A Test for State-Dependent Predictive Ability based on a Markov-Switching Framework

Sebastian Fossati

University of Alberta

2018 NBP Workshop on Forecasting, Warsaw

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

Macroeconomic Forecasting

Empirical fact: **predictability is unstable over time**.

- Many individual indicators exhibit predictive content for output growth and inflation but only sporadically
- Stock and Watson (2003), Rossi and Sekhposyan (2010), Rossi (2013), and Granziera and Sekhposyan (2017), among others

New (but related) empirical fact: **predictability varies across states**.

- Inflation rate: Dotsey et al. (2015), Gibbs and Vasnev (2017)
- Output growth: Chauvet and Potter (2013)
- Stock returns: Rapach et al. (2010), Henkel et al.(2011), Dangl et al.(2012)
- Bond excess returns: Gargano et al. (2014)
- Commodity prices: Gargano and Timmermann (2014)

This Paper

Current approach: use **exogenously provided shift dates**.

- NBER recession dates for the US economy

This paper: A test for comparing the out-of-sample forecasting performance of two competing models with state-dependent predictive content.

Main features:

- econometrician not required to observe when underlying states shift
- forecast loss differences modeled using Markov-switching process
- test equal and constant predictive ability vs. state-dependent ability
- heteroskedasticity and autocorrelation consistent (HAC) test

This Paper

Current approach: use **exogenously provided shift dates**.

- NBER recession dates for the US economy

This paper: A test for comparing the out-of-sample forecasting performance of two competing models with state-dependent predictive content.

Main features:

- econometrician not required to observe when underlying states shift
- forecast loss differences modeled using Markov-switching process
- test equal and constant predictive ability vs. state-dependent ability
- heteroskedasticity and autocorrelation consistent (HAC) test

This Paper

Current approach: use **exogenously provided shift dates**.

- NBER recession dates for the US economy

This paper: A test for comparing the out-of-sample forecasting performance of two competing models with state-dependent predictive content.

Main features:

- econometrician not required to observe when underlying states shift
- forecast loss differences modeled using Markov-switching process
- test equal and constant predictive ability vs. state-dependent ability
- heteroskedasticity and autocorrelation consistent (HAC) test

Main Results

Available tests (Diebold-Mariano, Giacomini-White, Giacomini-Rossi, etc.):

- can have low power when predictability varies across economic states
- even when they reject, can lead to incorrect inference

Proposed test:

- performs well with unequal but constant performance
- exhibits more power when predictability varies across states
- estimates when underlying states shift
- HAC estimators work well

Main Results

Available tests (Diebold-Mariano, Giacomini-White, Giacomini-Rossi, etc.):

- can have low power when predictability varies across economic states
- even when they reject, can lead to incorrect inference

Proposed test:

- performs well with unequal but constant performance
- exhibits more power when predictability varies across states
- estimates when underlying states shift
- HAC estimators work well

1 Introduction

2 State-Dependent Predictive Ability
- Environment
- State-dependent loss differences
- Asymptotic predictive ability tests

3 Monte Carlo Evidence
- Appendix - HAC estimators and asymptotic confidence intervals
- Size and power properties

4 Empirical Example: Forecasting Output

5 Conclusion

Environment

Objective: compare sequences of h-step ahead out-of-sample forecasts for y_t obtained from two competing models.

Set up (Giacomini and White, 2006):

- sample of size T, in-sample portion R, out-of-sample portion P
- out-of-sample forecast loss differences
  $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = L^{(1)}(y_t, \hat{\delta}_{t-h,R}) - L^{(2)}(y_t, \hat{\gamma}_{t-h,R})$
- parameters estimated using rolling scheme

Null hypothesis of **equal predictive ability** (Diebold and Mariano, 1995):

$$H_0^{(1)} : E\left[\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})\right] = 0 \text{ for all } t = R + h, \ldots, T$$

Environment

Objective: compare sequences of h-step ahead out-of-sample forecasts for y_t obtained from two competing models.

Set up (Giacomini and White, 2006):

- sample of size T, in-sample portion R, out-of-sample portion P
- out-of-sample forecast loss differences
  $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = L^{(1)}(y_t, \hat{\delta}_{t-h,R}) - L^{(2)}(y_t, \hat{\gamma}_{t-h,R})$
- parameters estimated using rolling scheme

Null hypothesis of **equal predictive ability** (Diebold and Mariano, 1995):

$$H_0^{(1)} : E\left[\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})\right] = 0 \text{ for all } t = R+h, \ldots, T$$

Environment

Objective: compare sequences of h-step ahead out-of-sample forecasts for y_t obtained from two competing models.

Set up (Giacomini and White, 2006):

- sample of size T, in-sample portion R, out-of-sample portion P
- out-of-sample forecast loss differences
  $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = L^{(1)}(y_t, \hat{\delta}_{t-h,R}) - L^{(2)}(y_t, \hat{\gamma}_{t-h,R})$
- parameters estimated using rolling scheme

Null hypothesis of **equal predictive ability** (Diebold and Mariano, 1995):

$$H_0^{(1)} : E\left[\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})\right] = 0 \text{ for all } t = R + h, \ldots, T$$

Unconditional Predictive Ability

Giacomini and White (2006): test of equal **(unconditional) predictive ability**.

Test based on the sample average of loss differences:

$$GW = \hat{\sigma}_P^{-1} P^{1/2} \overline{\Delta L}_P$$

with $\overline{\Delta L}_P = P^{-1} \sum_{t=R+h}^{T} \Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})$, and $\hat{\sigma}_P^2$ the sample variance or a HAC estimator of the long-run variance of $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})$.

Results:

- treat loss differences $\{\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})\}_{t=R+h}^{T}$ as observed data
- under the null hypothesis, $GW \xrightarrow{d} N(0,1)$ as $P \to \infty$
- can be applied to nested and non-nested models

Unstable Environments

Giacomini and Rossi (2010): tests of the joint hypothesis of equal and constant performance of two models against **local predictive ability**.

Fluctuation test based on a rolling average of loss differences:

$$\text{Fluct}_{t,m} = \hat{\sigma}_P^{-1} m^{-1/2} \sum_{j=t-m/2}^{t+m/2+1} \Delta L_j(\hat{\delta}_{j-h,R}, \hat{\gamma}_{j-h,R})$$

for $t = R + h + m/2, \ldots, T - m/2 + 1$, m the window size, and $\hat{\sigma}_P^2$ a HAC estimator of the long-run variance of $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R})$.

Results:

- test rejects when $\max_t |\text{Fluct}_{t,m}| > k_\alpha$
- other tests: Giacomini and Rossi (2010) and Martins and Perron (2016)

State-Dependent Loss Differences

Model out-of-sample forecast loss differences using **Markov-switching** mean plus noise **regression**:

$$\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = \mu_{s_t} + \sigma_{s_t} u_t$$

where s_t $(= 0, 1)$ is an unobserved two-state first-order Markov process with transition probabilities

$$\text{Prob}(s_t = j \,|\, s_{t-1} = i) = p_{ij} \quad i, j = 1, 2$$

and u_t is an unobservable moving-average (MA) process with zero mean and non-zero autocorrelations up to lag $h - 1$. Next, estimate

$\theta = (\mu_0, \mu_1, \sigma_0, \sigma_1, p_{00}, p_{11})'$ by quasi maximum likelihood (Q-ML) following Hamilton (1989), Hamilton (1990), or Kim and Nelson (1999).

