Survey density forecast comparison in small samples

Laura Coroneo (University of York), **Fabrizio Iacone** (University of Milan) and **Fabio Profumo** (University of York)

Abstract

We apply fixed-b and fixed-m asymptotics to tests of equal predictive accuracy and of encompassing for survey density forecasts. We verify in an original Monte Carlo design that fixed-smoothing asymptotics delivers correctly sized tests in this framework, even when only a small number of out of sample observations is available. We use the proposed density forecast comparison tests with fixed-smoothing asymptotics to assess the predictive ability of density forecasts from the European Central Bank's Survey of Professional Forecasters (ECB SPF). We find an improvement in the predictive ability of the ECB SPF since 2010, suggesting a change in the forecasting practice after the financial crisis.

Motivation

Traditional inference methods suffer from small sample size distortions, which can lead to spurious results, Clark (1999).

This paper

We apply fixed-b and fixed-m asymptotics to address the small sample bias of two density forecast comparison tests:

1. Equal predictive accuracy (Diebold and Mariano, 1995)
2. Forecast encompassing (Harvey, Leybourne, and Newbold, 1998)

Density forecast comparison

The h-step ahead density forecast i is given by

$$f_{t,i}^{k} = P_{t-h,i}(y_t \in k),\ k = 1, \dots, K$$

and the realisation by

$$y_t^k = I(y_t \in k),\ k = 1, \dots, K.$$

Denoting $\mathbf{f}_{t,i} = [f_{t,i}^1, \dots, f_{t,i}^k, \dots, f_{t,i}^K]'$ and $\mathbf{y}_t = [y_t^1, \dots, y_t^k, \dots, y_t^K]'$, the forecast error is

$$\mathbf{e}_{t,i} = \mathbf{y}_t - \mathbf{f}_{t,i}$$

and the cumulative (over k) error is $\mathbf{E}_{t,i} = \mathbf{Y}_t - \mathbf{F}_{i,t}$.

Loss functions for forecasts reported as histograms:

- Quadratic Probability Score (QPS) by Brier (1950)

$$QPS_{t,i} = \sum_{k=1}^{K} (y_t^k - f_{t,i}^k)^2 = \mathbf{e}_{t,i}'\mathbf{e}_{t,i}$$

- Ranked Probability Score (RPS) by Epstein (1969)

$$RPS_{t,i} = \sum_{k=1}^{K} (Y_t^k - F_{t,i}^k)^2 = \mathbf{E}_{t,i}'\mathbf{E}_{t,i}$$

Equal predictive accuracy test

Null hypothesis: two alternative forecasts have equal forecasting accuracy according to a user-chosen loss function.

Denote by L^i the loss function for $i = 1, 2$, so that $L_t^i = QPS_{t,i}$ or $L_t^i = RPS_{t,i}$, and the loss differential by

$$d_t = L_t^1 - L_t^2, \tag{1}$$

the null hypothesis of equal forecasting ability is

$$H_0 : \{E(d_t) = 0\}. \tag{2}$$

Forecast encompassing test

Null hypothesis: one forecast encompasses another one, in the sense that the accuracy of the (encompassing) forecast $\mathbf{f}_{t,1}$ cannot be improved through a linear combination with the (encompassed) forecast $\mathbf{f}_{t,2}$. So in the density forecast combination

$$\mathbf{f}_{t,c}(\lambda) = (1 - \lambda)\mathbf{f}_{t,1} + \lambda \mathbf{f}_{t,2}$$

the optimal weight associated with forecast $\mathbf{f}_{t,2}$ in the QPS (or RPS) sense is equal to zero.

We show that if we define d_t as

$$d_t = \begin{cases} \mathbf{e}_{t,1}'(\mathbf{e}_{t,1} - \mathbf{e}_{t,2}), \text{for } QPS, \\ \mathbf{E}_{t,1}'(\mathbf{E}_{t,1} - \mathbf{E}_{t,2}), \text{for } RPS, \end{cases} \tag{3}$$

then the null of density forecast encompassing can be expressed as

$$H_0 : \{E(d_t) = 0\}$$

against the one-sided alternative $E(d_t) > 0$ (i.e., $\lambda > 0$).

