Hurn, Stan, Tian, Jing, & Xu, Lina
(2021)
Assessing the Informational Content of Official Australian Bureau of Meteorology Forecasts of Wind Speed.
Economic Record, *97*(319), pp. 525-547.

https://doi.org/10.1111/1475-4932.12627

Assessing the informational content of official Australian Bureau of Meteorology forecasts of wind speed*

Stan Hurn
Queensland University of Technology
Jing Tian
University of Tasmania
Lina Xu
Queensland University of Technology

This Version: June 2021

Abstract

Understanding forecast revisions is critical for weather forecast users to determine the optimal timing for their planning decisions. A set of multi-horizon forecasts for wind speed produced by the Australian Bureau of Meteorology for 12 weather stations in eastern Australia are examined. The forecasts are examined in terms of the econometric definition of rationality and, as a robustness check, the economic value of the forecasts is also assessed using a cost-loss model. It is demonstrated that while the forecasts exhibit some of the characteristics of rational forecasts, when official testing is introduced forecast rationality is rejected at all the weather stations considered. Furthermore, the behaviour of the forecasts is shown to be very erratic over the course of the day and over forecast horizons. Although there is some evidence that the official forecasts can provide positive economic value, this metric also indicates that there is substantial room for improvement.

wind speed; multi-horizon forecasts; forecast rationality; economic value.

JEL classification: Q41; Q48.

*Corresponding author: <jing.tian.@utas.edu.au>. The authors thank two referees and an associate editor for useful comments on an earlier draft. Thanks are also due to Tom Pagano and Deryn Griffiths of the Australian Bureau of Meteorology for providing the wind speed forecast data and also for valuable interactions over the course of the research project. The views expressed here are, however, those of the authors and in no way reflect the views of the Australian Bureau of Meteorology.

1 Introduction

The Bureau of Meteorology (BOM) is Australia's national weather, climate and water agency. Through regular forecasts, warnings, monitoring and advice spanning the Australian region and Antarctic territory, the Bureau provides one of the most fundamental and widely used services of government. Of particular interest to this research are forecasts of hourly wind speed at weather stations across Australia which are made by the BOM up to one week in advance, and the public updates of these forecasts are provided once each day before the target date.

The reliability of forecasts across various forecast horizons is of paramount importance and cannot be overstated. Two important examples of the importance of forecasts to end users are in bushfire management and the energy sector. Forecasts of wind speed[1] are crucial to the control and management of bush fires including when and how to deploy fire fighting resources most efficiently (García-Portugués et al., 2014). The energy sector relies on forecasts to determine optimal bidding of generation, timing of maintenance and storage of gas for peaking plant production in the event of wind drought or shadow (Soman et al., 2010; Milligan et al., 1995).

Despite the significant economic importance of accurate wind forecasts, there has been no rigorous econometric evaluation of the reliability of the official BOM forecasts.[2] Lynch et al. (2014) calculate the correlation coefficient of wind speed forecast deviations from the seasonal and diurnal cycles in order to evaluate the skill inherent in forecasting anomalies. Pinson & Hagedorn (2012) compare ensemble wind speed forecasts with a benchmark that is derived using kernel density estimation. Similar to other weather forecasts, the precision of wind speed forecasts is often measured by accuracy scores such as the root mean squared error (RMSE) and mean

[1] In addition to wind speed, it should be noted that forecasts of wind direction, temperature and humidity are also important in practical settings. The focus in this paper, however, is entirely on forecasts of wind speed and these forecasts are evaluated in a univariate setting rather than in a multivariate context involving other weather variables.

[2] There is, however, substantial interest in forecast evaluation in terms of weather forecasting generally. McLay (2011) builds a dynamic decision model conditional on variability of revised weather forecasts rather than accuracy of revised forecasts. Motivated by a simple "flip-flop" decision rule, Griffiths et al. (2019) study the stability of a sequence of weather forecasts. Smith (2016) presents decision-makers in the energy sector with an example of how to utilise information contents in a sequence of revised weather forecasts that is known to be inaccurate.

absolute error (MAE), see for example, Sweeney et al. (2013), Pinson & Hagedorn (2012) and Xie et al. (2006) .

This paper addresses the problem of forecast evaluation from an econometric perspective and uses a suite of methods designed to test for the econometric rationality of point forecasts of wind speed. In essence, rational multi-horizon forecasts require that any revision to forecasts should incorporate newly available information efficiently and the literature in this area is well developed (Lovell, 1986; Nordhaus, 1987; Clements, 1997; Clements & Taylor, 2001; Patton & Timmermann, 2012). The official wind speed forecasts can be interpreted in multiple ways due to a two-stage production process that the BOM adopts for producing their official weather forecasts [3]. The forecasts are first computed using a global numerical weather prediction model, namely the Australian Climate Community Earth Systems Simulator or ACCESS model, and are then examined by meteorologists who use their knowledge and experience of local weather features to post-process the forecasts. By testing the rationality of wind speed forecast revisions, this paper sheds light on the combination of the efficacy of a large-scale numerical weather model together with the skill of the BOM's meteorologists in processing the information relevant to future local weather conditions.

As an additional robustness check, the paper also provides an evaluation of the wind speed forecasts from an end-user perspective using a cost-loss model (Murphy, 1977; Richardson, 2000; Foley & Loveday, 2020). This paper assumes that wind speed forecast users are primarily concerned about two wind anomaly events, namely, wind droughts and wind gusts. In order to assess the value of official forecasts in this context, point forecasts of wind speed are converted into event forecasts, that is, forecasts of the occurrence of each of the anomaly events. In this way an attempt is made to relate the evaluation of the forecasts to an economic loss function as opposed to the purely statistical metrics used in the tests of econometric rationality. The decision-based approach for forecast evaluation has been considered not only by meteorologists but also by economists. See, for example, Granger & Pesaran (2000a,b) and Pesaran & Skouras

[3] Details on how the BOM forecasts weather can be found at `http://media.bom.gov.au/social/blog/1696/explainer-how-meteorologists-forecast-the-weather/`.

(2004) for discussion of the links between statistical and economic measures of forecast accuracy.

The forecast evaluation is carried out using the offical BOM forecasts of hourly wind speed made at daily horizons out to 7 days. Wind speed forecasts of BOM are "fixed event" rather than "fixed horizon" in the sense that the target is fixed while the horizons of the forecasts change from 7 days to 1 day (see, for example, Yetman (2018)). Fixed-event forecasts provide an effective way to evaluate how expectations of the target change in response to information arrival. The hourly data set used in the paper is for 12 weather stations in eastern Australia, three for each of Queensland (QLD), New South Wales (NSW), Victoria (VIC) and South Australia (SA). Each station has been selected because of its proximity either to an existing or proposed wind farm, or to the bushfires of the 2019-20 Australian bushfire season (September 2019 to March 2020), which has become known colloquially as the black summer.

Very briefly, the results reported here demonstrate that the BOM's wind-speed forecasts cannot be regarded as rational forecasts. The results also illustrate differences in the extent of irrationality both across stations and the hours of a day. In addition, limited evidence from simple quantile regressions suggests that the BOM official wind speed forecasts perform better when the wind speed is relatively high. Furthermore, examination of internal consistency of a sequence of forecasts and comparing forecasts with realised wind speed, reveals that these wind-speed forecasts mimic problems encountered in the macroeconomic literature on forecasting with real-time data in the sense that the forecasts may reflect considerable judgmental information (Croushore & Stark, 2001; Croushore, 2006; Corradi et al., 2009). The major issue which is raised in these situations is the question of what exactly the forecasters are trying to predict. It certainly seems that the BOM considers objectives beyond those of simply providing the best wind forecasts across all horizons. These objectives may include providing advance warning of extreme bush fire conditions or maintaining consistency between wind speed forecasts and forecasts of other weather conditions also generated by the numerical weather prediction model.

2 Wind speed observations

Wind speed, or wind flow speed, is a fundamental atmospheric quantity caused by air moving from high to low pressure, usually due to changes in temperature. Official BOM data on hourly wind speed and multi-horizon forecasts for up to 7 days ahead have been collected from 12 Australian weather stations. The sample period for the targeted hourly wind speed is from 15:00 on 1 June 2015 to 14:00 on 6 March 2020 comprising 41089 targets (note that the realisation of the targets is observed up to 22:00 on 5 March 2020). The locations of the 12 weather stations are indicated in Figure 1 which shows that they should provide a good indication of the efficacy of forecasts in different positions across eastern Australia.[4]

In QLD, Cairns and Kingaroy are the closest BOM stations to the Mt. Emerald and Coopers Cap Wind Farms, while Applethorpe is both near to the planned Macintyre wind farm but was also at the centre of the bushfires in south east QLD in the summer of 2019. In NSW, Merriwa is adjacent to the largest proposed wind farm in Australia, the Liverpool Range wind farm which has a planned 267 turbines with a capacity of 1000 MW. Glen Innes is also close to a wind farm while Moruya was at the epicentre of the bush fires along the south coast of NSW. The three VIC stations are all associated with actual or approved wind farms with Stawell being close to the third largest farm in Australia at the time of opening, Ararat wind farm. Hamilton is also close to the Grampian National Park bushfire zone. Finally, SA has the highest proportion of wind generated electricity to total energy demand of all the Australian states and all three weather stations were chosen because of their proximity to a significant wind farm, namely Snowtown (Clare), Lincoln Gap (Whyalla) and Wattle Point (Edinburgh) wind farms.

The International System of Units (SI) measurement of speed and velocity is metres per second (m/s) while the BOM forecasts are published in terms of Knots, that is nautical miles per hour. The conversion factor of 1 Knot is equal to 0.514444 metres per second has been used to express all the data in terms of the SI standard.

[4] The data and code used to generate the results reported in this paper are available upon request from the corresponding author.

Figure 1: Illustration of the location of the 12 BOM weather stations in QLD, NSW, VIC and SA that are used in the analysis.

There seems to be relatively good spread to the statistics of wind speed between the 12 stations.[5] There are the low-mean, low-variance stations such as Applethorpe (QLD) and Clare (SA) and the high-mean, high-variance stations such as Glen Innes (NSW) and Hamilton (VIC). There are also interesting anomalies like Kingaroy (QLD) having a relatively low mean and variance but also having the second highest maximum speed recorded at $19.75\,m/s$. The highest maximum wind speed of $21.61\,m/s$ is recorded at Hamilton (VIC).[6] From the perspective of generating

[5]The data shows many hours with zero wind speed. A cup or propeller anemometer does not turn at a low wind speed ($< 0.5\,m/s$) because of internal mechanical friction. There is a small range after the low threshold value where measurements from this range are prone to errors (Brock & Richardson, 2001). The BOM confirmed that these observations were in fact recorded wind speed and were not missing observations. The presence of a large number of zeros precluded the use of a logarithms in the analysis.

[6]The fastest wind speed ever recorded was measured on Barrow Island, Australia. It was associated with Tropical Cyclone Olivia on 10 April 1996 and was measured at $113.3\,m/s$.