Predictive Ability Tests

Giacomini and White (2006): treat loss differences as **observed data**.

Standard **asymptotic normality** arguments: $\sqrt{P}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \Omega\right)$ as $P \to \infty$.

Null hypothesis of **equal and constant predictive ability**: $\mu_0 = \mu_1 = 0$.

Use **Wald test statistic** for linear hypotheses of the form $R_0\theta = 0$:

$$\text{Wald} = P\left(R_0\hat{\theta}\right)'\left(R_0\hat{\Omega}R_0'\right)^{-1}\left(R_0\hat{\theta}\right)$$

where R_0 a $r \times 6$ matrix, r the number of restrictions, and $\hat{\Omega}$ a consistent estimator of Ω.

Under the null hypothesis, Wald $\xrightarrow{d} \chi^2(r)$ as $P \to \infty$.

Predictive Ability Tests

Giacomini and White (2006): treat loss differences as **observed data**.

Standard **asymptotic normality** arguments: $\sqrt{P}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \Omega\right)$ as $P \to \infty$.

Null hypothesis of **equal and constant predictive ability**: $\mu_0 = \mu_1 = 0$.

Use **Wald test statistic** for linear hypotheses of the form $R_0\theta = 0$:

$$\text{Wald} = P\left(R_0\hat{\theta}\right)'\left(R_0\hat{\Omega}R_0'\right)^{-1}\left(R_0\hat{\theta}\right)$$

where R_0 a $r \times 6$ matrix, r the number of restrictions, and $\hat{\Omega}$ a consistent estimator of Ω.

Under the null hypothesis, Wald $\xrightarrow{d} \chi^2(r)$ as $P \to \infty$.

Predictive Ability Tests

Giacomini and White (2006): treat loss differences as **observed data**.

Standard **asymptotic normality** arguments: $\sqrt{P}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \Omega\right)$ as $P \to \infty$.

Null hypothesis of **equal and constant predictive ability**: $\mu_0 = \mu_1 = 0$.

Use **Wald test statistic** for linear hypotheses of the form $R_0\theta = 0$:

$$\text{Wald} = P\left(R_0\hat{\theta}\right)'\left(R_0\hat{\Omega}R_0'\right)^{-1}\left(R_0\hat{\theta}\right)$$

where R_0 a $r \times 6$ matrix, r the number of restrictions, and $\hat{\Omega}$ a consistent estimator of Ω.

Under the null hypothesis, Wald $\xrightarrow{d} \chi^2(r)$ as $P \to \infty$.

Estimators of Ω

Hessian estimator:

$$\hat{\Omega}^{H} = -\left[\frac{1}{P}\sum_{t=R+h}^{T} H_t(\hat{\theta})\right]^{-1}$$

where $\sum_{t=R+h}^{T} H_t(\hat{\theta})$ usually obtained numerically as in Hamilton (1989).

Outer-product estimator:

$$\hat{\Omega}^{OP} = \left[\frac{1}{P}\sum_{t=R+h}^{T} g_t(\hat{\theta})g_t(\hat{\theta})'\right]^{-1}$$

where $g_t(\theta)$ is the score vector for observation t as in Hamilton (1996).

But HAC estimator of the covariance matrix is required:

1. multi-step forecasts (Diebold and Mariano, 1995)
2. instabilities (Morley and Rabah, 2014; Martins and Perron, 2016)

Estimators of Ω

Hessian estimator:

$$\hat{\Omega}^{H} = -\left[\frac{1}{P}\sum_{t=R+h}^{T} H_t(\hat{\theta})\right]^{-1}$$

where $\sum_{t=R+h}^{T} H_t(\hat{\theta})$ usually obtained numerically as in Hamilton (1989).

Outer-product estimator:

$$\hat{\Omega}^{OP} = \left[\frac{1}{P}\sum_{t=R+h}^{T} g_t(\hat{\theta})g_t(\hat{\theta})'\right]^{-1}$$

where $g_t(\theta)$ is the score vector for observation t as in Hamilton (1996).

But HAC estimator of the covariance matrix is required:

1. multi-step forecasts (Diebold and Mariano, 1995)
2. instabilities (Morley and Rabah, 2014; Martins and Perron, 2016)

Estimators of Ω

An alternative estimator, the **'sandwich'** estimator:

$$\hat{\Omega}^S = \left[\frac{1}{P}\sum_{t=R+h}^{T} H_t(\hat{\theta})\right]^{-1} \hat{\Sigma}(\hat{\theta}) \left[\frac{1}{P}\sum_{t=R+h}^{T} H_t(\hat{\theta})\right]^{-1}$$

where Σ is the long-run variance of the score vector (Hayashi, 2000).

A HAC estimate of Σ using the kernel-based estimator:

$$\hat{\Sigma}(\hat{\theta}) = \sum_{j=-(q-1)}^{q-1} k(j/q) P^{-1} \sum_{t=R+h}^{T} g_t(\hat{\theta}) g_{t-j}(\hat{\theta})'$$

where $k(\cdot)$ a kernel weight function, q a bandwidth that grows with P.

For serially uncorrelated scores ($q = 1$), is the same estimator suggested in Hamilton (1996) to calculate robust standard errors.

Estimators of Ω

An alternative estimator, the **'sandwich'** estimator:

$$\hat{\Omega}^S = \left[\frac{1}{P}\sum_{t=R+h}^{T} H_t(\hat{\theta})\right]^{-1} \hat{\Sigma}(\hat{\theta}) \left[\frac{1}{P}\sum_{t=R+h}^{T} H_t(\hat{\theta})\right]^{-1}$$

where Σ is the long-run variance of the score vector (Hayashi, 2000).

A HAC estimate of Σ using the kernel-based estimator:

$$\hat{\Sigma}(\hat{\theta}) = \sum_{j=-(q-1)}^{q-1} k(j/q) P^{-1} \sum_{t=R+h}^{T} g_t(\hat{\theta}) g_{t-j}(\hat{\theta})'$$

where $k(\cdot)$ a kernel weight function, q a bandwidth that grows with P.

For serially uncorrelated scores ($q = 1$), is the same estimator suggested in Hamilton (1996) to calculate robust standard errors.