Diebold-Mariano framework

Denote $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$ and $\sigma_T^2 = var(\sqrt{T}\,\bar{d})$, then under H_0 and regularity conditions

$$\sqrt{T}\frac{\bar{d}}{\sigma_T} \to_d N(0, 1)$$

This test statistic is unfeasible as σ_T is unknown, but this may be replaced by an estimate, say $\hat{\sigma}$. If the latter is consistent, the feasible statistic obtained in this way retains the standard normal limiting distribution.

Consistent estimates for σ_T^2:

- **Weighted Covariance Estimate (WCE)**

$$\hat{\sigma}_{WCE}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} k(j/M)\hat{\gamma}_j$$

where $k(.)$ is a kernel function and M is a bandwidth parameter such that $M \to \infty$ and $M/T \to 0$ as $T \to \infty$.

- **Weighted Periodogram Estimate (WPE)**

$$\hat{\sigma}_{WPE}^2 = 2\pi \sum_{j=1}^{T/2} K_m(\lambda_j) I(\lambda_j)$$

where $K_M(.)$ is a symmetric kernel function and m a truncation parameter such that $m \to \infty$ and $m/T \to 0$ as $T \to \infty$.

The DM framework is subject to severe size distortion in small and medium-sized samples, as documented in Clark (1999). This is due to the fact that, in any finite sample, the ratio M/T (or m/T) is still non-zero, and in a moderate size sample this ratio may be non-negligible.

Fixed-smoothing asymptotics

Assumption 1: Partial sums of d_t are such that the functional central limit theorem (FCLT) holds

$$\frac{\sqrt{T}}{T}\frac{1}{\sigma_T}\sum_{t=1}^{\lfloor rT \rfloor} d_t \Rightarrow W(r)$$

where $\lfloor . \rfloor$ denotes the integer part of a number, $r \in [0, 1]$ and $W(r)$ is a standard Brownian motion.

Fixed-b asymptotic (Kiefer and Vogelsang, 2005)

Under H_0 and Assumption 1, when $M/T \to b \in (0, 1]$ as $T \to \infty$,

$$\sqrt{T}\frac{\bar{d}}{\hat{\sigma}_{WCE-B}} \to_d \Phi^B(b)$$

$\hat{\sigma}_{WCE-B}^2$ uses the triangular (Bartlett) kernel and $\Phi^B(b)$ is characterised in Kiefer and Vogelsang (2005).

Fixed-m asymptotic (Hualde and Iacone, 2017)

Under H_0 and Assumption 1, for m constant as $T \to \infty$,

$$\sqrt{T}\frac{\bar{d}}{\hat{\sigma}_{WPE-D}} \to_d t_{2m}$$

where $\hat{\sigma}_{WPE-D}$ uses the (rectangular) Daniell kernel.

Notice that with fixed-smoothing asymptotic, the limiting distribution depends on the kernel and the bandwidth used.

Simulation studies find that fixed-smoothing asymptotics yield better approximation of the empirical size, and that this improvement is stronger the larger is the bandwidth M (the smaller is m). These works also find that the finite sample power is decreasing with the bandwidth, therefore documenting the existence of a trade-off between correct size and power.

Monte Carlo study of size and power

We study the size and power properties for different bandwidths in a Monte Carlo exercise (we refer to the paper for a detailed description). We use a sample that replicates the dimension of the sample of our dataset. The autocorrelation in d_t is measured by a tuning parameter Q.