Table 1

Descriptive statistics of actual hourly wind speed for 12 weather stations in Australia. The sample period for the observed wind speed runs from 1 June 2015 to 5 March 2020. Wind speeds were provided in knots, but were translated into metres per second.

Station	State	Lat. °S	Long. °E	Mean	Std. Dev.	Max
Cairns	QLD	16.92	145.78	4.29	2.19	14.56
Kingaroy	QLD	26.54	151.84	2.95	1.92	19.75
Applethorpe	QLD	28.62	151.94	2.00	1.43	9.52
Glen Innes	NSW	29.67	151.69	4.31	2.44	17.54
Merriwa	NSW	32.18	150.17	3.64	2.27	15.48
Moruya	NSW	35.90	150.14	3.12	2.15	14.40
Sheoaks	VIC	37.90	144.13	3.74	2.21	15.95
Stawell	VIC	37.07	142.74	3.68	2.10	14.10
Hamilton	VIC	37.64	142.06	5.11	2.39	21.61
Edinburgh	SA	34.71	138.62	4.49	2.65	18.57
Clare	SA	33.82	138.59	2.78	2.03	11.99
Whyalla	SA	33.05	137.52	4.95	2.54	16.98

electricity from wind turbines, these statistics should be appraised bearing in mind that $3.5\,m/s$ is the typical speed at which small wind turbines start generating power (*cut-in speed*). This means that Kingaroy (QLD), Applethorpe (QLD), Moruya (NSW) and Clare (SA) all have average wind speeds below the cut-in speed. On the other hand it has been estimated that average speeds of at least $6-8\,m/s$ are required for a small wind turbine to be economically viable. At the other end of the range, turbines reach maximum power generation in the range $10-15\,m/s$ and most turbines are braked or stopped at $25\,m/s$ (*cut-out speed*). Finally, it should also be remembered that the BOM wind speeds are typically measured at height of 10 meters above ground level. The hub height of wind turbines may reach 100 meters and wind speed increases with height.

3 Econometric rationality and information flow

The BOM provides multi-horizon forecasts for hourly wind speed y_t with $t = 1, 2 \cdots$ at horizons $h_1 < h_2 < \cdots < h_H$. The forecasts for y_t made h periods earlier at time $t-h$ are denoted by $\widehat{y}_{t|t-h}$. Forecast errors are denoted $e_{t|t-h} = y_t - \widehat{y}_{t|t-h}$. The frequency of forecast revisions is daily which is lower than the hourly frequency for wind speed observations.[7] As the target date approaches, the wind forecast users observe a number of forecast revisions for the same target defined as $\widehat{y}_{t|t-h_s} - \widehat{y}_{t|t-h_l}$ with $h_s < h_l$. For forecasts to be rational, any revisions to the forecasts must incorporate newly arrived information efficiently.[8] Isiklar & Lahiri (2007) develop two measures of information flow in multi-horizon forecasts, the mean square forecast revision (MSFR_h) and the improvement in the forecast accuracy with the new information computed as the change in the mean squared forecast error (ΔMSFE_h).

The value of MSFR for a horizon h is computed as the difference in the forecast between horizon h and $h+1$

$$\text{MSFR}_h = \text{E}\left[\widehat{y}_{t|t-h} - \widehat{y}_{t|t-(h+1)}\right]^2, \tag{1}$$

and may be taken as a measure of the informational content of the revision made at horizon h. As the BOM's official wind speed forecasts reflect both the output of the ACCESS model and post-processing by BOM meteorologists, the value of MSFR_h incorporates not only information arrival between two adjacent points in time, but potentially also meteorologists' behavioural characteristics. In this sense the MSFR_h may be interpreted as the revision effort.

The ΔMSFE_h measures the difference in forecast accuracy between horizon h and $h+1$, that is

$$\Delta\text{MSFE}_h = E\left[y_t - \widehat{y}_{t|t-(h+1)}\right]^2 - E\left[y_t - \widehat{y}_{t|t-h}\right]^2. \tag{2}$$

Positive values indicate that forecast revisions result in improvement in forecast accuracy, whereas negative values suggest a worsened update compared with the forecast made previously.

The fundamental point made by Isiklar & Lahiri (2007) is that for forecasts to be rational, in the sense that each update incorporates new information efficiently, the revision effort must

[7] The difference of frequencies in wind speed observation t and forecast horizon h means an increment in h is equivalent to 24 increments in t. However, instead of using $t - 24h$ to denote the time of forecasting, the paper follows the general notation $t - h$, keeping in mind that t and h share different frequencies.

[8] Based on the assumption of a quadratic loss function there are a number of monotonicities in variance relating to the rationality of multi-horizon forecasts (Patton & Timmermann, 2012). These derived bounds for the second moments are implied by rational forecast and hence are only necessary conditions (Lahiri, 2012). Although not reported here the variance of the BOM multi-horizon wind forecasts largely satisfy these minimum (necessary) requirements. These results are available from the corresponding author on request.

be rewarded by an equivalent improvement in forecast accuracy, or in other words $\text{MSFR}_h = \Delta\text{MSFE}_h$. This equality between MSFR_h and ΔMSFE_h may be used to assess forecast rationality with any non-negative differential between MSFR_h and ΔMSFE_h providing evidence of forecast irrationality.

Figure 2 presents the values of MSFR_h and ΔMSFE_h across 6 horizons for all weather stations in the sample. The most striking result is that for all the weather stations and across all horizons without exception, the values of MSFR_h are much higher than the values of ΔMSFE_h. This is a quite remarkable result and is a clear indication that the BOM's forecast revisions do not only incorporate new information with a view to improving forecast accuracy, but perhaps also reflect behavioural characteristics that represent either overreaction to new information or are irrelevant to the realisation of wind speed but are judged to be important for other reasons.[9] The robust conclusion that emerges from Figure 2, therefore, is that forecasts of wind speed at these stations are irrational, at least when judged on this metric.

There are a number of other interesting observations to be made. First, in some cases (Applethorpe, Clare, Kingaroy, Moruya and Sheoaks), ΔMSFE_h for one day and/or two days ahead is negative. This result suggests short-horizon forecast revisions have actually led to more inaccurate forecasts. This result may be due to an increase in the bias of the forecasts at short horizons, a point that is taken up again later. Second, analysing the patterns in MSFR_h and ΔMSFE_h across horizons helps to identify the timing of information arrival. For some stations (Edinburgh, Hamilton, Merriwa and Whyalla) there is a non-increasing pattern in both revision effort (MSFR_h) and accuracy improvement (ΔMSFE_h) as the horizon shrinks, suggesting that forecast revisions are consistent with the purpose of improving forecast accuracy. Moreover, this observation suggests that the most valuable information to be incorporated in the BOM's revised forecast arrives 6 days before realisation. Cairns is the station where both MSFR_h and ΔMSFE_h exhibit a U shape, implying information adopted in one day and two days ahead forecasts is

[9] Isiklar & Lahiri (2007) and Lahiri (2012) model suboptimal forecast revisions using a parameter to control how forecasters perceive the impact of news. They show that $\Delta\text{MSFE}_h < \text{MSFR}_h$ can result from the overreaction of forecasters to news impact.

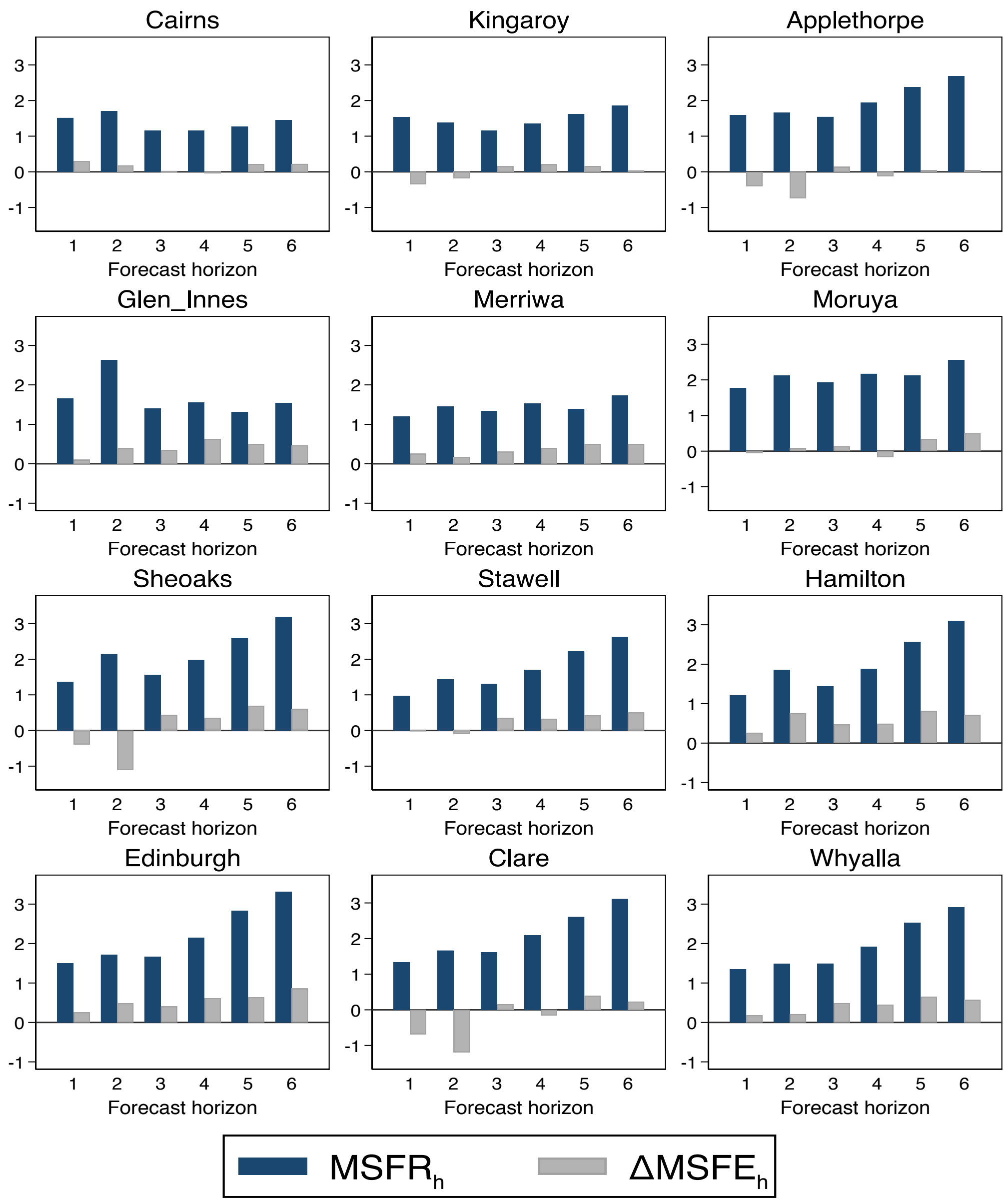

Figure 2: Illustrations of revision effort, MSFR_h and change in accuracy, $\Delta\mathrm{MSFE}_h$ for the 12 BOM stations. The measurement for h=7 is lost in computing the difference in MSFE.

important. Glen Innes is a perverse case where revision effort appears to be completely inconsistent with improvement in forecasting performance. The values of MSFR_h at short horizons

($h = 1$ and 2) are larger than the longer horizons but ΔMSFE_h in these short horizons gets smaller. Large revision effort accompanied with small accuracy improvement implies that BOM neither pursues forecast stability nor optimises skills when forecasting wind speed at some stations. This feature differs from BOM's official maximum temperature forecasts, which Griffiths et al. (2019) find more stable than automated guidance based on a numerical weather prediction model.