HAC Estimators

Sequences of loss differences ΔL_t generated assuming:

- $\mu_0 = -2$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$, and $p_{00} = p_{11} = 0.8$

The error term is the serially correlated MA(1) process:

$$u_t = (1 + \theta^2)^{-1/2}(1 + \theta L)\varepsilon_t$$

- $\varepsilon_t \sim \text{i.i.d.N}(0, 1)$

Set up:

- $T = 50, 100, 250$ and $\theta = 0, 0.5, 0.9$
- Q-ML estimates obtained using EM algorithm (Hamilton, 1990)
- covariance matrix of $\hat{\theta}$ estimated using:
  - Hessian (H) and outer-product (OP) estimators
  - 'sandwich' estimator using Bartlett (NW), Parzen (PK), and quadratic spectral (QS) kernels

Appendix: Small sample properties of ML estimates of μ_0

	θ	Bias	SD	H	OP	NW	PK	QS
$T = 100$	0	0.006	0.159	0.154	0.156	0.163	0.162	0.163
				0.933	0.938	0.946	0.942	0.945
	0.5	0.015	0.220	0.161	0.156	0.202	0.210	0.210
				0.846	0.838	0.896	0.909	0.910
	0.9	0.010	0.233	0.163	0.157	0.211	0.219	0.220
				0.824	0.817	0.904	0.908	0.913
$T = 250$	0.9	-0.001	0.135	0.102	0.098	0.129	0.134	0.134
				0.874	0.856	0.932	0.942	0.943

Notes: The true model is $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = \mu_{s_t} + \sigma_{s_t} u_t$ with $s_t = 0, 1$, $\mu_0 = -2$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$, and $p_{00} = p_{11} = 0.8$. The error term is an MA(1) process $u_t = (1 + \theta^2)^{-1/2}(1 + \theta L)\varepsilon_t$ where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$. The covariance matrix estimators are: Hessian (H), outer-product (OP), Bartlett (NW), Parzen (PK), and quadratic spectral kernel (QS). Bias is the average value of $(\hat{\mu}_0^i - \mu_0)$. SD is the square root of the average value of $(\hat{\mu}_0^i - \mu_0)^2$.

Appendix: Small sample properties of ML estimates of μ_0

	θ	Bias	SD	H	OP	NW	PK	QS
$T = 100$	0	0.006	0.159	0.154	0.156	0.163	0.162	0.163
				0.933	0.938	0.946	0.942	0.945
	0.5	0.015	0.220	0.161	0.156	0.202	0.210	0.210
				0.846	0.838	0.896	0.909	0.910
	0.9	0.010	0.233	0.163	0.157	0.211	0.219	0.220
				0.824	0.817	0.904	0.908	0.913
$T = 250$	0.9	-0.001	0.135	0.102	0.098	0.129	0.134	0.134
				0.874	0.856	0.932	0.942	0.943

Notes: The true model is $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = \mu_{s_t} + \sigma_{s_t} u_t$ with $s_t = 0, 1$, $\mu_0 = -2$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$, and $p_{00} = p_{11} = 0.8$. The error term is an MA(1) process $u_t = (1 + \theta^2)^{-1/2}(1 + \theta L)\varepsilon_t$ where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$. The covariance matrix estimators are: Hessian (H), outer-product (OP), Bartlett (NW), Parzen (PK), and quadratic spectral kernel (QS). Bias is the average value of $(\hat{\mu}_0^i - \mu_0)$. SD is the square root of the average value of $(\hat{\mu}_0^i - \mu_0)^2$.

Appendix: Small sample properties of ML estimates of μ_0

	θ	Bias	SD	H	OP	NW	PK	QS
$T = 100$	0	0.006	0.159	0.154	0.156	0.163	0.162	0.163
				0.933	0.938	0.946	0.942	0.945
	0.5	0.015	0.220	0.161	0.156	0.202	0.210	0.210
				0.846	0.838	0.896	0.909	0.910
	0.9	0.010	0.233	0.163	0.157	0.211	0.219	0.220
				0.824	0.817	0.904	0.908	0.913
$T = 250$	0.9	-0.001	0.135	0.102	0.098	0.129	0.134	0.134
				0.874	0.856	0.932	0.942	0.943

Notes: The true model is $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = \mu_{s_t} + \sigma_{s_t} u_t$ with $s_t = 0, 1$, $\mu_0 = -2$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$, and $p_{00} = p_{11} = 0.8$. The error term is an MA(1) process $u_t = (1 + \theta^2)^{-1/2}(1 + \theta L)\varepsilon_t$ where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$. The covariance matrix estimators are: Hessian (H), outer-product (OP), Bartlett (NW), Parzen (PK), and quadratic spectral kernel (QS). Bias is the average value of $(\hat{\mu}_0^i - \mu_0)$. SD is the square root of the average value of $(\hat{\mu}_0^i - \mu_0)^2$.

Appendix: Small sample properties of ML estimates of μ_0

	θ	Bias	SD	H	OP	NW	PK	QS
$T = 100$	0	0.006	0.159	0.154	0.156	0.163	0.162	0.163
				0.933	0.938	0.946	0.942	0.945
	0.5	0.015	0.220	0.161	0.156	0.202	0.210	0.210
				0.846	0.838	0.896	0.909	0.910
	0.9	0.010	0.233	0.163	0.157	0.211	0.219	0.220
				0.824	0.817	0.904	0.908	0.913
$T = 250$	0.9	-0.001	0.135	0.102	0.098	0.129	0.134	0.134
				0.874	0.856	0.932	0.942	0.943

Notes: The true model is $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = \mu_{s_t} + \sigma_{s_t} u_t$ with $s_t = 0, 1$, $\mu_0 = -2$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$, and $p_{00} = p_{11} = 0.8$. The error term is an MA(1) process $u_t = (1 + \theta^2)^{-1/2}(1 + \theta L)\varepsilon_t$ where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$. The covariance matrix estimators are: Hessian (H), outer-product (OP), Bartlett (NW), Parzen (PK), and quadratic spectral kernel (QS). Bias is the average value of $(\hat{\mu}_0^i - \mu_0)$. SD is the square root of the average value of $(\hat{\mu}_0^i - \mu_0)^2$.

Appendix: Small sample properties of ML estimates of μ_0

	θ	Bias	SD	H	OP	NW	PK	QS
$T = 100$	0	0.006	0.159	0.154	0.156	0.163	0.162	0.163
				0.933	0.938	0.946	0.942	0.945
	0.5	0.015	0.220	0.161	0.156	0.202	0.210	0.210
				0.846	0.838	0.896	0.909	0.910
	0.9	0.010	0.233	0.163	0.157	0.211	0.219	0.220
				0.824	0.817	0.904	0.908	0.913
$T = 250$	0.9	-0.001	0.135	0.102	0.098	0.129	0.134	0.134
				0.874	0.856	0.932	0.942	0.943

Notes: The true model is $\Delta L_t(\hat{\delta}_{t-h,R}, \hat{\gamma}_{t-h,R}) = \mu_{s_t} + \sigma_{s_t} u_t$ with $s_t = 0, 1$, $\mu_0 = -2$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$, and $p_{00} = p_{11} = 0.8$. The error term is an MA(1) process $u_t = (1 + \theta^2)^{-1/2}(1 + \theta L)\varepsilon_t$ where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$. The covariance matrix estimators are: Hessian (H), outer-product (OP), Bartlett (NW), Parzen (PK), and quadratic spectral kernel (QS). Bias is the average value of $(\hat{\mu}_0^i - \mu_0)$. SD is the square root of the average value of $(\hat{\mu}_0^i - \mu_0)^2$.

