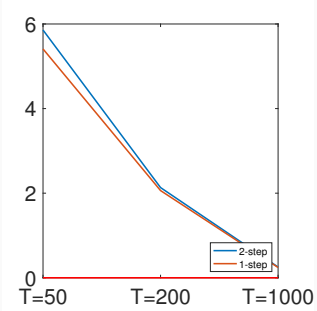
Exp.13



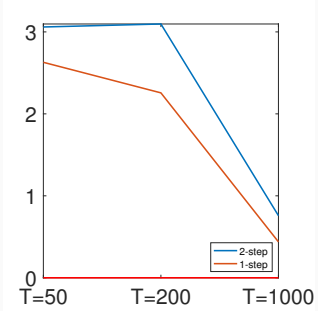
Exp.14



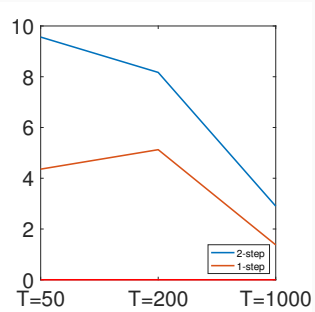
Exp.15



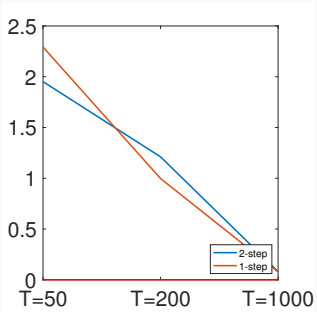
Exp.16



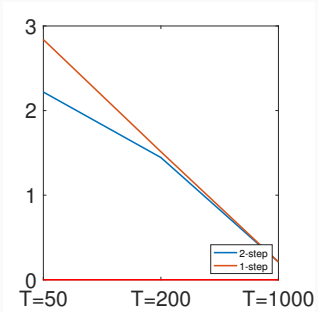
Exp.17



Exp.18



Exp.19



Algorithm for large dataset Application

Start from some initial values of $\boldsymbol{S}^0$ and repeat the following steps M times after a burn-in period long M_0.

1. parameter simulation conditional on the allocation $\boldsymbol{S}^{m-1}$ (as in algorithm (1)):
   - 1.1 Sample η_k from the conditional Dirichlet posterior $p(\eta_k|\boldsymbol{S})$ as in algorithm (1);
   - 1.2 Sample each regression coefficient $\phi = (\phi_{1,0}, \phi_{1,1}, \phi_{2,0}, \phi_{2,1}, \phi_{2,2})$ jointly from the posterior distribution $p(\phi|\sigma_k^2, y_t^o, \boldsymbol{S}^{m-1}) \sim \mathcal{N}(a_k, A_k)$ as in algorithm (1);
   - 1.3 Sample the random hyperparameter C_0 from $p(C_0|\boldsymbol{S}^{m-1}, \boldsymbol{x}_i\phi_k, \sigma_k^2, y_t^o \sim \mathcal{G}(g_N, G_N)$;
   - 1.4 Sample each variance σ_k^2 from the posterior distribution $\sigma_k|\phi, y_t^o, \boldsymbol{S}^{m-1} \sim \mathcal{G}^{-1}(c_k, C_k)$ Where $c_k = c_0 + \frac{N_k}{2}$ and $C_k = C_0 + \frac{1}{2}\epsilon_k'\epsilon$
2. Classification of each observation y_t conditional on $\boldsymbol{\theta}_k$: sample each element of S_i of $\boldsymbol{S}^m$ from the conditional posterior $p(S_t|\phi, \sigma_{\epsilon,k}^2, y_t^o)$ given by:

$$Pr(S_i = k|\phi, \sigma_k^2, y_t^o) \propto \eta_k f_N(y_t^o; \boldsymbol{x}_i\phi_k, \sigma_k^2)$$

The posterior density estimated from the MCMC draws is:

$$p(y_{t+1}|y_t^o, \boldsymbol{\theta}_k) = \frac{1}{M + M_0} \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \eta_k^m p(y_{t+1}|\theta_{k,t+1}^m) \right)$$

back

Preliminary Conclusions

1. Combination of two AR models suggests the superiority of one-step procedure with sufficiently large sample size;
2. In presence of breaks:
   - 2.1 The specification of individual models matters
     - Nonnested case: Two-step is more accurate;
     - Nested case: One-step is more accurate;
   - 2.2 The type of break matters:
     - One-step performs better when the break regards exogenous variables.
   - 2.3 The sample size matters:
     - In the nonnested case, two-step becomes more accurate than one-step
     - In the nested case, one-step becomes more accurate than two-step
   - 2.4 No clear information from the timing of the break.