4 Regression-based tests of rationality by horizon

Simple regression based tests of forecast rationality are based on the so-called Minzer-Zarnovitz (MZ) regression[10]

$$y_t = \alpha_h + \beta_h \widehat{y}_{t|t-h} + u_t \tag{3}$$

with the test of rationality being given by the joint test

$$\alpha_h = 0 \quad \text{and} \quad \beta_h = 1 \quad \forall h. \tag{4}$$

For $h > 1$ day the disturbance term u_t will exhibit serial correlation of order $24(h-1)$ because of the overlapping forecast problem. Consequently a Newey-West correction is required when estimating equation (3). Although the order of the serial correlation is known in theory, in practice there is likely to be remaining serial correlation in addition to that induced by the overlapping forecasts. As a result the Newey-West estimator is specified with a maximum lag of 168 (one week).

Consider imposing the restriction that $\beta_h = 1$ in equation (3). This restriction would leave the interpretation of α_h as the bias for each forecast horizon. A negative estimate of bias would mean that the forecast over-predicts the actual, while a positive value would imply that the forecast under-predicts the actual. A property of a rational forecast with a symmetric loss function is that it is unbiased. It is not unreasonable to expect therefore that in these multi-horizon forecasts,

[10]The MZ regression test and variance and information flow tests do not subsume one another. Rejection of forecast optimality by one test does not necessarily imply rejection by the other. Each approach may add value in terms of the interpretation of deviations from rationality.

the bias will approach zero as the forecast horizon shrinks. In other words, given rational forecast revisions, one would expect the bias of the forecast to approach zero, from above or from below, as the forecast horizon shrinks.

Figure 3 plots the estimated bias of the forecasts for all 12 of the BOM stations. Although the extent of the bias varies from station to station, many stations exhibit a startling pattern of increasing (over-) prediction with decreasing horizon. In 8 of the 12 stations the bias actually diverges from zero as the horizon shrinks.

Edinburgh (SA) experiences almost linear growth in bias which passes through zero at a forecast horizon of 2-days. Cairns (QLD) and Glen Innes (NSW) demonstrate an odd pattern with an almost constant positive bias (under-prediction) for up to 3 days out and then a sudden change to negative (over-prediction) bias at 1 day out. Only at Merriwa (NSW) is the pattern more or less as expected with the bias approaching zero as the horizon shrinks.

Note that the highest over-predictions across all seven horizons are observed in Applethorpe and Clare, and Table 1 suggests that these two stations have the lowest average wind speed (2.00 m/s at Applethorpe and 2.78 m/s at Clare). In fact the four stations, Kingaroy, Applethorpe, Moruya and Clare, that have the average wind speed under the cut-in speed, are all subject to over-prediction. This may imply that BOM tends to over-predict wind speed when wind speed is low. The high-wind-speed stations, such as Cairns, Glen Innes, Hamilton and Edinburgh (with an exception of Whyalla), are observed with under-predictions.

The results obtained by estimating the full MZ regression in equation (3) for the weather stations are given in Table 2. The scale of the rejection of rationality in this approach is unprecedented. None of the stations comes with estimated α_h and β_h being close to zero and one respectively at any horizon. The expected pattern is for the estimates to converge to the expected value of 1, either from below or from above, but in an orderly monotonic way as information arrival is processed rationally. This pattern is observed at 3 of the 12 stations, namely, Cairns (QLD), Glen Innes (NSW) and Hamilton (NSW). The other stations exhibit a wide variety of behaviour including, rather perversely, that of Applethorpe (QLD), Sheoaks (VIC) and Clare (SA) where

$\widehat{\beta}_h$ appears to converge to 0.

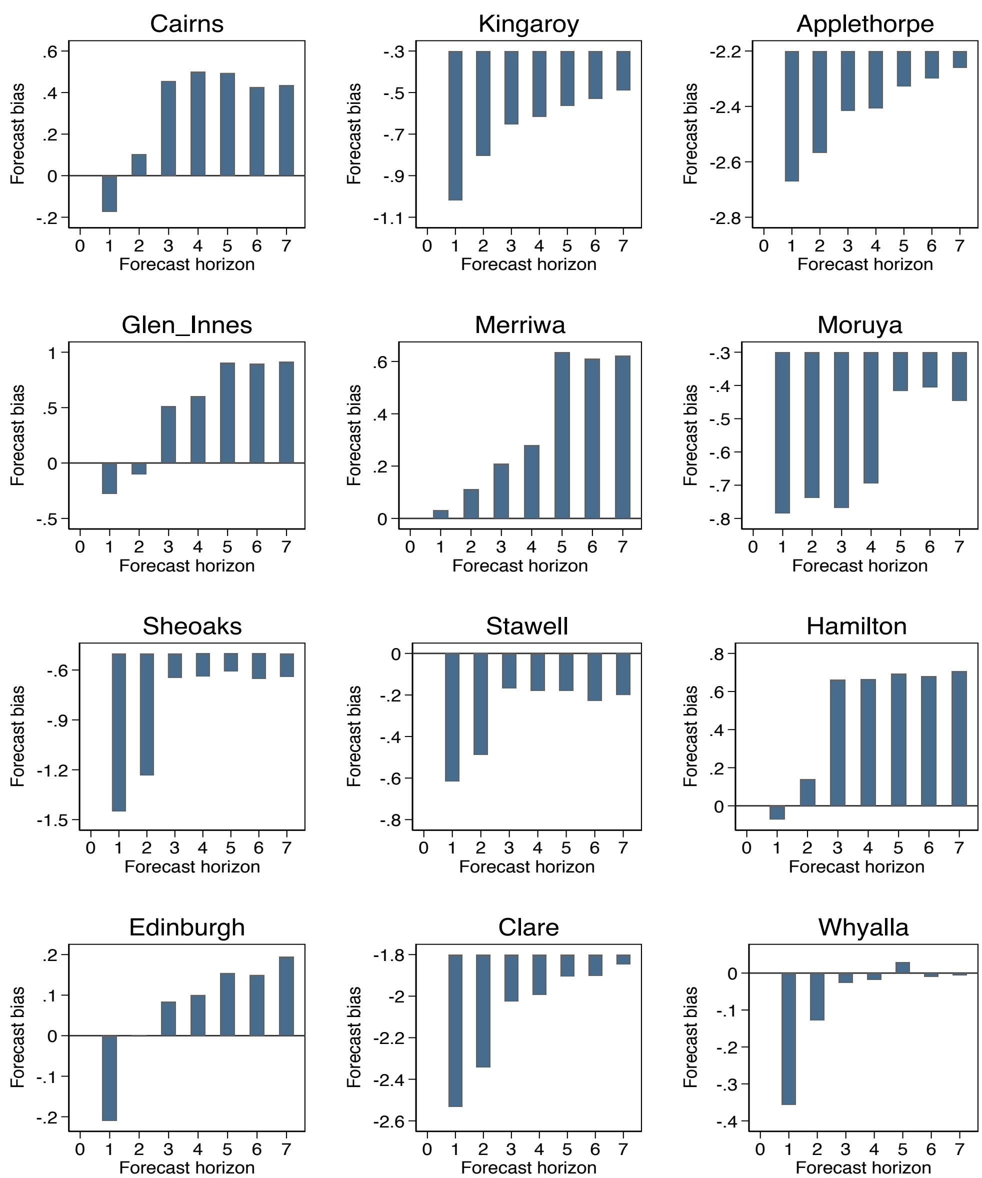

Figure 3: Plots of the bias of the forecasts for each of the 12 BOM stations for horizons of 1 to 7 days.

Table 2

The coefficients of MZ regressions in equation (3) for 12 BOM stations over horizons of 1 to 7 days. The sample period runs from 1 June 2015 to 6 March 2020. The standard errors shown in parentheses are computed using a Newey-West correction with a maximum lag of 168.