Set-up 1: Unequal but Constant Performance

Data generating process:

$$\begin{aligned} y_t &= \beta x_t + \sigma_\varepsilon \varepsilon_t \\ x_t &= \phi x_{t-1} + \sigma_\nu \nu_t \end{aligned}$$

- ε_t and ν_t are i.i.d.$N(0,1)$, $\phi = .5$, $\sigma_\varepsilon = \sigma_\nu = 1$

The time-t one-step ahead forecasts of y_{t+1}:

$$\begin{aligned} \hat{f}^1_{t,R} &= 0 \\ \hat{f}^2_{t,R} &= \hat{\beta}_{t,R} x_{t+1} \end{aligned}$$

- $\hat{\beta}_{t,R}$ the in-sample estimate of β, x_{t+1} known at time t
- same DGP as Giacomini and White (2006) and Martins and Perron (2016)

Set-ups 2 & 3: Different Performance in Different States

Data generating process:

$$\begin{aligned} y_t &= -\beta s_t + \sigma_\varepsilon \varepsilon_t \\ x_t &= \delta s_t + \sigma_\nu \nu_t \end{aligned}$$

- ε_t and ν_t are i.i.d.$N(0,1)$, $\delta = 1$, $\sigma_\varepsilon = \sigma_\nu = 0.5$

The time-t one-step ahead forecasts of y_{t+1}:

$$\begin{aligned} \hat{f}^1_{t,R} &= 0 \\ \hat{f}^2_{t,R} &= \hat{\gamma}_{t,R} x_{t+1} \end{aligned}$$

- $\hat{\gamma}_{t,R}$ the in-sample rolling estimate of γ, x_{t+1} known at time t
- Set-up 2: $\text{Prob}(s_t = j \,|\, s_{t-1} = i) = p_{ij}$ for $i, j = 0, 1$
- Set-up 3: s_t = actual NBER recession dates 1960Q1–2015Q4 ($T = 224$)

Size and Power

$\beta = 0$: the competing forecasting models are equally accurate in the population, then the smaller model would be preferable in the finite sample.

To have forecasting models **equally accurate in the finite sample**:

$$E\left[\left(y_{t+1} - \hat{f}^1_{t,R}\right)^2\right] = E\left[\left(y_{t+1} - \hat{f}^2_{t,R}\right)^2\right]$$

Size:

- Set-up 1: $\beta^0 \approx 1/\sqrt{R4/3}$
- Set-ups 2 & 3: $\beta^0 \approx \sqrt{\delta^2 \left[1 - \frac{\sigma^2_\nu}{R(\delta^2 p_1 + \sigma^2_\nu)}\right]^{-1} \frac{\sigma^2_\varepsilon}{R(\delta^2 p_1 + \sigma^2_\nu)}}$

Power:

- $\beta = \beta^0 + \beta^+$ with $\beta^+ = 0, .1, \ldots, 2$

Table 2: Empirical size under quadratic loss

R	P	GW	Fluct	SD-W	GW	Fluct	SD-W
		Set-up 1			Set-up 2		
50	50	0.044	0.043	0.084	0.053	0.056	0.093
	100	0.055	0.048	0.044	0.057	0.040	0.045
	250	0.031	0.037	0.034	0.076	0.044	0.055
100	50	0.059	0.043	0.103	0.047	0.053	0.097
	100	0.038	0.045	0.053	0.067	0.062	0.049
	250	0.038	0.051	0.030	0.064	0.030	0.046
					DGP-3		
50	174				0.104	0.053	0.056
100	124				0.081	0.061	0.051
150	74				0.058	0.065	0.062

Notes: The $\text{Fluct}_{t,m}$ test constructed using a window size of $.3P$. Nominal size 0.05.

Table 2: Empirical size under quadratic loss

R	P	GW	Fluct	SD-W	GW	Fluct	SD-W
		Set-up 1			Set-up 2		
50	50	0.044	0.043	0.084	0.053	0.056	0.093
	100	0.055	0.048	0.044	0.057	0.040	0.045
	250	0.031	0.037	0.034	0.076	0.044	0.055
100	50	0.059	0.043	0.103	0.047	0.053	0.097
	100	0.038	0.045	0.053	0.067	0.062	0.049
	250	0.038	0.051	0.030	0.064	0.030	0.046
					DGP-3		
50	174				0.104	0.053	0.056
100	124				0.081	0.061	0.051
150	74				0.058	0.065	0.062

Notes: The $\text{Fluct}_{t,m}$ test constructed using a window size of $.3P$. Nominal size 0.05.

Table 2: Empirical size under quadratic loss

R	P	GW	Fluct	SD-W	GW	Fluct	SD-W
		Set-up 1			Set-up 2		
50	50	0.044	0.043	0.084	0.053	0.056	0.093
	100	0.055	0.048	0.044	0.057	0.040	0.045
	250	0.031	0.037	0.034	0.076	0.044	0.055
100	50	0.059	0.043	0.103	0.047	0.053	0.097
	100	0.038	0.045	0.053	0.067	0.062	0.049
	250	0.038	0.051	0.030	0.064	0.030	0.046
					DGP-3		
50	174				0.104	0.053	0.056
100	124				0.081	0.061	0.051
150	74				0.058	0.065	0.062

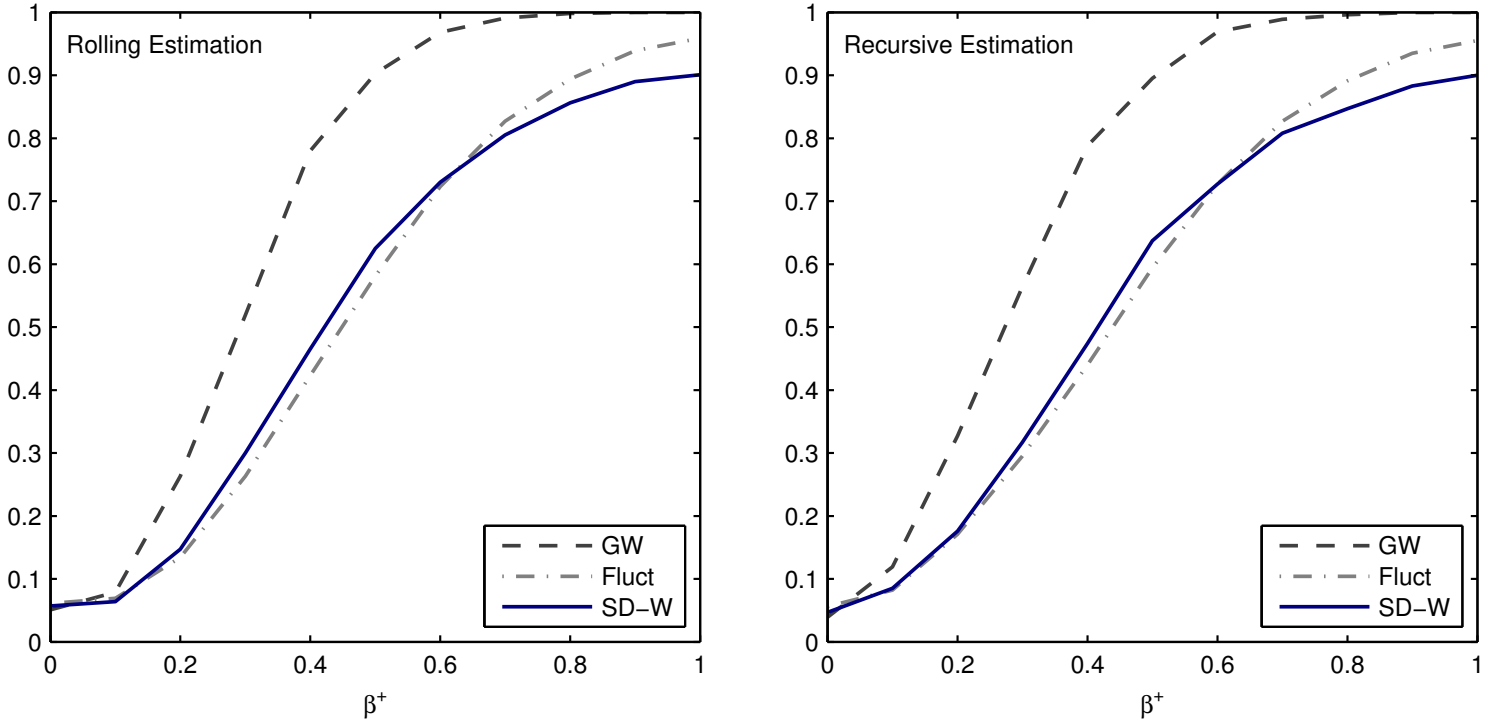
Notes: The $\text{Fluct}_{t,m}$ test constructed using a window size of .3P. Nominal size 0.05.