Table 1: Empirical size of the equal predictive accuracy test

Standard asymptotics

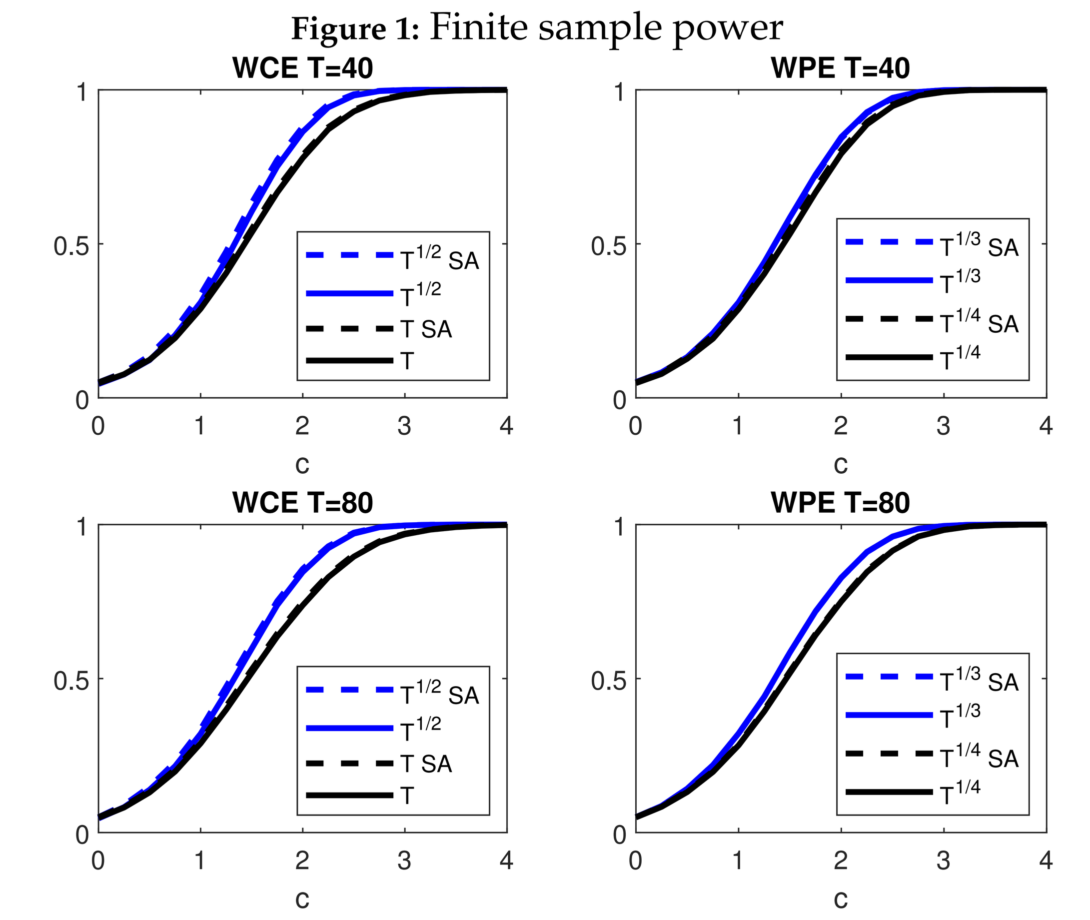
		WCE			WPE			
T	Q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
40	0	0.067	0.079	0.204	0.089	0.075	0.064	0.061
	2	0.102	0.098	0.218	0.084	0.078	0.082	0.117
	4	0.159	0.129	0.234	0.101	0.096	0.130	0.192
	6	0.197	0.152	0.244	0.109	0.117	0.176	0.232
80	0	0.060	0.068	0.201	0.085	0.069	0.058	0.056
	2	0.082	0.080	0.206	0.086	0.067	0.066	0.092
	4	0.117	0.098	0.220	0.090	0.077	0.087	0.160
	6	0.148	0.111	0.221	0.092	0.083	0.112	0.194

Fixed-smoothing asymptotics

		WCE			WPE			
T	Q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
40	0	0.047	0.045	0.049	0.052	0.051	0.051	0.055
	2	0.076	0.058	0.057	0.047	0.051	0.067	0.108
	4	0.129	0.083	0.070	0.058	0.069	0.113	0.183
	6	0.166	0.105	0.079	0.066	0.082	0.157	0.222
80	0	0.048	0.047	0.048	0.050	0.051	0.050	0.053
	2	0.067	0.053	0.052	0.050	0.049	0.057	0.085
	4	0.100	0.072	0.059	0.054	0.058	0.077	0.152
	6	0.131	0.082	0.061	0.053	0.062	0.100	0.188

Note: the theoretical size is 5%, for a one-sided alternative hypothesis.

Figure 1: Finite sample power



Note: the theoretical size is 5% and a one-sided alternative hypothesis. The dashed lines refer to power performances using size-adjusted critical values while solid lines use fixed-smoothing asymptotics. The parameter c indicates the distance from the null hypothesis.