Station	Coef.	$\widehat{y}_{t\|t-1}$	$\widehat{y}_{t\|t-2}$	$\widehat{y}_{t\|t-3}$	$\widehat{y}_{t\|t-4}$	$\widehat{y}_{t\|t-5}$	$\widehat{y}_{t\|t-6}$	$\widehat{y}_{t\|t-7}$
Cairns	α_h	1.0701	1.3371	1.3586	1.3188	1.2165	1.3459	1.5128
		(0.0841)	(0.0943)	(0.1709)	(0.1507)	(0.0936)	(0.0904)	(0.1003)
	β_h	0.7218	0.7058	0.7649	0.7843	0.8101	0.7624	0.7213
		(0.0182)	(0.0235)	(0.0474)	(0.0417)	(0.0244)	(0.0232)	(0.0252)
Kingaroy	α_h	0.5202	0.6201	0.6444	0.8294	1.0660	1.2433	1.3438
		(0.0613)	(0.0597)	(0.0610)	(0.0721)	(0.0801)	(0.0762)	(0.0746)
	β_h	0.6120	0.6203	0.6392	0.5943	0.5359	0.4903	0.4679
		(0.0162)	(0.0159)	(0.0170)	(0.0210)	(0.0240)	(0.0230)	(0.0213)
Applethorpe	α_h	-0.0922	0.0623	0.2089	0.3575	0.5940	0.7594	0.9133
		(0.0476)	(0.0519)	(0.0629)	(0.0686)	(0.0672)	(0.0660)	(0.0626)
	β_h	0.4474	0.4233	0.4047	0.3724	0.3242	0.2880	0.2547
		(0.0114)	(0.0126)	(0.0165)	(0.0181)	(0.0179)	(0.0176)	(0.0167)
Glen Innes	α_h	1.1196	1.2015	1.2387	1.3027	1.2800	1.6052	1.9809
		(0.0556)	(0.0591)	(0.0832)	(0.0847)	(0.0967)	(0.1022)	(0.0982)
	β_h	0.6950	0.7047	0.8084	0.8116	0.8892	0.7919	0.6851
		(0.0121)	(0.0128)	(0.0242)	(0.0242)	(0.0298)	(0.0319)	(0.0316)
Merriwa	α_h	0.7747	0.8469	0.7714	0.8160	0.7856	1.0787	1.4398
		(0.0367)	(0.0402)	(0.0480)	(0.0537)	(0.0609)	(0.0622)	(0.0687)
	β_h	0.7941	0.7916	0.8362	0.8405	0.9500	0.8451	0.7287
		(0.0104)	(0.0110)	(0.0156)	(0.0166)	(0.0190)	(0.0210)	(0.0233)
Moruya	α_h	0.7621	0.7430	0.6594	0.8044	0.9643	1.2201	1.5321
		(0.0436)	(0.0447)	(0.0544)	(0.0724)	(0.0736)	(0.0725)	(0.0760)
	β_h	0.6046	0.6168	0.6333	0.6074	0.6098	0.5388	0.4463
		(0.0132)	(0.0134)	(0.0167)	(0.0219)	(0.0237)	(0.0230)	(0.0238)
Sheoaks	α_h	0.2133	0.3458	0.5049	0.6702	0.8469	1.1380	1.4851
		(0.0382)	(0.0390)	(0.0496)	(0.0501)	(0.0592)	(0.0649)	(0.0778)
	β_h	0.6792	0.6822	0.7372	0.7005	0.6657	0.5928	0.5146
		(0.0082)	(0.0087)	(0.0130)	(0.0136)	(0.0158)	(0.0171)	(0.0200)
Stawell	α_h	0.6093	0.7123	0.8389	0.9834	1.1511	1.3797	1.7268
		(0.0396)	(0.0391)	(0.0467)	(0.0503)	(0.0592)	(0.0659)	(0.0785)
	β_h	0.7149	0.7123	0.7385	0.6987	0.6554	0.5890	0.5041
		(0.0101)	(0.0100)	(0.0133)	(0.0137)	(0.0162)	(0.0169)	(0.0200)
Hamilton	α_h	1.0740	1.2395	1.4989	1.6695	1.8837	2.3191	2.7050
		(0.0497)	(0.0502)	(0.0653)	(0.0683)	(0.0751)	(0.0857)	(0.1022)
	β_h	0.7783	0.7787	0.8119	0.7739	0.7308	0.6306	0.5460
		(0.0095)	(0.0093)	(0.0148)	(0.0153)	(0.0171)	(0.0191)	(0.0236)
Edinburgh	α_h	0.5165	0.7248	0.8113	0.9277	1.2132	1.4690	1.9104
		(0.0438)	(0.0448)	(0.0573)	(0.0615)	(0.0726)	(0.0861)	(0.1041)
	β_h	0.8453	0.8382	0.8348	0.8115	0.7558	0.6960	0.6001
		(0.0106)	(0.0108)	(0.0140)	(0.0153)	(0.0179)	(0.0207)	(0.0257)
Clare	α_h	-0.2906	-0.1090	0.1151	0.2753	0.4320	0.6799	1.0208
		(0.0546)	(0.0603)	(0.0668)	(0.0689)	(0.0776)	(0.0790)	(0.0933)
	β_h	0.5781	0.5642	0.5547	0.5246	0.5012	0.4486	0.3801
		(0.0095)	(0.0109)	(0.0137)	(0.0145)	(0.0171)	(0.0176)	(0.0211)
Whyalla	α_h	0.4854	0.6528	0.7078	0.9382	1.1539	1.4058	1.7058
		(0.0359)	(0.0411)	(0.0471)	(0.0558)	(0.0604)	(0.0706)	(0.0807)
	β_h	0.8413	0.8462	0.8524	0.8074	0.7713	0.7144	0.6546
		(0.0074)	(0.0091)	(0.0106)	(0.0128)	(0.0131)	(0.0153)	(0.0172)

5 Regression-based tests of rationality by hour

Considering that the ability of meteorological models to forecast local wind speed may vary depending on time of a day (Monk et al., 2019), this section focuses on a task of breaking down the forecasts hour-by-hour to see if any additional performance information can be gleaned. Consequently, the MZ regression is now applied to each hour

$$y_{it} = \alpha_{ih} + \beta_{ih}\widehat{y}_{it|t-h} + u_{it}. \tag{5}$$

The null hypothesis of rationality requires that $\alpha_{ih} = 0$ and $\beta_{ih} = 1$ for all $i = 1, \cdots, 24$ and $h = 1, \cdots, 7$ days. The following analysis focuses on the forecasts made at the longest horizon (7-day ahead) and two short horizons (1-day and 3-day ahead).

Figures 4 and 5, respectively plot the intercept and slope coefficients from the hourly MZ regressions for each of the three horizons considered. The estimates of the intercepts portrayed in Figure 4 are all far too large at midnight and in the early hours of the morning, decline over the course of the morning and remain there until 8pm before starting to increase again. This midnight effect coincides with the finding in Monk et al. (2019) who find systematic overnight biases in the wind speed forecasts produced by meteorological models. There is less of a structured pattern in the slope coefficients in Figure 5 but the one promising sign is that the solid curve for the 1-day forecast horizon is usually closest to the optimal value of 1.

Based on these hourly MZ regressions, it should come as no surprise that the null hypothesis of the econometric rationality of the multi-horizon forecasts is rejected for every station, at every hour and for every horizon.[11] The best performing stations over the period 8am to 8pm are Cairns, Glen Innes, Merriwa and Hamilton, although all the formal tests still indicate rejection of the rationality of the forecasts. Interestingly, these are the stations with relatively high wind speeds. Together with the observation made in Section 4 that the bias is most marked at low wind speed stations, the general conclusion appears to be that the higher the average wind speed,

[11] These results are not presented for reasons of space. They are available from the corresponding author on request.

the better the BOM forecasting performance.

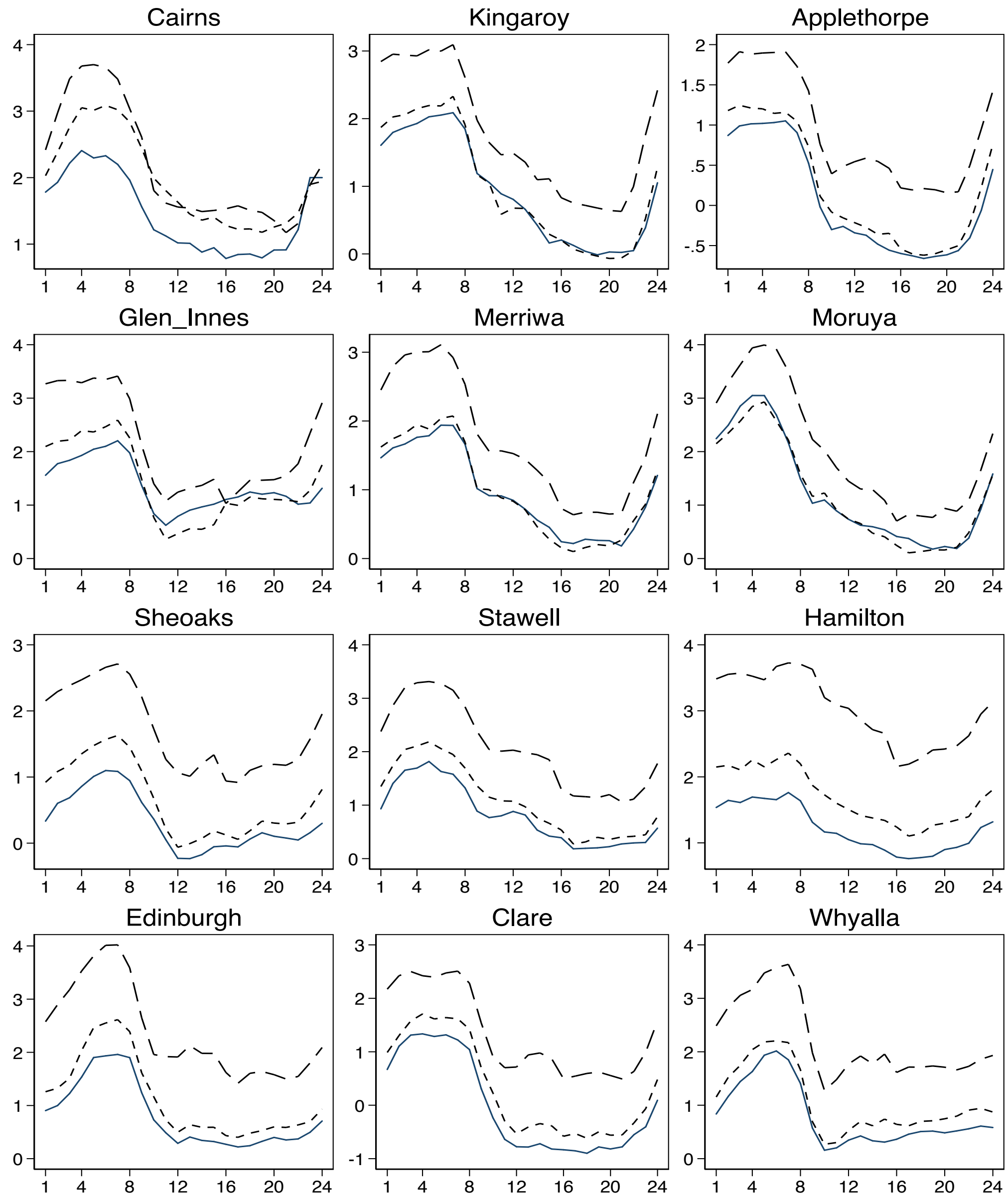

Figure 4: The intercepts, α_{ih} of hourly MZ regressions for horizons of 1 (solid line), 3 (short dashed line) and 7 (long dashed line) days. Forecast rationality requires that $\alpha_{ih} = 0$ for all i and h.