Table 2: Empirical size under quadratic loss

R	P	GW	Fluct	SD-W	GW	Fluct	SD-W
		Set-up 1			Set-up 2		
50	50	0.044	0.043	0.084	0.053	0.056	0.093
	100	0.055	0.048	0.044	0.057	0.040	0.045
	250	0.031	0.037	0.034	0.076	0.044	0.055
100	50	0.059	0.043	0.103	0.047	0.053	0.097
	100	0.038	0.045	0.053	0.067	0.062	0.049
	250	0.038	0.051	0.030	0.064	0.030	0.046
					DGP-3		
50	174				0.104	0.053	0.056
100	124				0.081	0.061	0.051
150	74				0.058	0.065	0.062

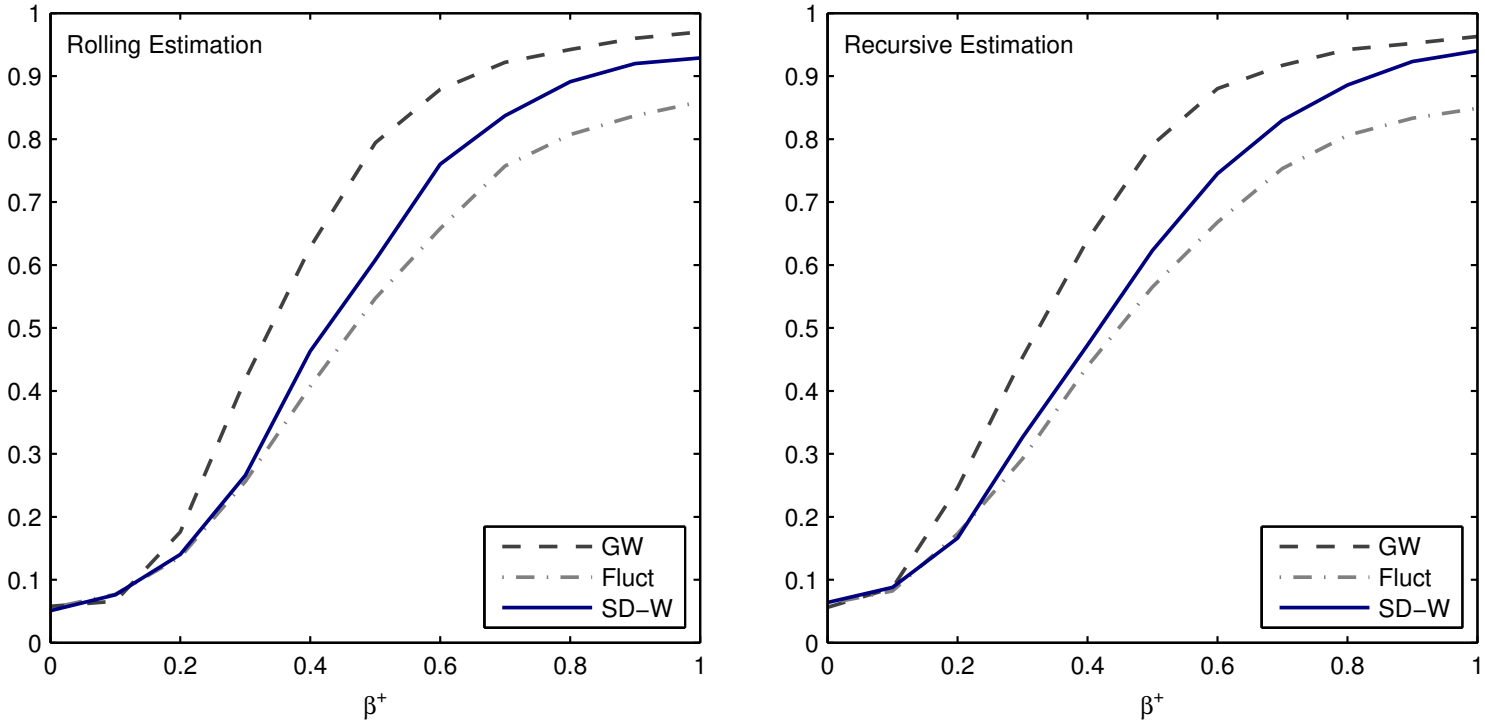
Notes: The $\text{Fluct}_{t,m}$ test constructed using a window size of $.3P$. Nominal size 0.05.

