WILEY

Enhancing survey-based investment forecasts

Ciaran Driver[1] | Nigel Meade[2]

[1] School of Finance and Management, SOAS, University of London, London, UK

[2] The Business School, Imperial College London, London, UK

Correspondence
Nigel Meade, The Business School, Imperial College London, London SW7 2AZ, UK.
Email: n.meade@imperial.ac.uk

Funding information
British Academy, Grant/Award Number: SG150307

Abstract

We investigate the accuracy of capital investment predictors from a national business survey of South African manufacturing. Based on data available to correspondents at the time of survey completion, we propose variables that might inform the confidence that can be attached to their predictions. Having calibrated the survey predictors' directional accuracy, we model the probability of a correct directional prediction using logistic regression with the proposed variables. For point forecasting, we compare the accuracy of rescaled survey forecasts with time series benchmarks and some survey/time series hybrid models. In addition, using the same set of variables, we model the magnitude of survey prediction errors. Directional forecast tests showed that three out of four survey predictors have value but are biased and inefficient. For shorter horizons we found that survey forecasts, enhanced by time series data, significantly improved point forecasting accuracy. For longer horizons the survey predictors were at least as accurate as alternatives. The usefulness of the more accurate of the predictors examined is enhanced by auxiliary information, namely the probability of directional accuracy and the estimated error magnitude.

KEYWORDS
business surveys, directional forecasting, forecasting accuracy, forecasting evaluation, time series

1 | INTRODUCTION

Business surveys, if timely and accurate, enable a first view of emerging trends that may not be apparent from a time series analysis of official statistics or econometric models. We examine the accuracy of various predictors contained in a national business survey and investigate whether the variations in accuracy of directional and point forecasts can be explained by information available at the time the forecasts were made. We further study how that accuracy changes over a range of horizons, as different business decisions necessitate different response times so that forecast horizons useful, say, for pricing or inventory decisions will be shorter than those relevant to longer term capacity decisions.

Our data source is a long-established, reputable business survey of South Africa carried out by the Bureau

of Economic Research (BER) at the University of Stellenbosch. The survey contains a variety of business information, but we consider only indicators of capital investment intentions. Assessing and improving the accuracy of investment forecasts is important because the investment spend affects both the demand side and the supply side of the economy. However, investment is also one of the most difficult series to predict accurately.[1] Our paper assesses the added value of the BER survey series in respect of directional and point forecasts of fixed investment. In addition, we enhance

---

[1] Investment reflects jumps in expectations regarding events far in the future and is thus more autonomous (and more variable) than other aggregates. The average 1-year-ahead root mean square error (RMSE) for fixed investment taken over a number of forecasting models has been shown to be four times larger than for gross domestic product (GDP) growth (Granger, 1994).

the survey predictions with a probability of directional accuracy and a predicted magnitude of the error in a point forecast, both based on contemporaneous information.

There are at least four relevant predictors of manufacturing capital investment in the BER quarterly survey: (a) intentions for the current quarter (the Nowcast) (b) intentions for the next quarter; (c) intentions for a year ahead in respect of a narrow investment category; and (d) a general indicator of business conditions for a year ahead. The definitive values of total manufacturing capital investment are published after some delay and revision by the South African Reserve Bank. Throughout our study, to replicate forecasting in real time, we pay particular attention to the timing of the publication of the survey and the official data. We address both directional accuracy and point accuracy of the survey-based forecasts; in both cases a time series forecast is used as a benchmark. Having assessed the directional accuracy of the survey forecasts using a battery of tests, we model the probability of a correct directional prediction using data available at the time of prediction. This gives an objective measure of confidence in the survey prediction, which is an important issue as turning points are notoriously difficult to predict. Addressing the accuracy of point forecasts, we consider several ways in which survey forecasts might be improved: rescaling the predictions; including the predictions in time series regressions; and by combining time series and survey results to produce hybrid forecasts. In parallel with providing a confidence measure for directional accuracy, we model the magnitude of survey forecast errors using data available at the time of prediction.

The structure of the paper is as follows. In Section 2, we review the context of the business survey—provided by the South African Bureau of Economic Research—and report on related literature. Methodologies and measures for assessing forecast accuracy are assessed in Section 3. We describe our data set in Section 4. In Section 5, we assess the directional accuracy of the BER survey predictions; we investigate the consistency of model accuracy before, during, and after the financial crisis; we also propose several variables likely to affect the stability of the survey predictors using data available at the time of the survey and, using these stability variables, we model the probability of directional accuracy. In Section 6, we investigate the accuracy of survey-based point forecasts. We also examine the effect of the financial crisis on our results and we model the magnitude of survey forecast errors using the stability variables proposed earlier. We summarize our findings and offer our conclusions in Section 7.

2 | CONTEXTUAL BACKGROUND AND RELATED WORK

The context of our study is the South African economy, whose trajectory is in some respects a puzzle. Its recent growth rate has not been above 4% for any sustained period, except for the 4 years before the financial crisis, when it averaged about 5%. The economy is widely perceived as being held back by a variety of poorly understood "soft" factors inhibiting faster investment and growth. Evidence suggests a positive relationship between investment and the growth of GDP per worker in non-OECD countries (Bond, Leblebicioğlu, & Schiantarelli, 2010). However, a particular concern for South Africa is that capital investment in the decade since independence only contributed about a quarter of GDP growth compared with almost a third in a panel of 10 similar countries selected on the basis of income and population (Eyraud, 2009); such a pattern has been shown to have been detrimental to the country's productivity growth (Arora & Bhundia, 2003).

Investment did accelerate somewhat in the decade before the financial crisis (Fedderke, 2009) but there has been concern that this was concentrated in the nontradable sectors (Frankel, Smit, & Sturzenegger, 2008). Comparatively, the overall investment-to-GDP ratio for the decade from 2000 was far lower than for similar countries (Viegi, 2014). The financial crisis adversely affected South Africa from the first quarter of 2009 and resulted in lower growth for a considerable period of time (OECD, 2017). The manufacturing investment rate declined sharply and, despite a considerable devaluation, the sector's employment fell by nearly 15%. The main constraints on corporate investment appear to be nonfinancial factors, reflecting fears over matters such as corruption, crime, and infrastructure, and also the challenges posed by the dual economy structure and extreme inequality that characterize the country (World Bank, 2010, 2018). These constraints may also affect foreign direct investment (FDI): In recent years FDI has averaged less than 5% of overall investment, and the manufacturing sector tends to receive only about a fifth of these funds (South African Reserve Bank, 2018).

Some improvement is now expected. Reflecting recent changes in the administration and political reforms, the OECD currently estimates that real GDP growth will accelerate over the next 2 years to a little over 2%, with the more volatile capital expenditure rising at about twice this rate (OECD, 2018). Nevertheless, these forecasts do not represent a fundamentally changed trajectory for South Africa.

Establishing, and possibly improving, the accuracy of the BER data is important for at least three reasons. First,

accurate business survey data improves the analytic toolkit for policymakers charged with setting interest rates and other macroeconomic determinants of investment. Previous forecasting exercises for South Africa have suggested difficulties occasioned by structural breaks (Aron & Muellbauer, 2002). Secondly, improved forecasting for investors would lower the cost of capital for firms; in South Africa, an elevated cost of funding costs reflects higher than normal economic uncertainty, and this has been shown to be detrimental to capital investment (Fedderke, 2004; Fedderke & Simkins, 2012). Finally, there may be an additional direct boost to capital investment from lowering uncertainty, independent of the cost of capital (Bond, Soderbom, & Wu, 2011; Chirinko & Schaller, 2009).

There is a substantial international literature on the forecast accuracy of business surveys and on their use in helping to explain movement in variables such as prices, output, employment, or investment.[2] For South Africa, some analytical assessments have been made of published forecasts. Van Walbeek (2013) reviewed the comparative accuracy of surveys and several commercial macroeconomic models for 2000–2012, with results that were not encouraging; adaptive predictions were found to beat the survey forecasts except after 2009, when the errors of the former increased for the set of variables studied. The accuracy of the BER macroeconomic forecasts—based partly on its surveys—has also been tested for an earlier period for some variables (measures of output, expenditure, and interest rates) with a finding that it predicts best up to two or three quarters ahead and beats a naïve model at four quarters ahead only for gross domestic expenditure and the prime interest rate (Van Walbeek & Sessions, 2007).

3 | MEASURES OF FORECASTING ACCURACY

Various properties have been used to describe a good forecast, rationality, value, efficiency, and unbiasedness. According to Muth's (1961) definition, rational expectations (forecasts) fully reflect currently available information. Merton (1981) defines a directionally rational forecast as one that would cause nobody to expect a direction of change opposite to the forecast. The problem of measuring the accuracy of expectations survey data in forecasting economic variables has been addressed by several authors. In Section 3.1, we will describe several tests of the accuracy and value of a set of directional

forecasts. We further discuss tests of unbiasedness and efficiency in Section 3.2.