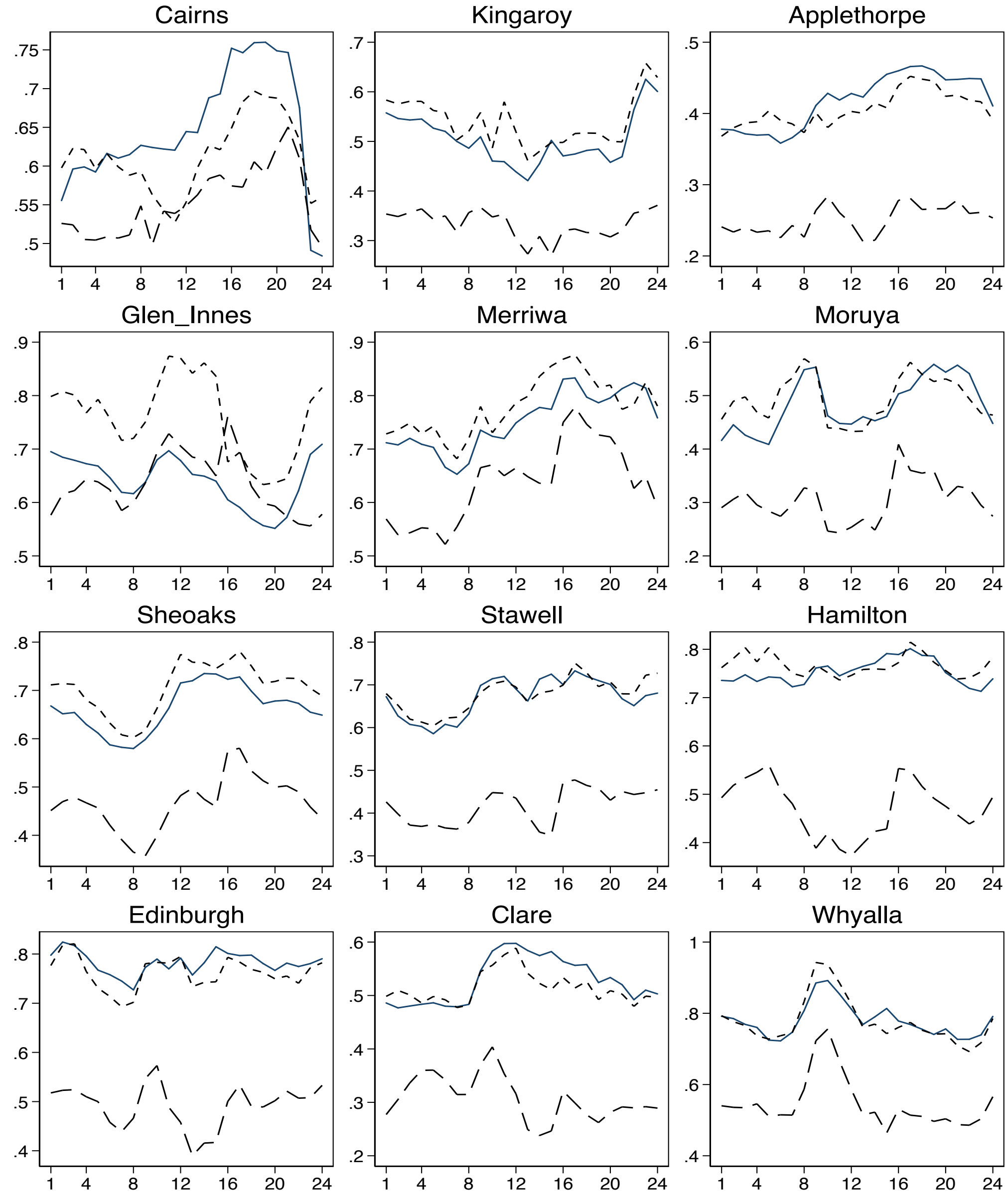

Figure 5: The slope coefficients, β_{ih} of hourly MZ regressions for horizons of 1 (solid line), 3 (short dashed line) and 7 (long dashed line) days. Forecast rationality requires that $\beta_{ih} = 1$ for all i and h.

6 Discussion

There are several fundamental questions that arise out of the rationality test results for the official BOM forecasts of wind speed. The first of these is simply this: why are the characteristics of the forecasts so different across stations? The tentative reason that emerges from these results is that the BOM is better at forecasting in places where average wind speeds are higher. This question is worthy of research in its own right, but as a first attempt to explore the topic it is useful to use quantile regression techniques aimed at assessing forecasting performance at low and high wind speeds respectively.[12] Consequently, MZ quantile regressions for 7 forecasting horizons were implemented for the 25th and 75th quantiles to see if there was any significant difference in performance. For reasons of space only the results for the slope coefficient are plotted in Figure 6, in which the long dashed line represents the $\hat{\beta}$ coefficient for the 75th percentile of wind speed and the short dashed line is the same quantity for the 25th percentile. The solid line represents the coefficient from the ordinary least squares regression. In many instances the coefficient estimated from the 75th quantile regression is significantly closer to 1, providing some support to the conjecture.

The puzzling aspect of these results is why the characteristics of the forecasts differ so greatly across the hours of the day. Generally, the Australian Digital Forecast Database (ADFD) files with the official forecasts are updated twice a day (around 6am and 6pm each day). While these discrete revision times may hint at granularity in the resolution of forecast performance, they do not really explain the consistently poor performance around midnight and the early hours of the day. In general, forecast performance is determined by two factors, namely, the data generating process of the realisation and forecaster's effort. There are, therefore, two possible explanations for differences across hours. The first is that forecasting effort is not expended equally across the day by the BOM. On this argument, the BOM would seem to be expending minimal effort to forecasting wind speed in the hours around midnight and just after. The second, and possibly

[12] For reasons of space it is not possible to include a description of quantile regression but interested readers unfamiliar with the method will find that Koenker & Hallock (2001) is a useful survey while Koenker (2005) provides a comprehensive treatment.

more persuasive, is that wind speed around midnight is very complicated to model and hence difficult to forecast.

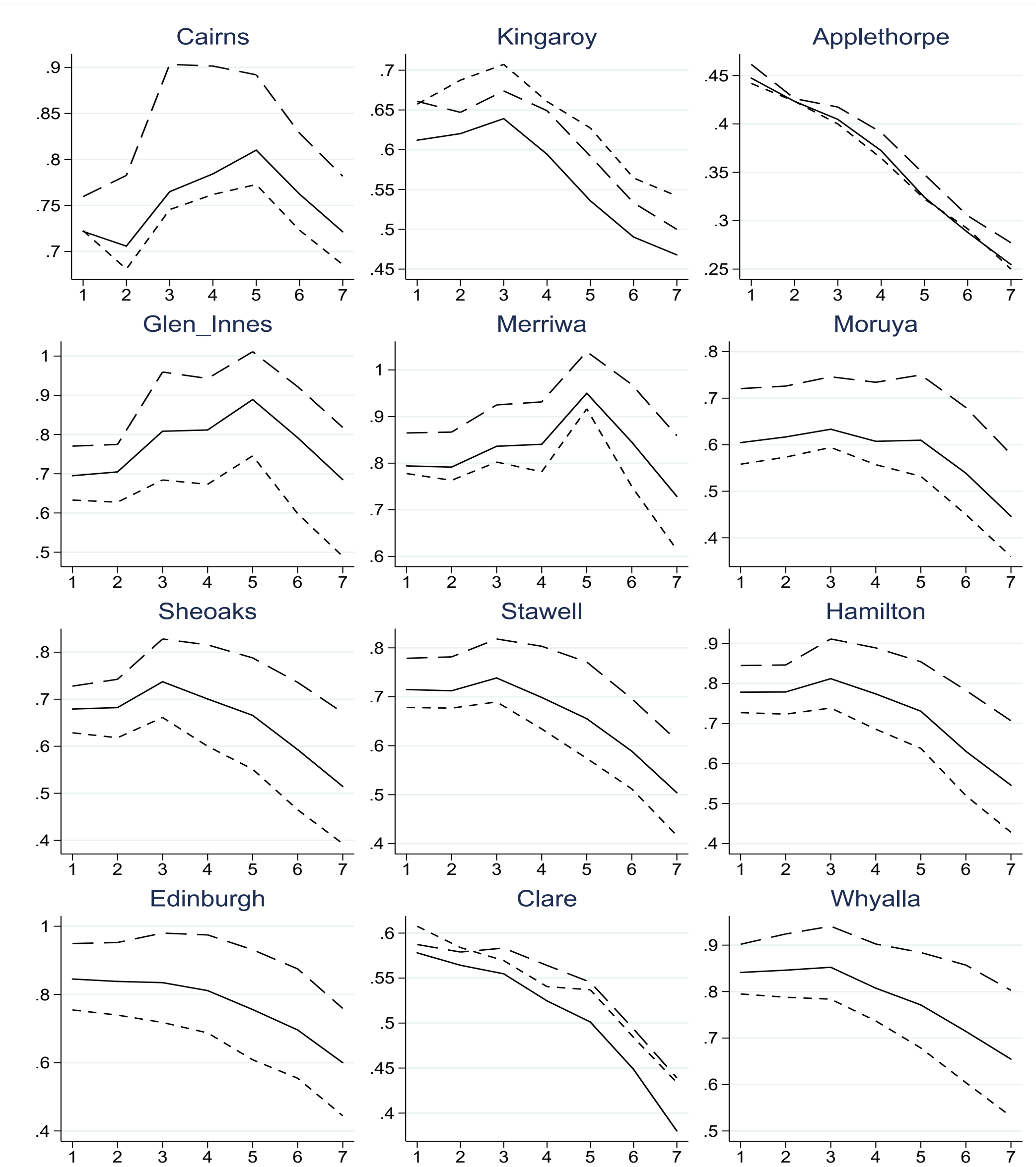

Figure 6: The slope coefficients of quantile MZ regressions for different horizons. The solid line is the coefficient from the ordinary least squares MZ regression. The long dashed line is the coefficient for the 75th quantile regression and the short dashed line is the coefficient for the 25th percentile.

The next fundamental question raised by these results is why the performance of the forecasts at $h = 1$ is not significantly better than those at longer horizons across all stations? A possible explanation is that achieving accurate weather forecasts is not the only purpose of BOM. In the case of wind speed forecasts for a particular location at a given time, the forecast values may also be used, for example, as an input into a model to predict maximum fire danger with a view to providing a warning to the public. The two-stage procedure for producing the official wind speed forecasts enables the meteorologists to adjust the forecast values made by the ACCESS model in order to achieve this objective. The results show that 1- and 2-day ahead forecasts tend to over-predict compared with longer horizon forecasts.

On the one hand, it has been argued in the literature that large and rapid revisions of wind speed forecasts (so-called run-to-run volatility) at short horizon may impose substantial costs on other wind speed forecast users (McLay, 2011). In this scenario, poor performance at shorter horizons stems from a desire not to flip-flop (Griffiths et al., 2019). On the other hand, the public may react more effectively to late warnings than to early warnings. Therefore, it may be that BOM forecasters intentionally and artificially raise wind speed forecasts (produced by the ACCESS model) a couple of days before the realisation in order to alert the public effectively.

From this perspective, the shorter the horizon, the more likely BOM forecasters will purposefully add a bias that results in a forecast much higher than the future realised wind speed. While both reasons may be an admirable traits from the point of view of public policy, it is clear that the official forecasts of wind speed contain a substantial amount of judgemental adjustment. The major issue which is raised in these situations is the question of what exactly the forecasters are trying to predict and whether or not the forecast figure should in fact be treated as an optimal point forecast at all.

Finally, during the almost 5 years of data, the BOM has experienced changes in forecasting practice which have been introduced at different times in different states. For instance, a procedure of manual adjustment to forecasts for fire warning purpose could be introduced by individual states at various points of time. Such a procedure often targets forecasts at specific horizons and for specific hours which may vary from state to state. Figure 7 plots the forecast biases for each of the stations by year, but only for the 1-day horizon because it should be the most accurate. Given the results reported so far, it should certainly not be expected that the value of the bias exhibits minimal random fluctuations on either side of zero, but it is interesting to ascertain if there is any evidence of improved performance or changes in forecast practice over the period.

Figure 7 does not present a consistent pattern in how bias changes over time across stations. This result implies that changes in forecast practice indeed has been different across states and even across stations within the same state. A limited number of stations, including Kingaroy, Applethorpe, Glen Innes, Moruya and Clare, have experienced reduction in biases in recent years.