Figure 2: Unequal but constant performance



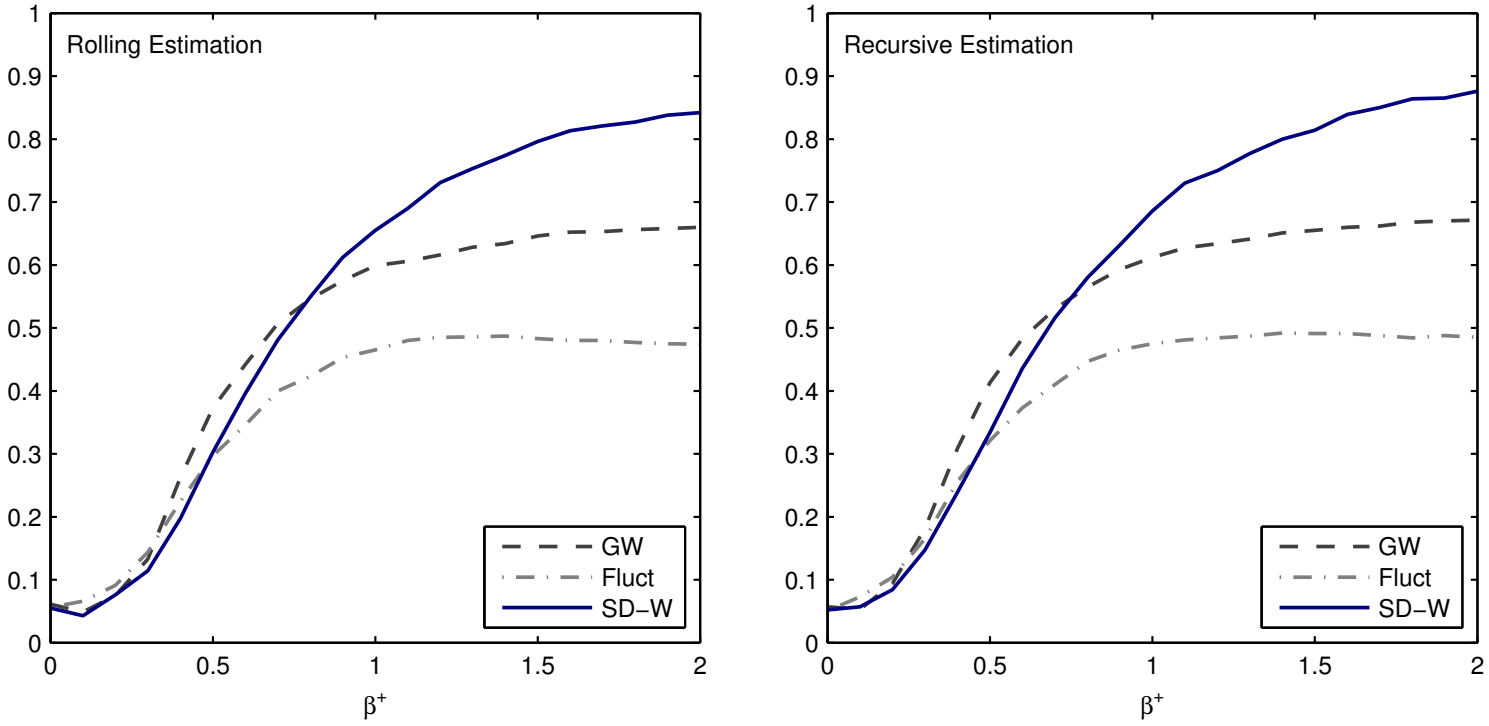
- empirical power (rejection frequencies) of the GW, $\text{Fluct}_{t,m}$ ($m = .3P$), and SD-Wald tests with $R = P = 100$

Figure 3: Different performance in different states (1)



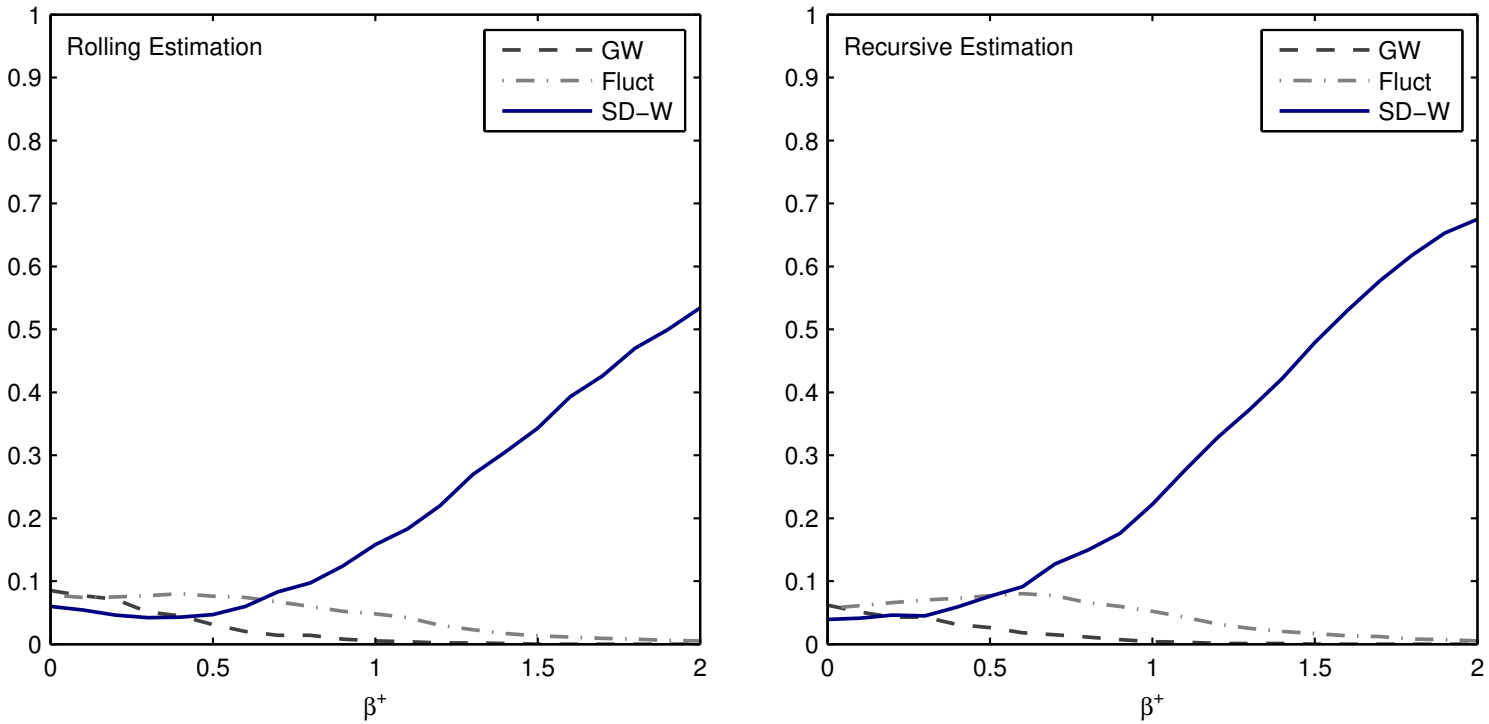
- empirical power (rejection frequencies) of the GW, $\text{Fluct}_{t,m}$ ($m = .3P$), and SD-Wald tests with $p_{00} = p_{11} = 0.8$ and $R = P = 100$

Figure 4: Different performance in different states (2)



- empirical power (rejection frequencies) of the GW, $\text{Fluct}_{t,m}$ ($m = .3P$), and SD-Wald tests with $p_{00} = 0.90$ and $p_{11} = 0.75$ and $R = P = 100$

Figure 5: Different performance in different states (3)



- empirical power (rejection frequencies) of the GW, $\text{Fluct}_{t,m}$ ($m = .3P$), and SD-Wald tests with NBER recession dates, $R = 100$, and $P = 124$

Forecasting Output

Chauvet and Potter (2013): most output growth forecasting models exhibit a similar performance during economic expansions but **one model performs significantly better during recessions**.