3.1 | Tests of accuracy of directional forecasts

An early test of the value of a set of directional forecasts was proposed by Henriksson and Merton (1981) and it provides a useful framework for later tests. If the variable to be forecast is A_t and the forecast is F_t, where $F_t = E(A_t | \Theta)$ and Θ is the relevant information set, then the success of the directional forecasting exercise can be described by the following table, where n denotes the count in each cell and P denotes the probability of falling in the cell.

		Forecast: F_t		
		< 0	**≥0**	
Actual: A_t	<0	n_{11} (P_{11})	n_{21} (P_{21})	n_{01} (P_{01})
	≥0	n_{12} (P_{12})	n_{22} (P_{22})	n_{02} (P_{02})
		n_{10} (P_{10})	n_{20} (P_{20})	N (1)

Henriksson and Merton argued that for the forecast to have value the sum of these conditional probabilities, $\Pr(F_t < 0 | A_t < 0)$ and $\Pr(F_t \geq 0 | A_t \geq 0)$, should exceed unity (if the forecast and the actual were independent, then this sum would equal unity). The value of the forecast is proportional to the amount by which unity is exceeded. Under the null hypothesis that $\Pr(F_t < 0 | A_t < 0) + \Pr(F_t \geq 0 | A_t \geq 0) = 1$, the distribution of n_{11} is hypergeometric. Thus, if

$$p - \text{value} = 1 - \sum_{x=\max(0,n_{10}-n_{02})}^{n_{11}-1} \frac{\binom{n_{01}}{x}\binom{n_{02}}{n_{10}-x}}{\binom{N}{n_{10}}}$$

is less than a critical value, say 5%, then the null hypothesis can be rejected and the forecast has value. This test is equivalent to Fisher's exact test for independence in a 2×2 contingency table. A related approach is to use a straightforward χ^2 test, usually with Yates' correction, where under the null hypothesis of independence

$$\frac{N\left(|n_{11}n_{22} - n_{12}n_{21}| - \frac{N}{2}\right)^2}{n_{10}n_{20}n_{01}n_{02}} \sim \chi_1^2.$$

Pesaran and Timmermann (1992) developed a nonparametric test for measuring the directional accuracy of forecasts and demonstrated it using CBI Industrial Trends Survey data. They derive a predictive failure test for the case where A_t and F_t have m different states—

that is, an $m \times m$ contingency table. Their test statistic is derived as follows.

The "observed" probability of a correct forecast $p_c = P_{11} + P_{22}$; under their H_0, the probability of a correct forecast is $p_1 = P_{01}P_{10} + P_{02}P_{20}$. The variances of these terms are

$$V(p_c) = \frac{p_1(1 - p_1)}{N}$$

and

$$V(p_1) = \frac{(2P_{02}-1)^2 P_{20}(1 - P_{02}) + (2P_{20}-1)^2 P_{02}(1 - P_{20}) + \frac{4P_{02}P_{20}(1 - P_{02})(1 - P_{20})}{N}}{N}.$$

The test statistic

$$\frac{(p_c - p_1)^2}{(V(p_c) - V(p_1))} \sim \chi_1^2.$$

Their null hypothesis is that $H_0^{PT}: \sum_{i=1}^{m} (P_{ii} - P_{i0}P_{0i}) = 0$, which is more general than that of Henriksson and Merton, $H_0^{HM}: P_{ij} = P_{i0}P_{0i}$ for all (i, j). However, for $m = 2$, the two hypotheses coincide.

In the case where there is an asymmetry between $\Pr(A_t < 0)$ and $\Pr(A_t \geq 0)$, goodness-of-fit measures may fail to identify the ability of a forecasting system to evaluate the odds of the occurrence of a low-probability event against its nonoccurrence. A measure (rather than a test) is Kuiper's score (see, for example, Doswell, Davies-Jones, & Keller, 1990).

Kuiper's score $=$ Hit rate $-$ False alarm rate, where the

Hit rate $= \Pr(F_t < 0 \text{ and } A_t < 0 | A_t < 0)$ and False alarm rate $= \Pr(F_t < 0 \text{ and } A_t \geq 0 | A_t \geq 0)$. From the table

$$\text{Kuiper's score} = \frac{n_{11}}{n_{01}} - \frac{n_{12}}{n_{20}},$$

but a higher score is preferable.

3.2 | Tests of value, unbiasedness and efficiency

Stepping away from the contingency table approach, Mincer and Zarnowitz (1969) pioneered the evaluation of economic forecasts. They suggested that the realized observation be regressed on the forecast: $A_t = \alpha_1 + \beta_1 F_t + \varepsilon_{1t}$. The forecast is efficient if $H_0^{MZ1}: \beta_1 = 1$ cannot be rejected, and is unbiased if $H_0^{MZ2}: \alpha_1 = 0$ cannot be rejected. Cumby

and Modest (1987) consider the accuracy of the directional forecast. The forecast is regressed on a binary variable indicating the actual direction of change. $F_t = \alpha_2 + \beta_2 \delta(A_t) + \varepsilon_{2t}$, where

$$\delta(A_t) = \begin{cases} 1 & \text{if } A_t \geq 0 \\ 0 & \text{if } A_t < 0 \end{cases}.$$

Under $H_0^{CM}: \beta_2 = 0$; that is, the forecast has no value. Holden and Peel (1990) test the efficiency of a set of forecasts using a regression that follows on from the Mincer and Zarnowitz approach. Defining the forecast error as $e_t = A_t - F_t$, their test looks for first-order autocorrelation in the errors, $e_t = \alpha_3 + \beta_3 e_{t-1} + \varepsilon_{3t}$, if $H_0^{HP}: \beta_3 = 0$ is rejected, then the forecast is inefficient because forecasts could be improved by correcting for the autocorrelation.

3.3 | Tests of comparative accuracy

During our analysis, we use root mean squared error (RMSE) as a summary measure of accuracy and we use absolute errors to highlight the changing magnitude of errors in different circumstances. To ascertain the significance of differences in forecasting accuracy between alternative forecasting models, we use the following tests.

The Diebold and Mariano (1995) test compares the value of two sets of forecasts of the same variable. At time t, two forecasts are available, F_{1t}^h and F_{2t}^h, for a horizon h, where $h = 1, ..., H$ generating errors, $e_{it+h}^h = A_{t+h} - F_{it}^h$. If the loss function is measured in terms of mean squared error, then, at a horizon of h, the benefit of using F_{1t} over F_{2t} is

$$\bar{d}_h = \frac{1}{N} \sum \left(|e_{2t}^h|^p - |e_{1t}^h|^p \right),$$

where p is either 1 or 2. The test statistic is

$$S_h = \frac{\bar{d}_h}{\sqrt{V(\bar{d}_h) \sim N(0, 1)}}$$

and $V(\bar{d}_h)$ is estimated by

$$\frac{1}{N}\left(\gamma_0 + 2\sum_{j=1}^{h-1}\gamma_j\right),$$

where γ_j is the j lagged autocovariance. Harvey, Leybourne, and Newbold (1997) modified the test statistic to

$$S_1^* = \left(\frac{N + 1 - 2h + \dfrac{h(h-1)}{N}}{N}\right)S_1 \sim t_{N-1}$$

to improve the distributional accuracy of the test.

Another test that can be used to compare the pairwise accuracy of forecasting methods is the Friedman test (see Conover, 1999). The test is nonparametric and is based on the rankings of the accuracy of each method for each observation. The overall null hypothesis is that the ranking of each method is equally likely for each observation; that is, there is no difference between methods. If the null hypothesis is rejected, then at least one method is better or worse than the others; in this case a pairwise comparison can be made to detect where method X is significantly better than method Y.

4 | THE DATA SET

Here, we first describe the objectives of the BER survey and the way the data are collected; we identify the specific questions relating to the predictions to be studied. Secondly, we explain the timing of the publication of information during a quarter, as the data available to correspondents at the time they complete the survey are an important constituent of our study. Thirdly, we carry out preliminary tests on the survey and official data, and present some stylized statistics.

4.1 | Purpose of the BER survey and details of the questions under study

The BER manufacturing business tendency survey data comprise qualitative series for manufacturing (ISIC code 3) on investment intentions, business climate indicators, and perceived business constraints, collected quarterly by mail with an industrial and geographical breakdown.[3] The manufacturing survey is part of a broader survey of the formal economy that includes building, financial,

———————————

[3]The format is similar to that of other surveys carried out in the UK by the employers' organization (CBI), in the EU by Eurostat, and in Australia by ACCI and the Westpac Bank. Further details on the BER survey are contained in: http://stats.oecd.org/mei/default.asp?lang=e&subject=6&country=ZAF.

and services sectors. In total, 4,100 questionnaires are sent out every quarter and the response rate has remained between 40% and 45% over the last three decades. On average, there are about 1,200 manufacturing units included in the survey. The sampling frame is updated every 2 years. A panel based on deliberate sampling is used as the same firms are approached from one survey to the next so that the majority of responses between consecutive surveys are from the same companies. The period over which the survey data is collected is 3 weeks. All the respondents that completed a questionnaire receive a copy of the survey results sent out about 6–8 weeks before they would receive the questionnaire for the next quarter. Nearly all the manufacturing survey questions require that the respondent indicate if a particular activity is "up," "remained the same" or "down," where the reference period is the current quarter compared to the same quarter 1 year ago.