Large over-predictions are observed in many of the stations in 2020, which perhaps can be attributed to the extraordinary circumstances of the bushfires and to the small sample size, given that the data end in March. Since it is highly unlikely that the ACCESS model generates systematic over-predictions over a specific period, it is possible to conjecture that the BOM may have purposefully manipulated wind forecasts in the first three months of 2020 during the intensive bushfire period. As a broad generalisation, up to this point, the Cairns' forecasts have performed relatively well in comparison to most other stations. Yet in Figure 7, a very worrying pattern emerges, namely, an almost linear progression over the period from significant under-prediction to significant over-prediction.

On a more positive note, there is weak evidence to suggest that the 2019 forecasting outcome, at least in terms of the forecast biases reported here, is marginally better than the preceding years. This improvement is observed particularly when benchmarked against the immediately preceding year 2018. Interpreted in a manner as favourable as possible, despite that BOM's official wind-speed forecasts serve multiple purposes, pursuing accurate forecasts is one of BOM's recent targets and better forecasting performance can be expected in the next few years.

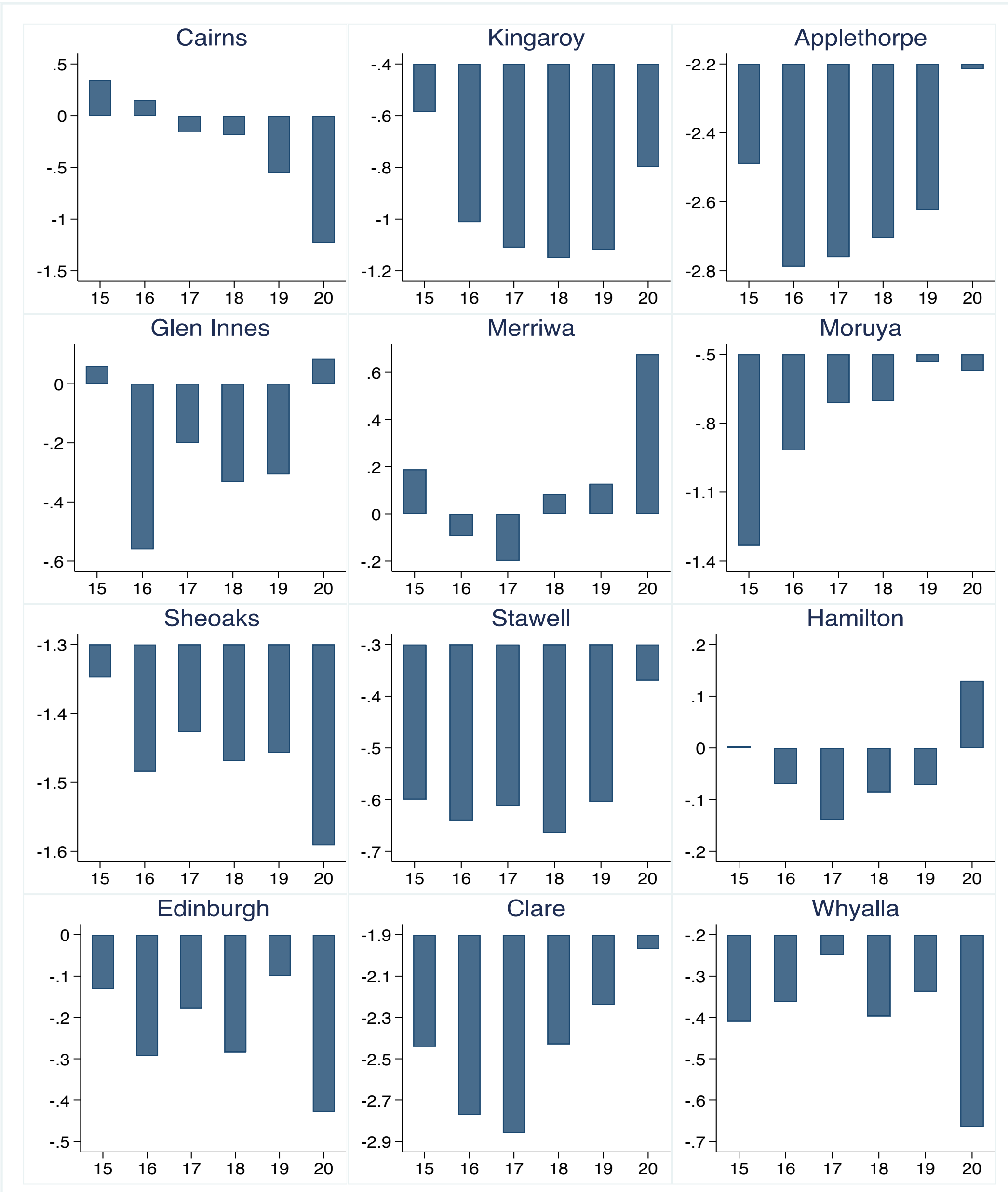

Figure 7: Estimated biases of the 1-day ahead forecasts for each of the 12 weather stations broken down by year.

Finally, all the properties of rational forecasts used in this paper are derived from a mean squared error loss function. This assumption is widely used but is certainly not universally accepted.

The significance of the over-prediction bias in the wind speed forecasts reported in Section 4 and particularly in Figure 3, raises the question of whether or not the BOM has an asymmetric loss function. From the perspective of fighting bushfires it can easily be argued that under-prediction of wind speed poses more of a problem than over-prediction. However, it is difficult to argue in favour of a lower cost to over-prediction in terms of the dispersion of pollutants or from the perspective of the wind energy sector. The properties of rational forecasts in presence of asymmetric loss and the estimation of the parameters of asymmetric loss functions are, however, well documented (Elliott et al., 2005; Patton & Timmermann, 2007; Elliott et al., 2008). This avenue for research looks promising in the light of the results reported here but is left for future work.

7 Economic value

The rationality assessments in the previous sections rely purely on statistical metrics. In this section an attempt is made to assess the economic value of the BOM's wind speed forecasts and their subsequent revisions based on a cost-loss ratio decision model (Murphy, 1977; Richardson, 2000; Foley & Loveday, 2020). Assume that wind speed forecast users face economic loss L when a wind anomaly event occurs. If they decide to take an action in advance to protect against the future wind anomaly event, they need to pay a cost of action C but are rewarded with a reduced loss L_1, with $C + L_1 < L$, when the event occurs. Of course the quantities C, L and L_1 are very problem specific and may in fact be difficult to quantify exactly. However, defining the cost-loss ratio $\alpha = C/(L - L_1)$, where α is now the cost of action relative to the loss prevented, allows this model to be made broadly applicable in the range $\alpha \in (0, 1)$.

The economic value of wind speed forecasts at a given horizon, V, is constructed using the expected expense of taking action based on three types of forecast, namely, the expected expenses due to taking action to mitigate the anomaly based on the unconditional probability of the event, E_U, hypothetical perfect forecasts, E_P, and the official BOM forecasts, E_F. The quantity V is now defined as

$$V = \frac{E_U - E_F}{E_U - E_P}\,. \tag{6}$$

In order to construct an estimate of economic value due to relying on official BOM forecasts, the three different types of expected expense must be computed.

Case 1: E_U

In this case, the decision of whether or not to take action is simply based on the unconditional probability of occurrence of the wind anomaly event, p. The expected expense, E_U, is given by

$$E_U = \min\{pL,\, C + pL_1\}. \tag{7}$$

Case 2: E_P

In a hypothetical scenario where users are provided with perfect wind speed forecasts and only take action when the wind anomaly event will occur, the user's expected expense, E_P, is

$$E_P = p(C + L_1). \tag{8}$$

Case 3: E_F

Now consider using the official BOM wind speed forecasts to inform whether to take action. The expected expense, E_F, depends on three of the four probabilities of the joint events listed in the following contingency table:

		Observed	
		No	Yes
Forecast	No	Free Hit	Miss
	Yes	False Alarm	Hit

The Free Hit is so named because the correct forecast is for no anomaly and thus no action is required and no expense is incurred. The expected expense E_F of using the BOM forecasts is then the weighted average of the expenses of the remaining three events,

$$\begin{aligned} E_F &= \Pr(\text{False Alarm})C + \Pr(\text{Miss})L + \Pr(\text{Hit})(C + L_1) \\ &= \Pr(\text{False Alarm})C + (p - \Pr(\text{Hit}))L + \Pr(\text{Hit})(C + L_1). \end{aligned} \tag{9}$$

Define the relative false alarm rate $F = \Pr(\text{False Alarm})/(1-p)$ and the relative hit rate $H = \Pr(\text{Hit})/p$. The expected expense of taking action based on the forecast is therefore rewritten as

$$E_F = F(1-p)C + p(1-H)L + H\,p(C + L_1). \tag{10}$$

Combing equations (7), (8) and (10), and using the definition of the cost-loss ratio, $\alpha = C/(L - L_1)$, equation (6) for V can be rewritten as

$$V = \frac{\min\{\alpha, p\} - F\alpha(1-p) + H\,p(1-\alpha) - p}{\min\{\alpha, p\} - p\alpha}. \tag{11}$$

This equation demonstrates that the relative economic value of wind speed forecasts is determined by both the values of F and H that represent the forecasting performance of wind anomaly events and the values of p and α that are independent to the wind speed forecasts. Given that $C + L_1 < L$ and hence α ranges between 0 and 1, it follows that V can be computed for all values of α in this range. Note, however, that irrespective of their accuracy, wind forecasts, can have negative relative economic values particularly when α gets close to 0 or 1. A very low or very high cost-loss ratio suggests that simply using climatological information to decide whether to take action or not for all targeted days is likely a better strategy than relying on the forecasts, that is $E_U < E_F$. Note that if $V > 0$, and the official forecasts do indeed provide positive value, the upper bound of V is 1. This limit provides a useful metric for comparison across different forecasts.

The economic value of the BOM forecasts will now be assessed by considering two different types of wind anomaly.

7.1 Wind drought

The first anomaly which is considered is wind drought. If wind speed drops below 3.5 m/s then most wind turbines stop generating power and if this occurs during the peak hours of electricity usage, namely 14:00 to 20:00, it is likely that gas operated peaking plants will have to be fired up to meet demand. Specifically, daily wind drought is observed (or forecast at a given horizon) if during a peak target period from 14:00 to 20:00, any five or more hourly wind speed observations (or forecasts at the given horizon) are lower than 3.5 m/s. Accurate wind drought forecasts are likely to be of critical importance to operators of peaking plants, for example, in helping to optimise gas storage decisions.

Table 3 presents the frequencies of observed wind drought, p, the relative false alarm rate, F, and the relative hit rate, H, for forecasts at the horizons of 1-day, 3-day and 7-day, respectively. Wind drought is observed remarkably frequently. Especially at four stations, Kingaroy, Applethorpe, Moruya and Clare, where average wind speed is low, wind drought occurs in more than 80% of the sampled days. Interestingly, the relative hit rate H of wind drought forecasts does not appear to increase significantly as the forecast horizon shrinks, and H_3 is often higher than H_1. As expected, however, the relative false alarm rate, F, does decrease as the horizon shortens.