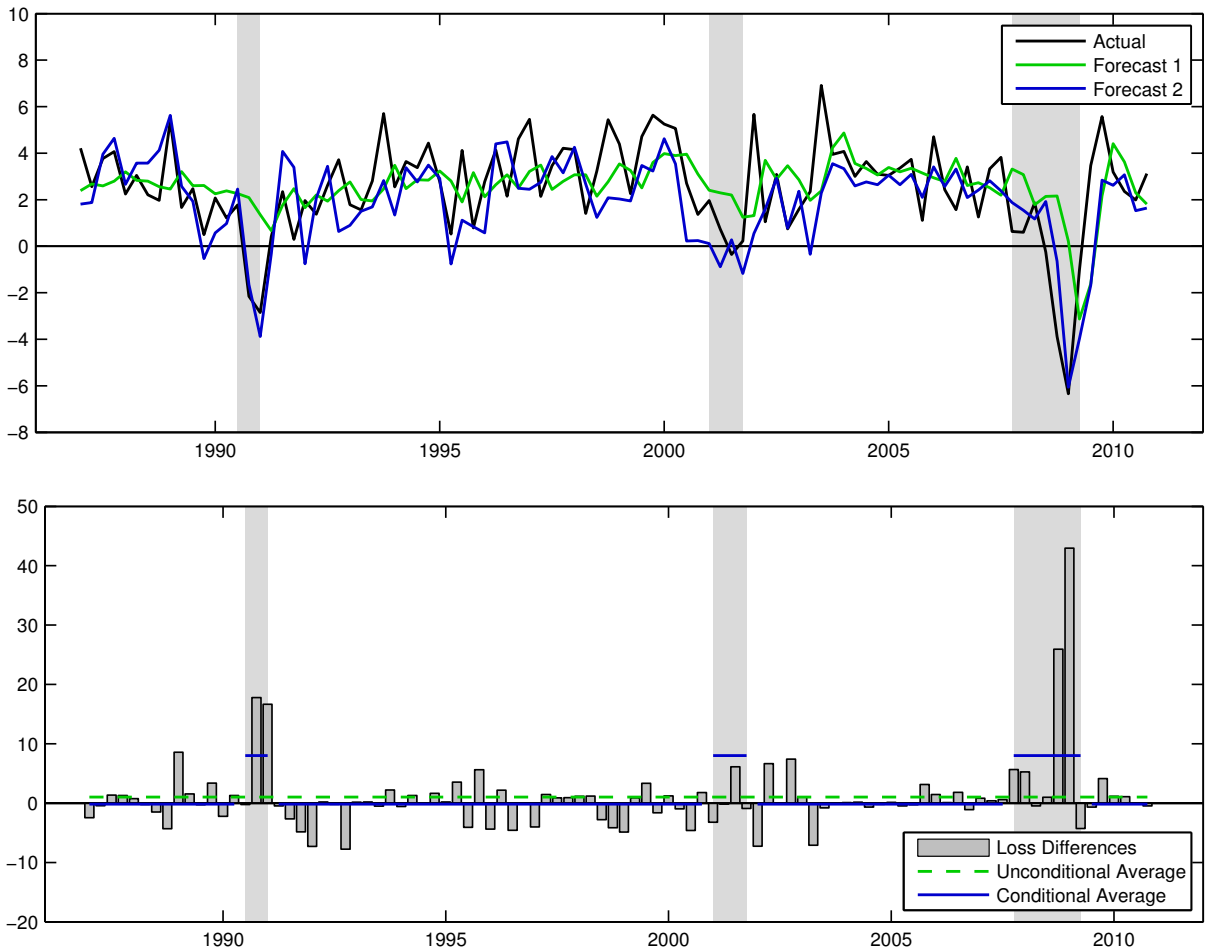
Two competing forecasting models for real GDP growth:

1. Benchmark model: $\hat{f}^1_{t,R} = \text{AR}(2)$
2. 1 + DF-MS model: $\hat{f}^2_{t,R} = \text{AR}(2) + \text{DF} + \text{prob of recession}$

Other details:

- one-step ahead out-of-sample forecasts
- parameters estimated using rolling window
- forecast comparisons using real-time data

Figure 6: Real GDP growth and forecasts



- real GDP growth and forecasts (top), loss differences (bottom)

Table 3: Loss difference statistics for GDP growth

	OOS	OOS_0	OOS_1
Observations	96	82	14
Average	1.032	-0.159	8.005
Standard Dev.	6.388	3.019	13.422
AR(1)	0.279**	-0.136	0.121
GW	1.582	-0.477	2.232**
GW(HAC)	1.360		
Fluctuation(HAC, $m = .1P$)	3.517**		
Fluctuation(HAC, $m = .3P$)	2.232		
SD-Wald(HAC)	16.484**		

Notes: Significance of the AR(1) coefficients is tested based on the asymptotic result $\sqrt{T}\hat{\rho}_1 \xrightarrow{d} N(0,1)$. The 5% (10%) critical value is 1.96 (1.645) for a two-sided GW test, 3.012 (2.766) for a two-sided Fluctuation test with a rolling window size of $m = .3P$, and 3.393 (3.170) for $m = .1P$. The 5% (10%) critical value for a SD-Wald test is 5.99 (4.61). ** (*) denotes rejection of the null hypothesis at the 5% (10%) level.

Table 3: Loss difference statistics for GDP growth

	OOS	OOS_0	OOS_1
Observations	96	82	14
Average	1.032	-0.159	8.005
Standard Dev.	6.388	3.019	13.422
AR(1)	0.279**	-0.136	0.121
GW	1.582	-0.477	2.232**
GW(HAC)	1.360		
Fluctuation(HAC, $m = .1P$)	3.517**		
Fluctuation(HAC, $m = .3P$)	2.232		
SD-Wald(HAC)	16.484**		

Notes: Significance of the AR(1) coefficients is tested based on the asymptotic result $\sqrt{T}\hat{\rho}_1 \xrightarrow{d} N(0,1)$. The 5% (10%) critical value is 1.96 (1.645) for a two-sided GW test, 3.012 (2.766) for a two-sided Fluctuation test with a rolling window size of $m = .3P$, and 3.393 (3.170) for $m = .1P$. The 5% (10%) critical value for a SD-Wald test is 5.99 (4.61). ** (*) denotes rejection of the null hypothesis at the 5% (10%) level.

Table 3: Loss difference statistics for GDP growth

	OOS	OOS_0	OOS_1
Observations	96	82	14
Average	1.032	-0.159	8.005
Standard Dev.	6.388	3.019	13.422
AR(1)	0.279**	-0.136	0.121
GW	1.582	-0.477	2.232**
GW(HAC)	1.360		
Fluctuation(HAC, $m = .1P$)	3.517**		
Fluctuation(HAC, $m = .3P$)	2.232		
SD-Wald(HAC)	16.484**		

Notes: Significance of the AR(1) coefficients is tested based on the asymptotic result $\sqrt{T}\hat{\rho}_1 \xrightarrow{d} N(0,1)$. The 5% (10%) critical value is 1.96 (1.645) for a two-sided GW test, 3.012 (2.766) for a two-sided Fluctuation test with a rolling window size of $m = .3P$, and 3.393 (3.170) for $m = .1P$. The 5% (10%) critical value for a SD-Wald test is 5.99 (4.61). ** (*) denotes rejection of the null hypothesis at the 5% (10%) level.