Each quarter, the survey respondents are asked to record their estimated development (up, same or down in comparison with 1 year ago) for the current quarter (e.g., the first quarter January to March) on a form to be returned at around the beginning of the last month (e.g., March). In addition, respondents are asked to fill in the same information for the expected development for the next quarter (e.g., Q2). Net balance statistics (i.e., percentage "up" less percentage "down") are constructed from both of these sets of data, being aggregated up from the individual responses using "number of factory workers" as weights. Specifically, the survey form asks:

> *Compared with the same quarter a year ago [is] fixed investment:*
> *up/same/down [estimated for current quarter]?*
> *up/same/down [expected for the next quarter]?*

We will term the aggregated balance (ups minus downs) for the estimated series for the current quarter XA and the corresponding balance for the same period made a quarter earlier as XE. XA may thus be thought of as an updating (Nowcast) of the previous quarter *XE* forecast. We further consider the survey questions:

> *In comparison to current levels in your sector, what do you expect the following to be in 12 months' time:*
> *Real investment in machinery and equipment (M&E)?*
> *General business conditions?*

We refer to the aggregated balance of these variables as XI and XC, respectively. Our time sample for these four survey variables is from 1992:Q3 to 2015:Q2.

We relate the BER survey forecasts of manufacturing fixed investment to corresponding data from the published national accounts, in particular the quarterly series of real gross capital formation (level) for manufacturing investment (seasonally adjusted) published by the South Africa Reserve Bank (KBP6082D).[4] These official data are collected early in each quarter for the previous quarter (e.g., collected in January of the current year for investment in Q4 of the previous year) and then published towards the end of the collection quarter (e.g., in March). Revisions are annual and are made when completing the third quarter statistics with periodic major revisions. A time line for the collection and publication of all these statistics is given later in Section 4.2.

It has been shown in Pesaran (1984, 1987) that the survey data balance statistic corresponds, under some restrictions, to a rate of change over the interval to which the survey question refers. The use of a balance statistic is but one method of transforming the qualitative directional data to a quantitative series, and there is a large literature on such transformations (Biau, Erkel-Rousse, & Ferrari, 2006; Lui, Mitchell, & Weale, 2011; Mitchell, Smith, & Weale, 2005). One critique of the balance measure is that it assumes symmetry and constancy of the indifference limens on either side of zero that are encompassed by a "no change" response. For some nominal series such as cost measures or stock market data one or other of these assumptions has been shown to be questionable (Breitung & Schmeling, 2013; Smith & McAleer, 1995). However, for production and investment data, the balance statistic representation has been shown to be adequate (Driver & Urga, 2004). We are obliged to use the balance statistics in this study as our data source does

not contain the individual ups and downs for all the sample period.[5]

As the BER survey data refer to an annual change, we constructed the corresponding annual growth rate for the official investment series as the fourth difference of the natural log of KBP6082D (D4LD). We use y_t to represent the dependent variable:

$$y_t = 100(1 - B^4) \ln(\text{KBP6082D}: \text{Gross fixed capital formation}).$$

The dependent variable is available quarterly from March 31, 1980, to June 1, 2015. As indicated earlier, we use XA_t to represent the Nowcast and XE_t to represent the Expected. Note that for both these variables the subscript t relates to the period that the forecast is made for, not the origin of the forecast. The quarterly data available for Nowcast and Expected are 92 observations from September 30, 1992, to June 1, 2015.

We show the time series used in Figure 1, indicating the sections of the data used for measuring directional accuracy and for point forecasting estimation and accuracy measurement. In Figure 1a, the actual annual growth rate is shown with the shorter-term forecasts, XA and XE (the time series for y_t before March 1992 is not shown). In Figure 1b, the actual annual growth rate is shown with the longer-term forecasts, XI and XC.
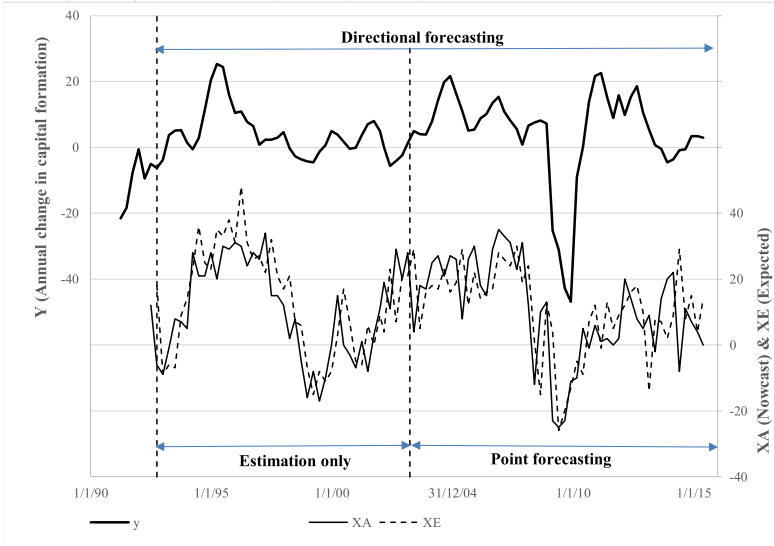
4.2 | The timing of information availability during each quarter

To make a comparison of forecasting accuracy that is useful in practice, we must take into account the timing of the publication of information. In Figure 2, we show the timing of publication for the official data and the short-term forecasts. This timing divides the quarter (the main time unit of interest) into three phases. The information set available during these phases for forecasting y_t is

Phase A	y_{t-2};	XA_{t-1};	XE_t
Phase B	y_{t-2};	XA_t;	XE_t
Phase C	y_{t-1};	XA_t;	XE_t

Our comparison between time series and survey-based models should be seen in the light of the revision policy for the official data used by time series forecasts. Minor amendments are made in the third quarter of each year and major rebasing takes place at approximately 5-year intervals. We have assumed, *faute de mieux*, that forecasters have access to the revised data, but at any point in time they may in fact be using several years of data less accurate than the revised figures. This gives a built-in advantage to the time-series model, and

---

[4]This choice of official data is partly based on interviews conducted with the Director and Manager of the BER, who confirmed that KBP6082D is the statistic they use to compare their forecasts. We also experimented with the nominal series published by the Reserve Bank but this series is trended whereas the survey series are mean reverting.

[5]For a shorter period (2001:Q1–2015:Q2), the micro data are available for XA and we were able to test the implied restriction in using the balance statistic and to reject the hypothesis that decomposing the balance series increased explanatory power in relation to the official investment series. In a regression of y_t on a constant and eight lags with XA included, we were unable to improve on the Akaike information criterion (AIC) by adding variables corresponding to the weighted up or the weighted down indicators of replies to the corresponding survey question, neither of which indicator was significant. The success of the balance statistic may be because, in respect of this survey question, the responses appear to be symmetric; a regression of XA on the ups and downs separately does not reject the null that the signs of the coefficients are equal and opposite ($p = 0.311$). For completeness we note here that the aggregated micro data series has some differences from the XA series due to it containing late replies that are not thought to be as accurate or informative as those that are timely.