Table 3

Sample unconditional probability of wind drought, p, together with the sample relative false alarm rate F and the sample relative hit rate H for forecasts at the horizons of 1-day, 3-day and 7-day, respectively.

	p	F_1	F_3	F_7	H_1	H_3	H_7
Cairns	0.45	0.44	0.52	0.50	0.89	0.88	0.84
Kingaroy	0.81	0.35	0.32	0.55	0.65	0.69	0.73
Applethorpe	0.94	0.00	0.01	0.08	0.42	0.42	0.42
Glen Innes	0.59	0.20	0.28	0.47	0.78	0.79	0.80
Merriwa	0.73	0.27	0.31	0.58	0.90	0.89	0.90
Moruya	0.89	0.12	0.10	0.33	0.78	0.69	0.72
Sheoaks	0.66	0.05	0.08	0.21	0.61	0.69	0.61
Stawell	0.71	0.12	0.21	0.34	0.73	0.78	0.68
Hamilton	0.37	0.14	0.28	0.35	0.80	0.82	0.66
Edinburgh	0.62	0.10	0.16	0.31	0.73	0.73	0.64
Clare	0.83	0.00	0.02	0.18	0.35	0.39	0.38
Whyalla	0.52	0.06	0.12	0.22	0.66	0.72	0.61

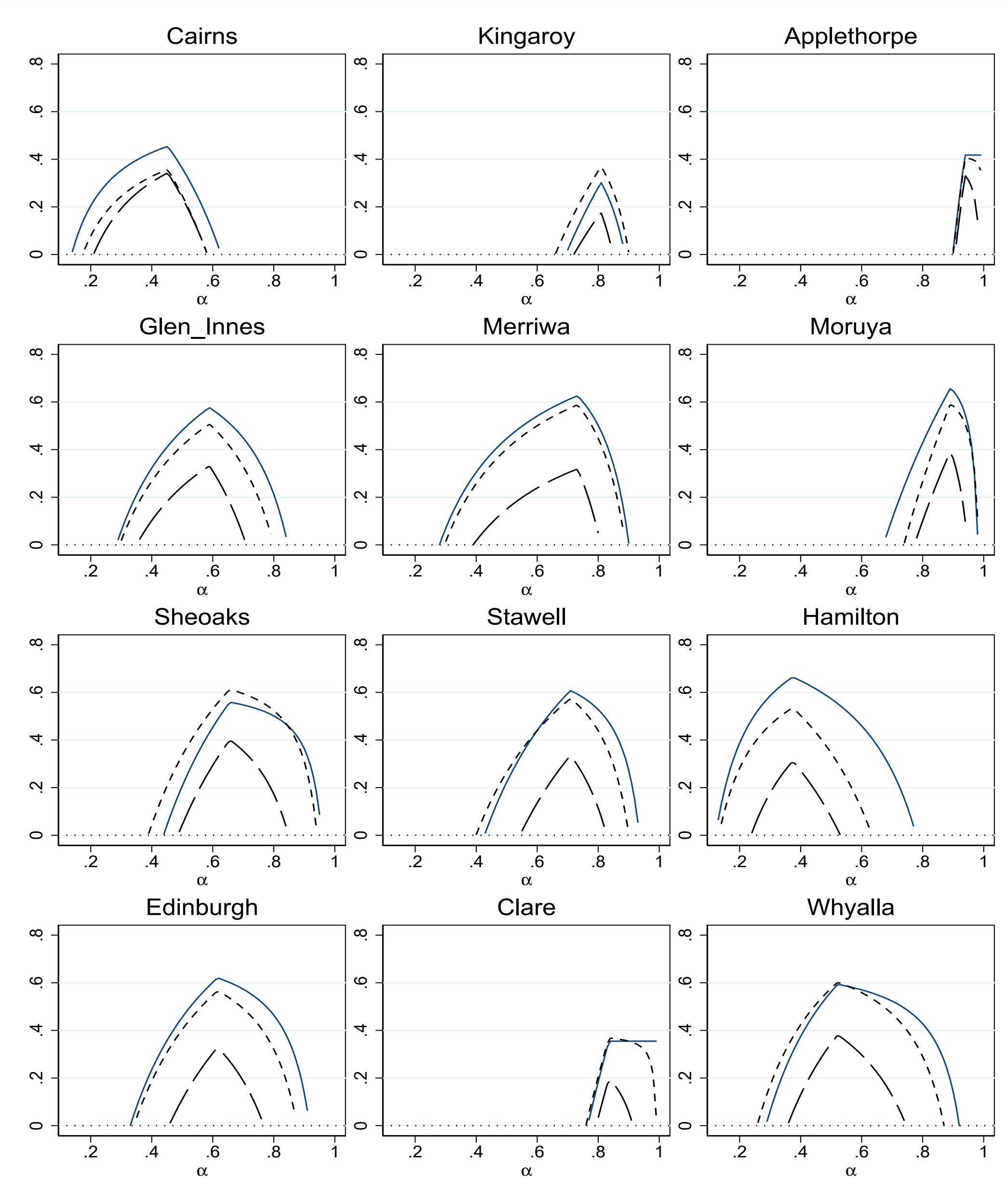

Figure 8: The relative economic value V of wind drought forecasts at the horizons of 1-day (solid line), 3-day (short dashed line) and 7-day (long dashed line).

The relative economic values of multi-horizon forecasts for wind drought, calculated using equation (11) for a cost-loss ratio $\alpha \in (0, 1)$, are shown in Figure 8. Note that only positive relative economic values are illustrated. It is apparent that using BOM forecasts of wind drought provide users in NSW, VIC and SA with positive relative economic value over a wide range of values for the cost-loss ratio. There are four stations, namely, Kingaroy, Applethorpe, Moruya and Clare, where the positive value is limited to cost-loss ratios greater than 0.8. It is clear from Table 3 that these stations all have very high unconditional probabilities of wind drought. It would appear therefore that because wind drought occurs very frequently, decision-makers favour taking action without reference to forecasts, except when the cost of taking action is almost as high as the prevented loss, that is $\alpha \approx 1$.

The economic value of forecast revisions can be ascertained by comparing the values of V across horizons. Given the same value of α, the relative economic value V of the initial forecasts made 7 days out is always less than that of the revised forecasts made 3 days out, suggesting a positive economic gain. However, further revisions made between the 3-day and 1-day horizons out do not necessarily result in a significant increase in economic value. There may therefore be little gain to forecast end-users in waiting for the 1-day forecasts to be published.

Finally, it is worth noting that taken as a group the forecasts for the QLD stations, Cairns, Kingaroy and Applethorpe appear to be of very limited economic value. The range of α for which any value is provided is particularly small by comparison with stations in other states. This result is particularly concerning given that all three stations have proximity to major wind farms. Of the three stations, the forecasts provided for Cairns are definitely superior to the other two in terms of the range of α over which the forecasts provide positive economic value. Interestingly, however, the 1-day horizon forecasts for Cairns appear to be significantly better than the 3-day horizon forecasts, suggesting a positive payoff for waiting to take any action.

7.2 Wind gusts

The second wind anomaly considered relates to bushfires. On hot and dry days, strong gusty wind increases the risk of loosing control of bushfires, thereby endangering nearby communities. Accurate forecasts of wind gusts during bushfire seasons therefore contribute to effective bushfire control and management, thereby having the potential to mitigate damages.[13] For the purposes of this paper, a wind gust is taken to be a wind speed of $4.17 m/s$ (equivalent to 15 km/h) because this speed represents a threshold that makes a significant difference in the behaviour of bushfires in the open.[14] An anomaly occurs if there are 3 or more wind gusts between 12:00 and 17:00 during the bushfire season from November to February.

Table 4

Sample unconditional probability of wind gust anomalies, p, together with the sample relative false alarm rate F and the sample relative hit rate H for forecasts at the horizons of 1-day, 3-day and 7-day, respectively.

	p	F_1	F_3	F_7	H_1	H_3	H_7
Cairns	0.17	0.13	0.10	0.16	0.68	0.50	0.39
Kingaroy	0.11	0.35	0.25	0.23	0.71	0.64	0.44
Applethorpe	0.02	0.61	0.61	0.61	0.83	0.75	0.83
Glen Innes	0.39	0.24	0.15	0.17	0.84	0.67	0.56
Merriwa	0.29	0.16	0.10	0.11	0.66	0.58	0.29
Moruya	0.08	0.23	0.26	0.21	0.72	0.65	0.44
Sheoaks	0.19	0.25	0.20	0.28	0.87	0.80	0.62
Stawell	0.30	0.27	0.21	0.38	0.84	0.80	0.71
Hamilton	0.50	0.21	0.16	0.30	0.75	0.54	0.56
Edinburgh	0.31	0.36	0.34	0.45	0.86	0.80	0.70
Clare	0.08	0.62	0.58	0.61	0.98	0.95	0.82
Whyalla	0.63	0.40	0.40	0.68	0.95	0.92	0.92

It is apparent from Table 4 that the wind gust anomaly is an infrequent event by comparison with the wind drought (see Table 3). In only two cases, Hamilton and Whyalla, is $p \geq 0.5$ observed. The behaviour of the false alarm rate, F, is counter-intuitive as it does not decrease significantly as the forecast horizon shrinks. In fact in many cases, though not in either Hamilton or Whyalla, the false alarm rate actually increases as the forecast horizon decreases. By contrast, the hit rate behaves very much as expected and increases as the forecast horizon decreases.

Figure 9 shows the economic value of wind gust forecasts at all the stations for the various forecast horizons. Once again the four stations at which forecasts appear to be least effective are Kingaroy, Applethorpe, Moruya and Clare, which also happen to be where wind gusts are most infrequent.

[13]Note that in addition to wind speed, wind direction, temperature and humidity are also prevailing weather conditions that affect bushfire behaviour.

[14]See, for example, `https://www.ga.gov.au/scientific-topics/community-safety/bushfire`.