Table 3: Loss difference statistics for GDP growth

	OOS	OOS_0	OOS_1
Observations	96	82	14
Average	1.032	-0.159	8.005
Standard Dev.	6.388	3.019	13.422
AR(1)	0.279**	-0.136	0.121
GW	1.582	-0.477	2.232**
GW(HAC)	1.360		
Fluctuation(HAC, $m = .1P$)	3.517**		
Fluctuation(HAC, $m = .3P$)	2.232		
SD-Wald(HAC)	16.484**		

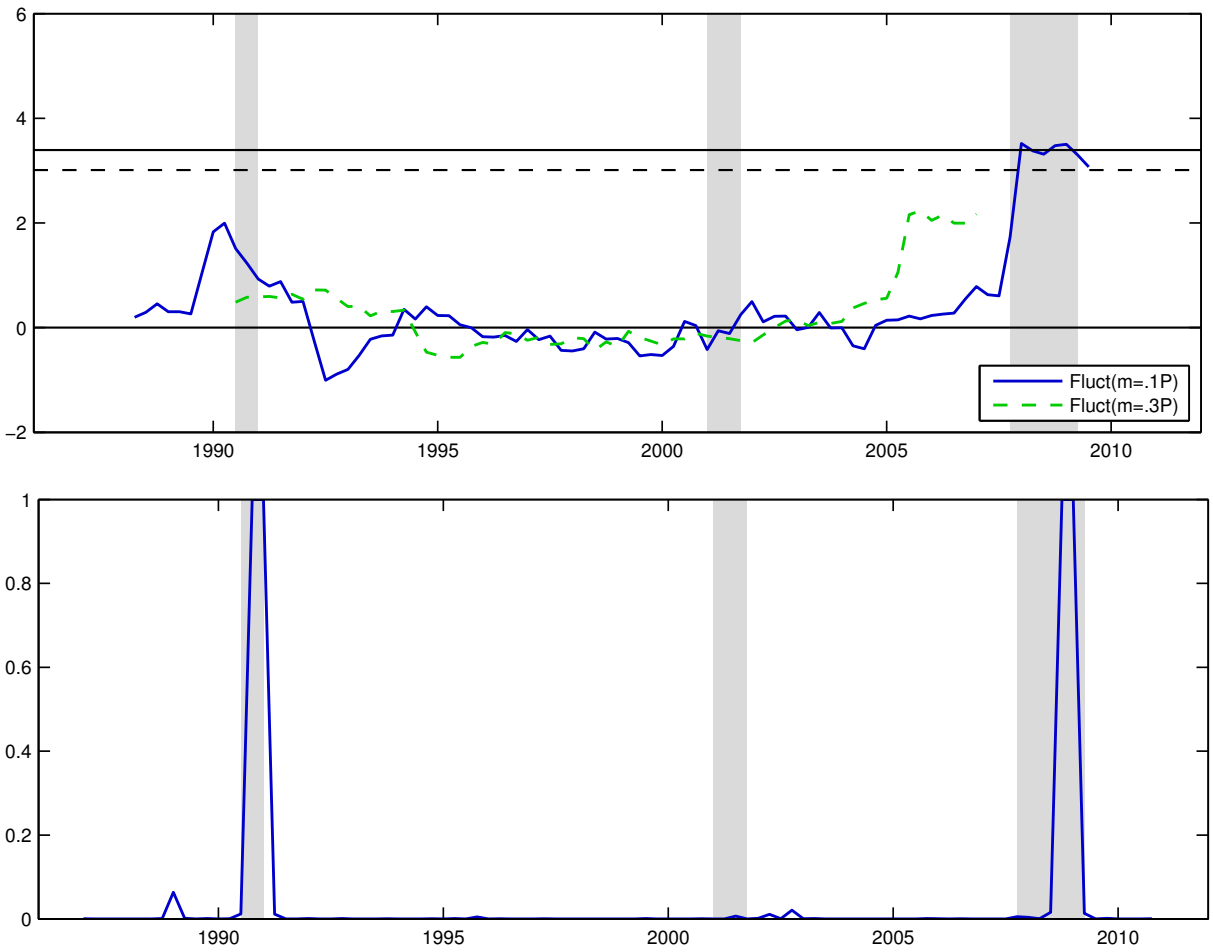
Notes: Significance of the AR(1) coefficients is tested based on the asymptotic result $\sqrt{T}\hat{\rho}_1 \xrightarrow{d} N(0,1)$. The 5% (10%) critical value is 1.96 (1.645) for a two-sided GW test, 3.012 (2.766) for a two-sided Fluctuation test with a rolling window size of $m = .3P$, and 3.393 (3.170) for $m = .1P$. The 5% (10%) critical value for a SD-Wald test is 5.99 (4.61). ** (*) denotes rejection of the null hypothesis at the 5% (10%) level.

Table 3: Loss difference statistics for GDP growth

	OOS	OOS_0	OOS_1
Observations	96	82	14
Average	1.032	-0.159	8.005
Standard Dev.	6.388	3.019	13.422
AR(1)	0.279**	-0.136	0.121
GW	1.582	-0.477	2.232**
GW(HAC)	1.360		
Fluctuation(HAC, $m = .1P$)	3.517**		
Fluctuation(HAC, $m = .3P$)	2.232		
SD-Wald(HAC)	16.484**		

Notes: Significance of the AR(1) coefficients is tested based on the asymptotic result $\sqrt{T}\hat{\rho}_1 \xrightarrow{d} N(0,1)$. The 5% (10%) critical value is 1.96 (1.645) for a two-sided GW test, 3.012 (2.766) for a two-sided Fluctuation test with a rolling window size of $m = .3P$, and 3.393 (3.170) for $m = .1P$. The 5% (10%) critical value for a SD-Wald test is 5.99 (4.61). ** (*) denotes rejection of the null hypothesis at the 5% (10%) level.

Figure 7: Test statistics and probabilities of $s_t = 1$



- Fluctuation test statistics (top), probabilities of $s_t = 1$ (bottom)

Forecasting Output

The parameter estimates (s.e. in parentheses) of the test equation are:

- $\hat{\mu}_0 = 24.72$ (6.10), $\hat{\sigma}_0^2 = 129.13$ (57.90), $\hat{p}_{00} = 0.48$ (0.25)
- $\hat{\mu}_1 = -0.06$ (0.42), $\hat{\sigma}_1^2 = 9.37$ (1.70), $\hat{p}_{11} = 0.98$ (0.02)

The results suggest:

- a regime in which the AR-DF model is more accurate ($\hat{\mu}_0 > 0$)
- a regime in which the two models exhibit a similar performance ($\hat{\mu}_1 \approx 0$)
- a substantial difference in variance across regimes with $\hat{\sigma}_0^2/\hat{\sigma}_1^2 \approx 13.78$

Conclusion

- Tests of equal predictive ability based on averages (unconditional or rolling) are inadequate when the predictive content is constrained to just a few observations (see also Casini, 2018)
- I proposed a new test for comparing the out-of-sample forecasting performance of two models for situations in which the predictive content may be state-dependent and of short duration
- The results could be used to improve the overall accuracy of macroeconomic forecasts:
  - Chauvet and Potter (2013), Kotchoni and Stevanovic (2018)
  - Gibbs and Vasnev (2017)