(a) **Time series of the dependent variable, annual growth in gross fixed capital formation, y, and the two survey variables, the Nowcast, XA, and the Expected, XE.**

(b) **Time series of the dependent variable and both one-year ahead survey variables, Real investment in machinery, XI, and General business conditions, XC.**
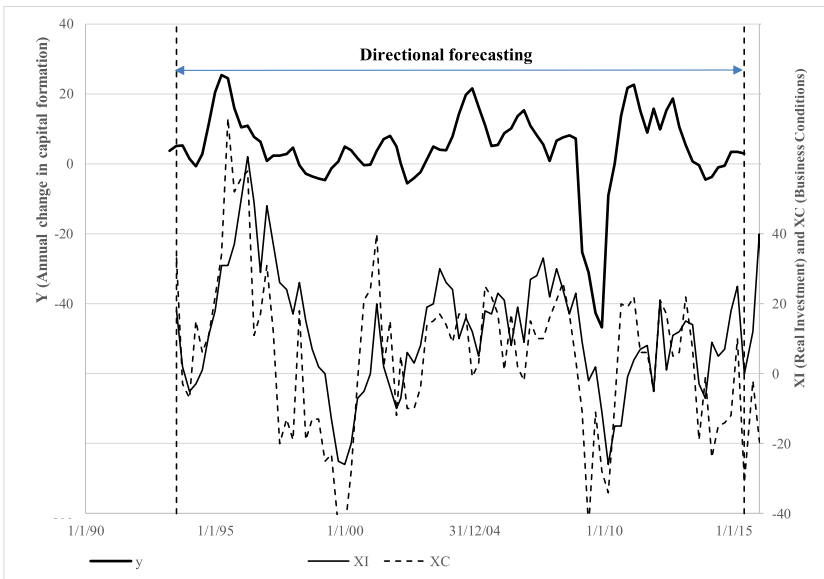

FIGURE 1 (a). time series of the dependent variable, annual growth in gross fixed capital formation, *y*, and the two survey variables, the Nowcast, XA, and the expected, XE. (b) Time series of the dependent variable and both 1-year-ahead survey variables, real investment in machinery, XI, and general business conditions, XC [Colour figure can be viewed at wileyonlinelibrary.com]
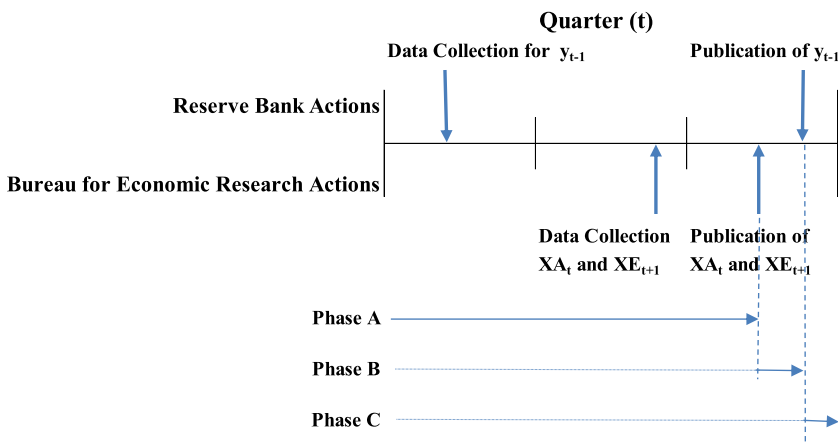
FIGURE 2 A timeline showing the times during each quarter when information is made available [Colour figure can be viewed at wileyonlinelibrary.com]

comparative results should probably be viewed as favoring the survey data if the forecasting performance is similar.

4.3 | Preliminary tests of the data

We performed initial tests for outliers on the two shorter term survey series, XA and XE, and the official data by visually inspecting the data before and after winsorization and checking for data points more than k IQRs outside the interquartile range (IQR). Only for the very narrow bands ($k = 1$) are there any data points that could count as outliers. However, these few observations are all clearly associated with the financial crisis post 2008 in the case of survey variables, and a few data points also for the Asian crisis of the late 1990s for the official series. We decided not to treat these observations as outliers, but to monitor their effect in some of our tests.

Stationarity tests for all variables were performed by selecting the appropriate lag order using the AIC and inspecting augmented Dickey–Fuller statistics with a constant (no trend was indicated). All the variables are stationary, which will allow combinations of these variables in the same model.

In Table 1, we give summary statistics of the five time series. The statistics for the shorter term survey series (XE and XA) are similar and the Pearson correlation between them is 0.69. Actual investment growth is on a different

scale and relatively more volatile than the survey series (using coefficients of variation). The two series considering a 1-year horizon, XI and XC, exhibit greater dispersion than those with a 1-quarter horizon, XE and XA, and have a slightly lower correlation: 0.65.

5 | ASSESSMENT OF THE SURVEY'S DIRECTIONAL FORECASTING ACCURACY

As indicated in Figure 1, for the analysis of directional forecasting accuracy we use the full time period for which survey data are available We assume throughout that the cost of errors is symmetric, and our measurement of forecasting accuracy reflects this; the wide range of policymakers using the survey forecasts makes it very difficult, if not impossible, to judge whether a false positive or a false negative forecast is more serious. We first consider the accuracy of the predictions of the Expected and Nowcasts over a short horizon. Secondly, we consider the two predictions with a 1-year horizon. Thirdly, we model the probability of a correct directional forecast, using the information available at the time of making the prediction.

## 5.1 | Accuracy of directional forecasts over a short horizon

The directional forecasts considered first are the shorter term survey-based predictions: Nowcast and Expected. In Figure 3, we show a scatter plot of these two survey variables against the actual growth, y_t. We can see from Figure 1 that actual growth is predominantly positive; this is reflected in Figure 3, where the majority of the points are on the right side of the plots. The top right quadrants for both XE and XA contain most points, indicating the quarters when the survey correctly forecast positive growth. The points in the bottom left quadrants show the correct survey forecasts of negative growth. Incorrect survey forecasts are indicated in the top left and bottom right quadrants.

In addition to the survey data, we use a time series model as a benchmark forecasting method. Our choice of an autoregressive model as the benchmark for judging forecasts is only one of many. Alternatives include vector autoregressions (VARs) or variants on univariate forecasting such as autoregressive integrated moving average, threshold, and switching models. Our decision to focus on the AIC selected AR(8) model follows first from previous results that univariate autoregression, evaluated using averaged RMSEs, is generally either best or second best of all other models where survey data are not

TABLE 1 Descriptive statistics of the two short horizon survey series (1992:Q3–2015:Q2, 92 observations), actual investment growth and the two longer term survey series (1993:Q3–2015 Q2, 88 observations)

	Nowcast (XA)	Expected (XE)	Actual (y)	M&E Inv. (XI)	Bus. Cond. (XC)
Minimum	−25.00	−26.00	−46.84	−26.00	−45.00
Maximum	35.00	48.00	25.33	62.00	73.00
Mean	10.68	11.91	4.12	11.57	4.08
SD	14.84	14.50	11.60	17.00	22.23
Skewness	−0.35	−0.28	−1.82	0.21	0.20
Coeff. Of variation	1.39	1.22	2.81	1.47	5.45
Correlations					
Nowcast (XA)	1	0.69	0.53	0.54	0.47
Expected (XE)		1	0.50	0.62	0.48
Actual (y)			1	0.29	0.52
M&E Inv. (XI)				1	0.65
Bus. Cond. (XC)					1
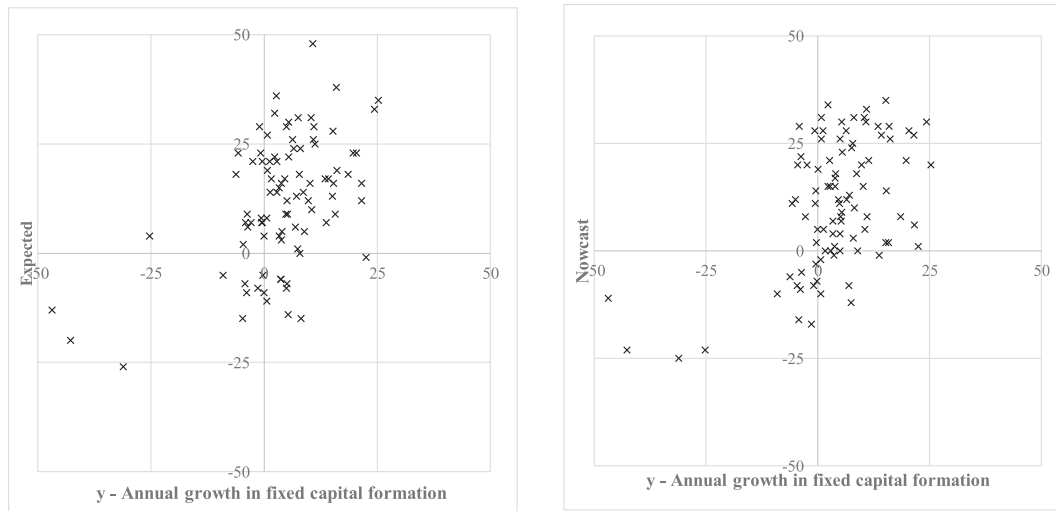
FIGURE 3 Scatter plots of the survey variables against the actual annual growth of fixed capital formation

included (Claveria, Pons, & Ramos, 2007); similar findings favoring univariate models are reported in Clar, Duque, and Moreno (2007). The time series model uses data from 1980:Q1. For this model, we considered recursively the minimum AICc autoregressive model using the data up to and including the forecast origin. (The AICc is a modified version of AIC appropriate for small samples; see Hurvich & Tsai, 1991.) For the first 76 cases this was AR(8); for the remainder it was AR(9). For simplicity, however, we used AR(8)[6] throughout; that is:

$$y_t = \mu + \sum_{k=1}^{8} \varphi_k y_{t-k} + \varepsilon_t.$$

The model is reestimated for each new quarter.