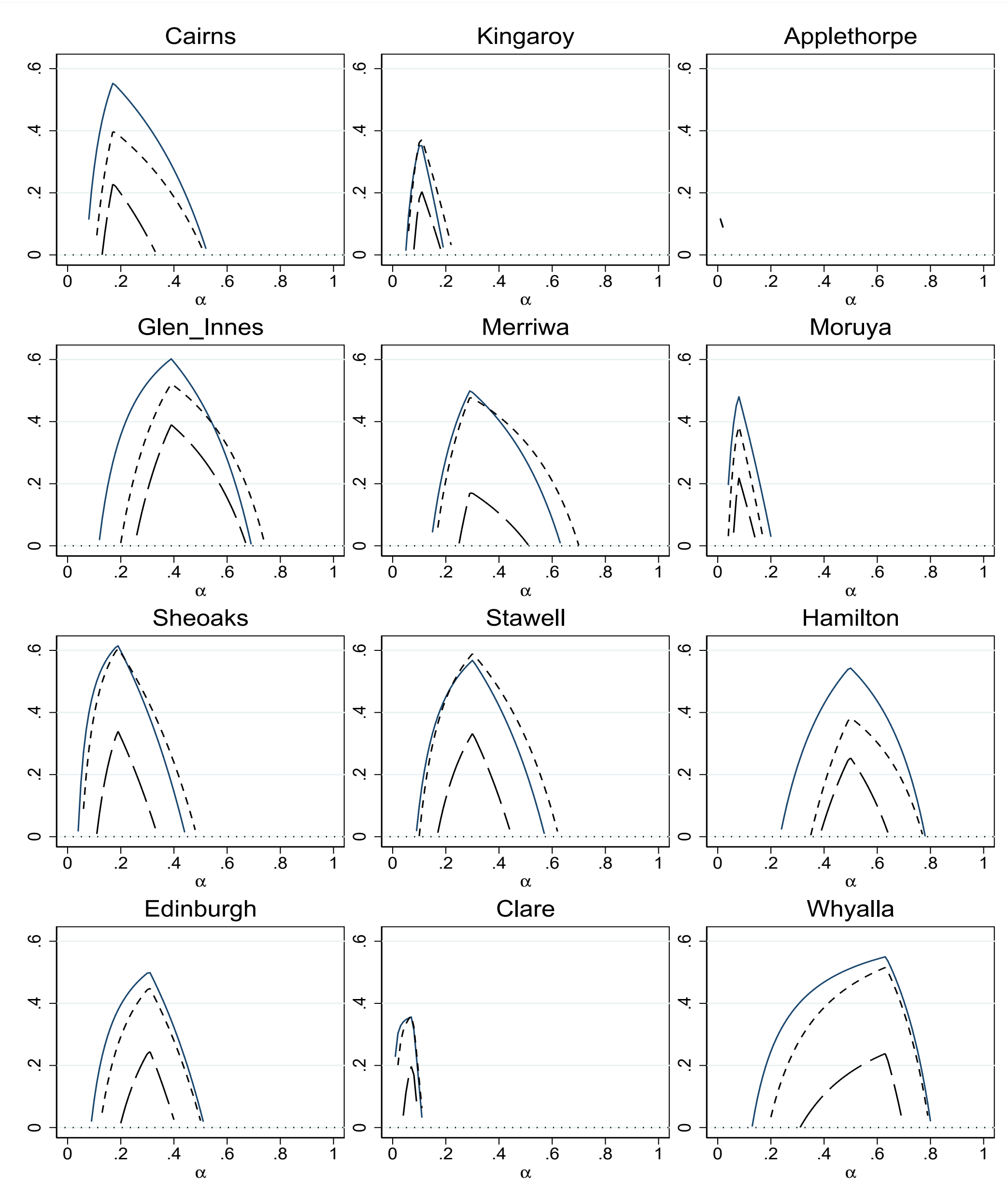

Figure 9: The relative economic value V of wind gust forecasts made at the horizons of 1-day (solid line), 3-day (short dashed line) and 7-day (long dashed line).

At these four stations, positive values of V are only associated with the values of α that are close 0. Since the anomaly almost never occurs, the best course of action for decision-makers is to ignore the wind gust forecasts unless the cost of taking action is negligible compared to the expected benefit of the intervention. As for the case of the wind drought forecasts, in general the 7-day forecasts are not competitive and more economic value can be obtained by waiting for the forecast revisions. There is, however, an added interesting twist to these results, namely, that for some stations as α increases the economic value of the 3-day forecasts is seen to be greater than that of the 1-day forecasts. This pattern is not observed in the wind drought forecasts. Finally, as with the previous case, the QLD forecasts appear to be less useful than those for stations in other states.

To conclude the discussion of the economic value of the official BOM forecasts a number of comments of a general nature may be made. It is clear that despite failing the tests of econometric rationality, the forecasts do offer some positive economic value to mitigate the effects of wind anomalies. It is clear, however, that in some cases the positive value is confined to very small ranges in the cost-loss ratio, α. In particular, the forecasts for QLD do not provide as much value as the forecasts for stations in other areas. Moreover, the maximum value attained by V is of the order of 0.6 which is substantially below its maximum value of 1. Even when the metric of economic value is used, there is substantial room for official BOM forecasts to improve.

8 Conclusion

This paper has examined the performance of the official wind-speed forecasts produced by the Australian Bureau of Meteorology for 12 weather stations along the eastern seaboard of Australia. The production of the official wind speed forecasts is a two-stage procedure that involves adding a human touch to the output of a complex large-scale numerical weather prediction model. The hourly forecasts are produced up to 7 days in advance of the target and are updated daily. The stations chosen represent a mix of locations important to existing and potential wind farms and for deploying resources in the event of bush fires. Although the stations are broadly speaking located in eastern states, there is a mix of coastal and inland positioning.

The first framework within which the forecasts are examined is that of econometric rationality, which in essence requires that information flow over the forecast period is used efficiently when revising the forecasts. On the face of it, the evidence is completely conclusive: the BOM multi-horizon forecasts are not rational forecasts in the econometric sense. While the task of passing econometric tests of rationality is quite demanding, the concept of forecasting rationality provides a formal framework within which it is useful to examine forecasting performance. The second metric used to judge the forecasts is that of economic value. Here the evidence in terms of a cost-loss model suggests that the BOM forecasts do indeed provide some positive value, but the performance across the various stations is patchy and there is room for improvement both in absolute terms and in the consistency across geographical areas.

This research demonstrates the importance of understanding BOM forecast revisions for various wind-speed forecast users. The over-predicted BOM forecasts may help with making a successful plan for controlling bushfires and protecting the public, but they can also mislead wind energy generators in making suboptimal commitment decisions or maintenance schedules, and hence cause large operating losses. BOM forecast users may need to re-adjust the official forecasts or choose forecasts made at longer horizons in their decision making to achieve an optimal level of individual benefits. Note that the non-increasing pattern in both revision effort measure and accuracy improvement measure as horizon shrinks are observed in some stations, implying that the most valuable information are incorporated in the longest horizon forecasts rather than in short horizon forecasts.

In terms of the way forward, at the very least this research should prompt serious thinking in terms of forecasting service provided by the BOM as an input into the current strengthening of services currently in progress in terms of the BOM's *Strategy 2017–2022* roadmap (Bureau of Meteorology, 2017) for maximising the value and impact that the BOM delivers for Australia.

References

Brock, F., & Richardson, S. (2001). *Meteorological Measurement Systems*. Oxford: Oxford University Press.

Bureau of Meteorology (2017). *Strategy 2017-2022*. `http://www.bom.gov.au/inside/downloads/Bureau-of-Meteorology-Strategy-2017-2022.pdf`.

Clements, M. P. (1997). Evaluating the rationality of fixed-event forecasts. *Journal of Forecasting*, *16*, 225–239.

Clements, M. P., & Taylor, N. (2001). Robust evaluation of fixed-event forecast rationality. *Journal of Forecasting*, *20*, 285–295.

Corradi, V., Fernandez, A., & Swanson, N. (2009). Information in the revision process of real-time datasets. *Journal of Business & Economic Statistics*, *27*, 455–467.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In G. Elliot, C. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting*. Amsterdam: North-Holland.

Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, *105*.

Elliott, G., Komunjer, I., & Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, *2*, 1107–1125.

Elliott, G., Komunjer, I., & Timmermann, A. (2008). Biases in macroeconomic forecasts: Irrationality or asymmetric loss? *Journal of the European Economic Association*, *6*, 122–157.

Foley, M., & Loveday, N. (2020). Comparison of single-valued forecasts in a user-oriented framework. *Weather and Forecasting*, *35*, 1067–1080.

García-Portugués, E., Barros, A. M., Crujeiras, R. M., González-Manteiga, W., & Pereira, J. (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stochastic Environmental Research and Risk Assessment*, *28*, 1261–1275.

Granger, C. W., & Pesaran, M. H. (2000a). A decision-based approach to forecast evaluation. In W. Chan, W. Li, & H. Tong (Eds.), *Statistics and Finance: An Interface* (pp. 261–278). London: Imperial College Press.

Granger, C. W., & Pesaran, M. H. (2000b). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, *19*, 537–560.

Griffiths, D., Foley, M., Ioannou, I., & Leeuwenburg, T. (2019). Flip-flop index: Quantifying revision stability for fixed-event forecasts. *Meteorological Applications*, *26*, 30–35.

Isiklar, G., & Lahiri, K. (2007). How far ahead can we forecast? Evidence from cross-country surveys. *International Journal of Forecasting*, *23*, 167–187.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Koenker, R., & Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*, 143–156.

Lahiri, K. (2012). Comment. *Journal of Business & Economic Statistics*, *30*, 20–25.

Lovell, M. C. (1986). Tests of the rational expectations hypothesis. *The American Economic Review*, *76*, 110–124.

Lynch, K. J., Brayshaw, D. J., & Charlton-Perez, A. (2014). Verification of European subseasonal wind speed forecasts. *Monthly Weather Review*, *142*, 2978–2990.

McLay, J. G. (2011). Diagnosing the relative impact of "sneaks," "phantoms," and volatility in sequences of lagged ensemble probability forecasts with a simple dynamic decision model. *Monthly Weather Review*, *139*, 387–402.

Milligan, M. R., Miller, A. H., & Chapman, F. (1995). *Estimating the economic value of wind forecasting to utilities*. Technical Report National Renewable Energy Lab., Golden, CO (United States).

Monk, K., Guérette, E.-A., Paton-Walsh, C., Silver, J. D., Emmerson, K. M., Utembe, S. R., Zhang, Y., Griffiths, A. D., Chang, L. T.-C., Duc, H. N. et al. (2019). Evaluation of regional air quality models over Sydney and Australia: Part 1—meteorological model comparison. *Atmosphere*, *10*, 374.

Murphy, A. H. (1977). The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review*, *105*, 803–816.

Nordhaus, W. D. (1987). Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics*, *69*, 667–674.

Patton, A., & Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, *140*, 884–918.

Patton, A. J., & Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, *30*, 1–17.

Pesaran, M. H., & Skouras, S. (2004). Decision-based methods for forecast evaluation. In *A Companion to Economic Forecasting* chapter 11. (pp. 241–267). John Wiley and Sons, Ltd.

Pinson, P., & Hagedorn, R. (2012). Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Meteorological Applications*, *19*, 484–500.

Richardson, D. S. (2000). Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, *126*, 649–667.

Smith, L. A. (2016). Integrating information, misinformation and desire: improved weather-risk management for the energy sector. In *UK Success Stories in Industrial Mathematics* (pp. 289–296). Springer.

Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium 2010* (pp. 1–8). IEEE.

Sweeney, C. P., Lynch, P., & Nolan, P. (2013). Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. *Meteorological Applications*, *20*, 32–40.

Xie, L., Bao, S., Pietrafesa, L. J., Foley, K., & Fuentes, M. (2006). A real-time hurricane surface wind forecasting model: Formulation and verification. *Monthly Weather Review*, *134*, 1355–1370.

Yetman, J. (2018). *The perils of approximating fixed-horizon inflation forecasts with fixed-event forecasts*. Technical Report BIS Working Paper.