We suggest $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$, in line with Lazarus, Lewis, Stock and Watson (2018).

Application

We analyse the ECB SPF aggregate density forecasts at the rolling horizons of one and two years for HICP inflation, the unemployment rate and real GDP growth, for the surveys conducted between 2000.Q1 and 2019.Q4, corresponding to a total of 80 quarterly observations. We also split the sample in two equally sized subsamples: 2000.Q1-2009.Q4 and 2010.Q1-2019.Q4, of 40 observations each.

Benchmarks:

- Unconditional Gaussian (UG) density forecast
- Gaussian random walk (GRW) density forecast
- Naive density forecast: last available ECB SPF density forecast for the same horizon, i.e. $f_t^{k,Naive} = f_{t-1}^{k,SPF}$.

Equal Predictive Ability Test, Subsample Q1.2000 - Q4.2009

			1 year ahead			2 years ahead		
Variable	Loss	LRV	UG	GRW	Naïve	UG	GRW	Naïve
UN	RPS	WCE	3.52**	3.22**	2.80**	0.81	1.25	2.27**
		WPE	2.98**	2.90**	2.37**	0.71	1.25	1.87*
	QPS	WCE	2.69**	1.34	2.15**	0.89	0.20	2.31**
		WPE	2.25**	1.08	1.93*	0.78	0.16	2.15**
GDP	RPS	WCE	1.34	1.61*	2.20**	-2.21	1.24	1.04
		WPE	1.10	1.46*	2.11**	-1.98	1.12	1.31
	QPS	WCE	0.17	1.00	1.36	-4.24	-1.05	0.63
		WPE	0.14	0.81	1.35	-4.07	-0.92	0.62
HCPI	RPS	WCE	-0.46	1.62*	0.90	-0.83	1.85*	-0.27
		WPE	-0.38	1.47*	0.83	-0.71	1.53*	-0.26
	QPS	WCE	-1.47	1.40	-0.54	-0.71	1.56*	-0.49
		WPE	-1.30	1.11	-0.49	-0.58	1.51*	-0.48

Note: a negative sign implies that the benchmark performs better than the ECB SPF. [dark shading] and [light shading] indicate, respectively, one-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, one-sided significance at the 5% and 10% level.

Forecast Encompassing Test, Subsample Q1.2010 - Q4.2019

			1 year ahead			2 years ahead		
Variable	Loss	LRV	UG	GRW	Naïve	UG	GRW	Naïve
UN	RPS	WCE	0.15	-0.51	-4.80	0.46	-0.50	-1.84
		WPE	0.12	-0.45	-4.79	0.41	-0.46	-1.67
	QPS	WCE	1.63*	1.67*	-4.18	1.64*	2.58**	-1.41
		WPE	1.44*	1.53*	-4.32	1.45*	2.34**	-1.31
GDP	RPS	WCE	-0.50	-0.71	-2.25	1.35	1.54*	-0.53
		WPE	-0.42	-1.02	-2.12	1.28	1.25	-0.44
	QPS	WCE	1.04	1.15	-0.91	1.39	1.35	1.33
		WPE	0.97	1.14	-0.73	1.30	1.24	1.26
HCPI	RPS	WCE	-0.08	1.72*	-1.76	-0.33	0.03	-1.63
		WPE	-0.07	1.39	-1.41	-0.28	0.02	-1.65
	QPS	WCE	2.41**	2.59**	-0.66	2.00**	1.22	-0.82
		WPE	1.96**	2.10**	-0.72	1.78*	0.99	-0.91

Note: a negative value indicates that the unrestricted weight on the benchmark is negative. [dark shading] and [light shading] indicate, respectively, one-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, one-sided significance at the 5% and 10% level.

Conclusions

- We apply fixed-b and fixed-m asymptotics to address the small sample bias of density forecast evaluation tests for equal predictive accuracy and encompassing.
- Monte Carlo evidence shows that fixed smoothing asymptotics are reliable for density forecast comparison in small samples
- Fixed smoothing asymptotics allow to assess the accuracy of the ECB SPF density forecasts before and after the financial crisis.