In Phases A and B (See Figure 2), y_{t-1} is not available, so we use the 2-quarter-ahead forecast that becomes available in Phase C of the previous quarter. In Phase C, we use the newly available 1-quarter-ahead forecast.

The directional accuracy, value, unbiasedness, and efficiency of the forecasts are evaluated using the tests described in Section 3.1. To ensure comparability, the accuracy of the forecasts is evaluated over the same range of data, as indicated in Figure 1a. The results are summarized for distinct phases within a quarter, indicated as A, B, and C in Table 2.

In Phase A—the majority of the quarter—Kuiper's score shows that the one-period-ahead ranking in terms

of decreasing accuracy is: AR(8) (2 quarters ahead); lagged Nowcast; Expected. The current Nowcast becomes available in Phase B with a higher Kuiper's score than the lagged Nowcast and Expected. In Phase C at the very end of the quarter, the AR(8) forecast becomes available with the highest Kuiper's score. In the contingency table-based tests (chi squared; Henrikkson & Merton; and Pesaran & Timmerman), the null hypothesis of independence between forecast and actual is rejected for each of the five forecasts for all three tests. Thus all five forecasts have value; this is reinforced by the Cumby and Modest test for value, where the hypothesis, $H_0^{\text{CM}}: \beta_2 = 0$, is clearly rejected in all cases. For the first Mincer and Zarnowitz test for unbiasedness, the hypothesis, $H_0^{\text{MZ2}}: \alpha_1 = 0$ (i.e., the means of the actual and the forecasts are not significantly different) is only rejected for AR(8) (2 quarters ahead). For the survey-based forecasts, the magnitude of the forecast is, on average, about 40% of the actual; this is because the survey balance is only expected to mirror the actual growth rate up to an arbitrary constant. Thus it is not surprising that for the second Mincer and Zarnowitz test for efficiency the only forecast for which the hypothesis, $H_0^{\text{MZ1}}: \beta_1 = 1$, cannot be rejected is AR(8) (1 quarter ahead). For Holden and Peel's efficiency test, the errors are correlated for all the forecasts. So, although the AR(8) (1-quarter-ahead) model is efficient in its usage of information, its errors are correlated. Note that this finding cannot be used to improve the forecasts in practice because of the delay in publication of y_t.

5.2 | Accuracy of directional forecasts over a 1-year horizon

Here we consider the two survey questions that predict the change over the next 12 months: XI considers "real

---

[6]Given the long data period for the model, we investigated the stability of y_t using the sequential Bai–Perron approach tests for $L + 1$ versus L breaks; $L \leq 5$ (Bai & Perron, 2003). The global Bai–Perron test first performs sequential tests with a given maximum (1–5) and then combines the results to give weighted and unweighted test scores. Neither of these two sets of test results shows evidence of any break over the entire period from 1980 to 2015 using a 5% significance level and a trimming percentage of 15%.

TABLE 2 Summary of directional accuracy tests on the survey-based forecasts and a time series model. Using data from September 30, 1992, to June 1, 2015 (92 observations)

Forecast Kuiper's score		Phase A			Phase B	Phase C
		Expected	Lagged nowcast	AR(8) horizon, 2 quarters	Nowcast	AR(8) horizon, 1 quarter
		0.24	**0.33**	**0.57**	**0.50**	**0.74**
Chi squared test	Statistic	5.18	9.25	22.44	22.73	41.21
	p-value	0.023	0.002	0.000	0.000	0.000
H&M/Fisher's exact test	p-value	0.014	0.002	0.000	0.000	0.000
Pesaran and Timmermann's test	Statistic	6.71	11.20	25.10	25.84	45.01
	p-value	0.010	0.001	0.000	0.000	0.000
Mincer and Zarnowitz: Efficiency and unbiasedness	Beta	0.40	0.38	0.75	0.41	0.98
	Upper 95% limit	0.55	0.53	0.94	0.55	1.10
	Alpha	−0.65	−0.01	2.59	−0.29	1.05
	p-value	0.634	0.994	0.009	0.820	0.102
Cumby and Modest: Value	Beta	11.38	25.92	10.72	14.79	14.27
	p-value	0.001	0.000	0.000	0.000	0.000
Holden and Peel: Efficiency	Beta	0.55	0.59	0.61	0.58	0.27
	p-value	0.000	0.000	0.000	0.000	0.008

investment in machinery and equipment" and XC considers "general business conditions" (shown in Figure 1b). Here, we consider only Phase A and thus the appropriate time series benchmark is the 6-quarter ahead (4-quarter horizon plus 2-quarter publication delay) forecasts from the AR(8) model. The results of the analysis are shown in Table 3. Kuiper's score ranks Expected Business Conditions, XC, as most accurate and Expected Real Investment, XI, as least accurate. The time series forecast, AR(8), is more accurate than XI. Applying a 5% significance level to the contingency based tests, the value of XC as a directional forecast has significant value

according to all three tests; the value of XI is not significant at this level for any test; the value of AR(8) is significant for Fisher's exact test and for Pesaran and Timmerman's test. The Mincer and Zarnowitz test finds all the forecasts inefficient and all but XI significantly biased upwards. The Cumby and Modest test shows significant value for XI and XC but not for AR(8). All the forecasts fail the Holden and Peel efficiency test; that is, the forecasts are inefficient as the errors are correlated.

In summary, as a directional forecast Expected Business Conditions, XC, performed convincingly well in the contingency table-based tests and has significant value

TABLE 3 Summary of directional accuracy tests on the 1-year-ahead survey-based forecasts and a time series model. Using data from September 30, 1992, to June 1, 2015 (89 observations)

Forecast Kuiper's score		Real investment (XI) 0.18	Business conditions (XC) 0.58	AR(8) horizon, 6 quarters 0.38
Chi squared test	Statistic	2.26	21.16	8.33
	p-value	0.132	0.000	0.004
H&M/Fisher's exact test	p-value	0.069	0.000	0.002
Pesaran and Timmermann's test	Statistic	3.27	23.82	9.93
	p-value	0.070	0.000	0.002
Mincer and Zarnowitz: Efficiency and unbiasedness	Beta	0.20	0.28	0.18
	Upper 95% limit	0.34	0.37	0.48
	Alpha	2.05	2.92	4.29
	p-value	0.158	0.009	0.001
Cumby and Modest: Value	Beta	13.46	21.28	3.51
	p-value	0.001	0.000	0.082
Holden and Peel: Efficiency	Beta	0.74	0.48	0.82
	p-value	0.000	0.000	0.000

but is inefficient and biased. In contrast, the prediction of Real Investment, XI, performed poorly in most of the tests; this result may be due to the likely asynchronicity of machinery and equipment investment (M&E) with gross fixed capital investment, which includes construction.

5.3 | Sensitivity of directional forecast accuracy to the effects of the financial crisis

Here, we investigate whether the directional accuracy of the survey-based methods was unduly influenced by the effects of shocks to the economy. The period over which the forecasts are evaluated is divided into three parts: "pre-crisis" from 1992:Q2 to 2008:Q4, "crisis" from 2009:Q1 to 2011:Q4, and "post-crisis" from 2012:Q1 to 2015:Q2. We calculated the Pesaran–Timmerman statistic for each of these subperiods for the shorter term variables: the Nowcast, XA, the Expected, XE, and the AR(1) forecast; and for the longer term variables: real investment, XI, and business conditions, XC. The results are shown in Table 4.

The accuracy of the Nowcast, XA, as measured by the Pesaran and Timmerman statistic, persists during the crisis but deteriorates post-crisis. Although the results for Expected, XE, are not as strong, the pattern is similar with a post-crisis deterioration. The AR(1) forecast is included for comparison and shows no effect due to the crisis. For the longer term survey forecasts, the accuracy of real investment, XI, is not significant in any subperiod; the accuracy of business conditions, XC, suffers a slight drop in accuracy during the crisis.

TABLE 4 Pesaran and Timmerman statistics for the survey-based variables calculated before, during, and after the financial crisis

Variable		Pre-crisis	Crisis	Post-crisis	Overall
XA	Statistic	18.41	9.35	0.22	25.84
	p-value	0.00	0.00	0.64	0.00
XE	Statistic	4.71	3.37	0.64	6.71
	p-value	0.03	0.07	0.42	0.01
XI	Statistic	0.40	1.67	2.78	3.27
	p-value	0.53	0.20	0.10	0.07
XC	Statistic	14.25	3.57	4.76	23.82
	p-value	0.00	0.06	0.03	0.00
AR(1)	Statistic	31.07	5.65	8.38	45.01
	p-value	0.00	0.02	0.00	0.00

Thus we see that the most noticeable effect of the crisis on directional accuracy was to destabilize the shorter term survey variables after the event.

5.4 | Estimating the probability of a directionally correct prediction

Here we attempt to calibrate the probability of a correct directional prediction in terms of the information available at the time the prediction is made; this will allow us to associate an objective measure of confidence with the prediction. We use logit regression to model the success of the four survey predictions. The dependent variables are binary (0, 1), where 1 indicates a *correct* directional prediction for the official series y_t. To define the problem, we identify the information set available to the survey panel when they complete their response (on the assumption that the latest four known values of y_t are most relevant).

For the Nowcast:

$$PA_t = \text{Prob}(XA_t \text{ directionally correct}|XA_{t-1}, XE_t, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}).$$

For the 1-quarter-ahead and each of the 4-quarter-ahead forecasts:

$$PE_t = \text{Prob}(XE_t \text{ directionally correct}|XA_{t-2}, XE_{t-1}, y_{t-3}, y_{t-4}, y_{t-5}, y_{t-6}),$$

$$PI_t = \text{Prob}(XI_t \text{ directionally correct}|XA_{t-5}, XE_{t-4}, y_{t-6}, y_{t-7}, y_{t-8}, y_{t-9}),$$

$$PC_t = \text{Prob}(XC_t \text{ directionally correct}|XA_{t-5}, XE_{t-4}, y_{t-6}, y_{t-7}, y_{t-8}, y_{t-9}).$$

In general, if the horizon of the forecast is H quarters, then

$$P?_t = \text{Prob}(H \text{ quarter} - \text{ahead} - \text{forecast directionally correct}|XA_{t-1-H}, XE_{t-H}, y_{t-2-H}, y_{t-3-H}, y_{t-4-H}, y_{t-5-H}).$$

We consider different groups of possible explanatory variables that may affect the success of the survey predictions. One focus is on stability. First, we consider the subjectively assessed acceleration in growth by the forecaster represented by $Z_1 = |XE_{t-H} - XA_{t-1-H}|$. This represents the absolute difference between the Nowcast and 1 quarter ahead published in the latest available set of results at $(t - 1 - H)$. Second, we attempt to capture the objective degree of instability of the underlying

environment that may influence the forecast, with stability and continuity in the officially recorded investment series favoring accuracy. Our measure of stability is the magnitude of any sign switch from or towards the normal pattern of positive annual growth. If

$$\delta_t = 1 \quad \text{when } y_t \geq 0,$$
$$\delta_t = 0 \quad \text{when } y_t < 0,$$

then

$$Z_2 = |\delta_{t-2-H} - \delta_{t-3-H}|.|y_{t-2-H}|.$$

This variable would be expected to worsen accuracy because large sign shifts in any direction make forecasting difficult.

Our next two sets of logit regressors capture information in the lag structure of the official data. Given the competitiveness of the AR(8) forecasting model, it is reasonable to hypothesize that there may be some information in the lag structure that signifies ease of forecasting. We therefore include the magnitude of the latest known values of four lagged values, $M_k = |y_{t-1-H-k}|$ for $k = 1, ..., 4$. Since positive directional change is more frequent, forecasters may be able to intuit signals from a *positive* sign on the latest known values of lagged values of y_t, $D_k = |\delta_{t-1-H-k}|$ *for* $k = 1, ..., 4$. Finally, we include seasonal effects (S_1, S_2, S_3) for completeness.

The logit regression was performed for each of the four survey predictions, initially including all variables. In view of our focus on the effects of instability on forecast accuracy, the variables, Z_1 and Z_2, were included in all cases; other lagged variables or seasonal dummies were dropped if their p-values exceeded 20%. The results are shown in Table 5.

The percentage of deviance explained can be interpreted similarly to R^2 in linear regression. We see that explanatory variables only make a worthwhile contribution to explaining the probability of directional accuracy for the Nowcast (corresponding to PA) and General Business Conditions (corresponding to PC). For the year-ahead narrow investment category (corresponding to PI), no variable achieves 5% significance and we do not consider it further in this section.

Focusing on the effects of instability, we find that Z_1, the subjective measure of instability at the time of making the response, is significant for both PA and PC. As is intuitively reasonable, the probability of the Nowcast being directionally accurate, PA_t, decreases with Z_1. Counterintuitively, Z_1 has the opposite effect on PC; this apparent anomaly may indicate that the effect of the instability dies out during the longer horizon. Variable Z_2, representing environmental instability, does not have a significant effect on any of the prediction probabilities. Further, no significant seasonal effects were found.

Focusing on the regressors for the lag structure, M_3 has a significantly negative coefficient for PA_t, PE_t and PC_t; this is the magnitude of the official series 4 quarters before the survey response is made—the greater this magnitude, the lower the probability of a correct prediction. This effect is compensated by significant positive coefficients on the magnitude of the official values a quarter before (for PA_t) or a quarter after (for PC_t). The absolute values of the successive coefficients are similar, so that a consistent magnitude of the official growth series has a neutral effect on the probability of a directionally correct prediction. An increase in magnitude between 4 quarters before and 3 quarters before the response, $|y_{t-1-H-2}| > |y_{t-1-H-3}|$, will tend to increase the probability of a correct directional prediction for PC_t. In contrast, an increase in magnitude from 5 quarters before and 4 quarters

TABLE 5 Results of the logistic regression on the probability of a directionally correct survey prediction

Variable	PA		PE		PI		PC	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Intercept	1.60	0.08	0.64	0.19	−0.03	0.96	−1.01	0.15
Z_1	−0.27	0.00	−0.03	0.68	0.00	0.99	0.24	0.03
Z_2	2.41	0.17	0.21	0.39	−0.30	0.11	0.22	0.33
M_1	0.40	0.00			0.06	0.08		
M_2			0.15	0.02			0.21	0.00
M_3	−0.32	0.01	−0.19	0.03			−0.20	0.00
M_4	0.25	0.03	0.12	0.07				
D_1	2.06	0.07			0.73	0.19	1.82	0.01
D_3	−1.92	0.14						
% deviation explained		37.19		2.24		5.26		20.27

before the response, $|y_{t-1-H-3}| > |y_{t-1-H-4}|$, will tend to decrease the probability of a correct directional prediction for PA_t. It may also be noted that a positive latest known official value, D_1, has a positive effect for the cases corresponding to PA and PC.

Focusing on the Nowcast and Business Conditions where the logit regression has most explanatory power, we show how the out-of-sample estimates of PA and PC give an objective estimate of the probability of a directionally accurate forecast and can thus constitute a measure of confidence. The coefficients identified in Table 5 are reestimated quarter by quarter using contemporaneous data to predict PA and PC.

The predictions of PA and PC are shown alongside the predictions, XA_t and XC_t, and the actual growth, y_t, in Table 6. Note that there is not an exact correspondence between low values of PA and PC and incorrect directional predictions—that would be too good to be true. However, the values of PA and PC do give the decision maker an indication of how much faith to place in the Nowcast and the prediction of Business Conditions in a year's time. For example, the Nowcast of positive growth on March 1, 2014, or September 1, 2014, has a low estimated probability of being correct, suggesting caution.

6 | ACCURACY OF POINT FORECASTS

Here, we first present how the survey data are used to make point forecasts and we identify the time series models to be used as comparators. We then examine the forecasting accuracy of these model over horizons up to 8 quarters. Secondly, we investigate whether the

occurrence of the financial crisis during our sample period influences our results. Thirdly, we make pairwise comparisons of the short-term forecasting accuracy of the models considered to evaluate the significance of differences in accuracy. Fourthly, we mirror our analysis of directional accuracy by modeling the magnitude of point forecast errors using data available at the time of prediction.

6.1 | Using survey data for point forecasts, a time series benchmark, and some hybrid models

As shown in Figure 1a, for point forecasting we use some of the data for estimation only, the remainder being used for measuring forecasting accuracy. In addition to models using survey data, we consider time series data and two hybrid models using data from both sources. The survey-based forecasts are adjusted for scale by regressing the actual values, y_t, on the survey variables:

$$y_t = \mu_E + \eta XE_t + \varepsilon_t \tag{1}$$

and

$$y_t = \mu_A + \gamma XA_t + \varepsilon_t. \tag{2}$$

The time series model

$$y_t = \mu_3 + \sum_{k=1}^{8} \varphi_k y_{t-k} + \varepsilon_t \tag{3}$$

is used as a benchmark. The hybrid models are extensions of the time series model in Equation 3 to include each of the survey-based forecasts: the Expected and the Nowcast:

TABLE 6 Conditional probabilities of directionally correct predictions, PA and PC, using only data available at the time of prediction

Date	y	XA	XA directionally correct (1, 0)	PA	XC	XC directionally correct (1, 0)	PC
Sep 1, 2012	18.62	8	1	0.997	5	1	0.965
Dec 1, 2012	10.58	5	1	0.998	6	1	0.986
Mar 1, 2013	5.34	9	1	0.979	22	1	0.947
Jun 1, 2013	0.69	−2	0	0.982	7	1	0.675
Sep 1, 2013	−0.38	14	0	0.724	−19	1	0.973
Dec 1, 2013	−4.51	20	0	0.832	−1	1	0.510
Mar 1, 2014	−3.72	22	0	0.324	−24	1	1.000
Jun 1, 2014	−0.95	−8	1	0.749	−15	1	0.845
Sep 1, 2014	−0.56	11	0	0.102	−14	1	0.926
Dec 1, 2014	3.38	7	1	0.375	−12	0	0.987
Mar 1, 2015	3.44	4	1	0.814	10	1	0.758
Jun 1, 2015	2.91	0	1	0.897	−31	0	0.745

$$y_t = \mu_6 + \sum_{k=1}^{8} \varphi_{3k} y_{t-k} + \lambda_3 XE_t + \varepsilon_t, \quad (4)$$

$$y_t = \mu_4 + \sum_{k=1}^{8} \varphi_{1k} y_{t-k} + \lambda_1 XA_t + \varepsilon_t. \quad (5)$$

To ensure comparability of results, all the models (Equations 1–5) are estimated using data from September 1992 up to March 2003. The models were then reestimated every quarter and used to compute 1- to 8-quarter-ahead forecasts. The accuracy of these forecasts is evaluated for data from June 2003 to June 2015. The forecasts for all models are computed recursively. The lagged dependent variable in an h-step-ahead forecast, y_{t+h-k}, is replaced by $\hat{y}_{t+h-k} = E(y_{t+h-k}|\Theta_t)$ when $h > k$. For the explanatory variables, univariate time series forecasts were prepared; both Nowcasts and Expected were best represented (minimum AICc) by AR(1) models. In addition to the two hybrid models, we also consider two combined forecasts using both data sources; for phases A and B, we compute equally weighted combinations of forecasts using the available time series forecasts and the available survey forecasts.

RMSE was used to measure the accuracy of each model for each horizon; the results are shown in Table 7. In addition, for a 1-quarter-ahead horizon, the absolute error of each forecasting method was ranked for each phase of the forecasting quarter and the average rank is shown in the right-hand column of the table (a low value is preferable). In our analysis, we consider the latest forecasts available from each model in each phase, A, B or C. For the time series-based models, this means

using two or more quarter-ahead forecasts in Phases A and B. For example, in Phase A, the current Nowcast is not available and the one-period lag is used. In phases A and B, y_{t-1} is not yet available and the second lag is used. We denote a forecast prepared in the current quarter using the model described in Equation M by (M); a 2-quarter-ahead forecast prepared in the previous quarter is denoted by $(M)^*$.

Considering Phase A, we see that the Expected is more accurate than the lagged Nowcast for horizons of 1–7 quarters. For horizons of 1 and 2 quarters, the time series model $(3)^*$ is more accurate than the survey-based forecasts, even though the latest available observation is y_{t-2}. However, the use of both data sources, either via the hybrid regression model $(5)^*$ or via the combined forecasts from (1) and $(3)^*$, leads to a noticeable improvement in accuracy over either single data source. $(5)^*$ has the lowest RMSE for horizons up to 5 quarters.

In Phase B, the Nowcast (2) becomes available and these up-to-date forecasts are more accurate than the Expected for horizons of 1, 2, and 3 quarters; over later horizons the differences in accuracy are marginal. To use both data sources we combine the Nowcast with the AR(8) time series model. This combination $((2) + (3)^*)$ leads to a forecast that is more accurate for 1 and 2 quarters ahead than the Phase A combination $((1) + (3)^*)$. However, model $(5)^*$, from Phase A, remains the most accurate for horizons of 1–5 quarters. Note that $(5)^*$ using data up to $\{y_{t-2}, XA_{t-1}\}$ is more accurate (RMSE) than the simple combination $((2) + (3)^*)$, using data up to $\{y_{t-2}, XA_t\}$.

TABLE 7 Measures of the accuracy of forecasts using survey models, time series models and models using both information sources

	Forecast models	Horizon in quarters								Quarter 1 average ranking over observations
		1	2	3	4	5	6	7	8	
Phase A (1)	Expected (via LR)	13.21	14.16	14.97	15.33	15.50	15.64	15.59	15.49	7.00
(2)*	Lagged Nowcast	13.90	14.78	15.38	15.64	15.69	15.69	15.59	14.91	7.06
(3)*	AR(8) ≥ 2 quarters ahead	11.80	14.04	18.19	17.62	16.83	16.48	17.35	16.91	5.96
(5)*	AR(8) + Nowcast ≥2 quarters ahead	10.03	11.55	14.28	15.17	15.38	15.68	15.72	16.19	5.38
(1) + (3)*	Combination: Expected and ≥ 2-quarter-ahead AR forecast	10.70	12.67	15.56	15.59	15.50	15.45	15.57	15.40	5.06
Phase B (2)	Nowcast (via regression)	12.65	13.90	14.78	15.38	15.64	15.69	15.69	15.59	6.52
(2) + (3)*	Combination: Nowcast and 2-quarter-ahead AR forecast	10.35	12.50	15.39	15.60	15.54	15.47	15.66	15.45	5.33
Phase C (3)	AR(8)	8.53	11.80	14.04	18.19	17.62	16.83	16.48	17.35	4.33
(4)	AR(8) + Expected	7.47	10.20	12.08	14.71	15.22	15.48	15.45	15.60	4.38
(5)	AR(8) + Nowcast	7.51	10.03	11.55	14.28	15.17	15.38	15.68	15.72	3.98

Note. The RMSE of 1–8 quarters ahead is given. In the right-hand column, for a 1-quarter-ahead horizon, the absolute error of each forecasting method was ranked for each phase of the forecasting quarter, and the average rank is shown. The forecasts are shown according to the phase within the quarter that they become available. Accuracy is evaluated using data from June 2003 to June 2015 (49 observations for 1-quarter-ahead forecasts).

*Denotes the use of two or more quarter-ahead forecasts with an origin within the previous quarter.

In Phase C—the short interval at the end of the quarter when y_{t-1} becomes available—we see that the regression models considering both time series and survey information, (4) and (5), are more accurate over all horizons than the pure survey-based or pure time series models. Although Phase C is very short, the greater than 1-quarter-ahead forecasts from these regression models, denoted by (3)* and (5)*, prove to be dominant in Phases A and B of the next quarter.

6.2 | Investigating the effects of the financial crisis

In Section 5.2, we investigated the effect of the crisis on directional accuracy; here we consider point forecasting accuracy. Granger (1996) suggests that a reasonable strategy during a structural break is to use an adaptive approach such as an autoregressive moving-average model until a structural model can be recalibrated. In our case, we do not have a structural model, but it is of interest to discover whether the survey panel reacts more quickly to breaks than the AR model. The results in Table 8 summarize 12 years including the 2008 financial crisis. To investigate whether the comparison of forecasting methods was unduly influenced by these shocks to the economy, we looked at the behavior of the forecast errors over time.

In Figure 4, we plot the minimum and maximum (over the 10 forecasting methods considered) absolute error of the 1-quarter-ahead forecast for each quarter. Observing the increase in error magnitude in the middle of the

interval used for forecast evaluation, as in Section 5.3, we divided this period into three subperiods: pre-crisis from 2003:Q2 to 2008:Q4; crisis from 2009:Q1 to 2011:Q4; post-crisis from 2012:Q1 to 2015:Q2. We use the breakpoint of 2009:Q1 because the investment response of the financial crisis is not apparent in the data until then. These subperiods are shown in Figure 4. We calculated the RMSE for each method in each subperiod and overall. We show these results in Table 8; in addition, we show the ranking of each method (1 for the most accurate).

For each method, the RMSEs for the pre-crisis and post-crisis subperiods are broadly similar, whereas those for the crisis subperiod are much larger. However, the ranking of each method differs very little in each subperiod from the overall ranking (which corresponds to the 1-quarter horizon column in Table 7). The forecasts using both data sources predominate in each subperiod. The ranking of the survey-based forecasts remains the same during the crisis and thus there is no evidence that the survey panel anticipated the effects of the financial crisis better than the time series model.

6.3 | Pairwise comparisons of accuracy over a short horizon

Having established that the relative accuracy of the forecasting methods is very similar across the forecast subperiods; we now make pairwise comparisons between the methods. First, we used the Friedman test, as described in Section 3.2, and we found that there was a significant

TABLE 8 The RMSEs of 1-quarter-ahead forecasts calculated for each method for each subperiod and overall. The ranking in terms of ascending RMSE is also given

			Pre-crisis		Crisis		Post-crisis		Overall		Overall rank within phases A, B, C
		Forecast models	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank	
Phase A	(1)	Expected (via LR)	6.84	10	23.93	9	6.53	8	13.21	9	4
	(2)*	Lagged Nowcast	6.57	9	25.20	10	7.11	10	13.90	10	5
	(3)*	AR(8) ≥ 2 quarters ahead	4.73	5	22.26	7	4.20	6	11.80	7	3
	(5)*	AR(8) + Nowcast ≥2 quarters ahead	4.46	4	18.32	4	5.07	7	10.03	4	1
	(1) + (3)*	Combination: Expected and ≥ 2-quarter-ahead AR forecast	5.14	7	19.91	6	3.36	4	10.70	6	2
Phase B	(2)	Nowcast (via LR)	5.98	8	22.97	8	7.02	9	12.65	8	2
	(2) + (3)*	Combination: Nowcast and 2-quarter-ahead AR forecast	4.74	6	19.20	5	4.01	5	10.35	5	1
Phase C	(3)	AR(8)	2.76	2	16.52	3	2.83	1	8.53	3	3
	(4)	AR(8) + Expected	2.79	3	14.24	1	2.96	2	7.47	1	1
	(5)	AR(8) + Nowcast	2.48	1	14.36	2	3.23	3	7.51	2	2

*Denotes the use of two or more quarter-ahead forecasts with an origin within the previous quarter.
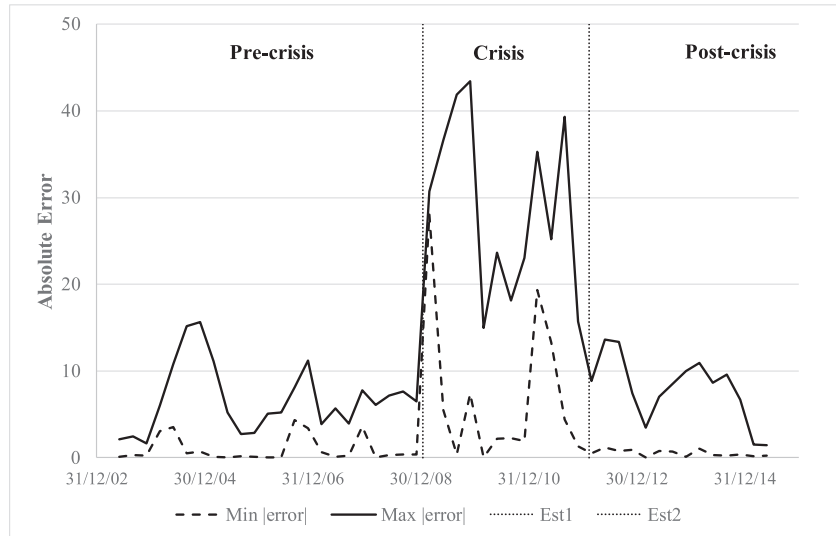
FIGURE 4 A time series of the minimum and maximum absolute errors of the 10 forecasting methods considered in Table 5

difference between at least some methods. In the second stage of this test we computed *p*-values showing the significance of the difference between pairs of methods. Subsequently, we found that these values were closely mirrored by the Diebold–Mariano test, and so for brevity we will confine ourselves to discussing the results of this test.

We begin by comparing the value of forecasts with the modified Diebold–Mariano test using the difference of squared errors from methods X and Y; that is, $e_X^2 - e_Y^2 = (|e_X| - |e_Y|)(|e_X| + |e_Y|)$.

However, we found that these measures suffered from very high kurtosis (around 10), rendering the implicit assumption of normality (kurtosis equal to zero) in the test invalid. This is mainly due to large errors during the crisis. Thus we used the difference of absolute errors $(|e_X| - |e_Y|)$ as a more credible alternative, as the value is only concerned with the difference in error size and not affected by the magnitude of the errors $(|e_X| + |e_Y|)$. In this case, the kurtosis is -0.3, far closer to zero. The results are shown in Table 9.

For the methods available in Phase A, we see that the two methods using both time series and survey methods, (5)* and ((1) + (3)*), are significantly more accurate than the survey forecasts alone. In Phase B, the combination of Nowcast and time series ((2) + (3)*) is significantly more accurate than Nowcast alone. Note also that the Nowcast is not significantly more accurate than the Expected. For both phases A and B we find that forecasts using survey data with time series data are significantly more accurate than the latest available survey data. In particular, (5)*, using the Nowcast and time series data, is more accurate than the time series only model, (3)*, with a *p*-value of 8%. In Phase C, there is no significant difference between the accuracy of the methods. However, these methods—(3), (4), and (5)—are significantly more accurate than

the pure survey or pure time series methods available in Phases A and B.

In panel 2, we compare over longer horizons the methods that use both data sources with those using survey data only or time series only; the *p*-values for forecasts over 8 quarters ahead are shown. The first four comparisons consider the time series/survey combination with a pure survey forecast. We see that the greater accuracy of the combined methods is confined to a 2-quarter horizon. Comparing combined forecasts with time series we see that the greater accuracy of the combined forecasts becomes more significant around the 1-year horizon (rows 5 and 6 of panel 2). This pattern is also visible in the RMSE statistics of Table 7.

In summary, if accuracy of point forecasts over a 1-quarter horizon is the primary objective, then for all three phases the best strategy is to enhance the latest available survey data with the latest available AR(8) time series model. This enhancement is achieved in Phases A and B either by combining forecasts or by using model (5)* estimated in the previous quarter; this last enhancement from model (5)* is the most significant according to the Diebold–Mariano test. In Phase C the survey data can be used with the time series models (4) and (5). If the objective is accuracy over longer horizons of 3 quarters or more, the advantage of combining time series data with the survey-based data tends to decrease in favor of the survey-only forecasts, given that the longer term time series forecasts are relatively less accurate.

6.4 | Predicting the magnitude of a survey-based point forecast error

In an analogous fashion to our calibration of the probability of a correct directional forecast, we investigate whether

TABLE 9 One-sided p-values for the Diebold–Mariano test, modified by Harvey et al. (1997)

			Phase A					Phase B		Phase C		
Panel 1	**Forecast models**		(1)	(2)*	(3)*	(5)*	(1) + (3)*	(2)	(2) + (3)*	(3)	(4)	(5)
Phase A	(1)	Expected (via LR)		0.15	0.11	0.01	0.00	0.19	0.00	0.00	0.00	0.00
	(2)*	Lagged Nowcast			0.09	0.01	0.00	0.06	0.00	0.00	0.00	0.00
	(3)*	AR(8) ≥ 2 quarters ahead				0.08	0.27	0.15	0.23	0.01	0.01	0.01
	(5)*	AR(8) + Nowcast ≥2 quarters ahead					0.31	0.02	0.35	0.04	0.02	0.02
	(1) + (3)*	Combination: Expected and ≥ 2-quarter-ahead AR forecast						0.01	0.34	0.03	0.01	0.01
Phase B	(2)	Nowcast (via LR)							0.01	0.00	0.00	0.00
	(2) + (3)*	Combination: Nowcast and 2-quarter-ahead AR forecast								0.02	0.01	0.00
Phase C	(3)	AR(8)									0.21	0.29
	(4)	AR(8) + Expected										0.36
	(5)	AR(8) + Nowcast										

Panel 2		**Horizon**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
(1) + (3)*	Better than	(1)	0.00	0.03	0.28	0.21	0.23	0.22	0.44	NA
(5)*	Better than	(1)	0.01	0.05	0.40	0.63	0.58	0.65	NA	NA
(2) + (3)*	Better than	(2)	0.01	0.02	0.19	0.20	0.20	0.22	0.44	NA
(5)*	Better than	(2)	0.02	0.04	0.33	0.62	0.52	0.59	0.90	NA
(1) + (3)*	Better than	(3)*	0.27	0.18	0.06	0.06	0.09	0.11	0.07	0.08
(2) + (3)*	Better than	(3)*	0.23	0.17	0.06	0.06	0.09	0.09	0.05	0.07
(5)*	Better than	(3)*	0.08	0.17	0.14	0.09	0.23	0.41	0.13	0.12

Note. Small p-values indicate that method X is more accurate than method Y, where method X corresponds to a lower RMSE (horizon 1 quarter) than method Y in Table 8.

Panel 1 shows a comparison of the accuracy of 1-quarter-ahead forecasts across all methods considered.

Panel 2 shows comparison of two methods, each from the same phase, showing the p-values for different horizons.

*Denotes the use of two or more quarter-ahead forecasts with an origin within the previous quarter.

the magnitude of a survey-based error can be predicted using the information available at the time the point forecast is made. We focus on the errors from the Expected (1) and the Nowcast (2). We use linear regression to model the magnitude of these two survey-based point forecasts. The explanatory variables are the same as those used in Section 5.4, with the addition of Z_2 lagged by 1 quarter. As in Section 5.4, the least significant variables were removed from the model until all p-values were less than 20%.

Using the adjusted R^2 values as a measure of the variation in the magnitude of the forecast errors explained by the two regressions, we see that those of Nowcast are far better explained by these variables than for Expected. The first measure of instability—Z_1, the last known difference between Expected and Nowcast—was significant in explaining the probability of a correct directional Nowcast. However, this variable did not significantly contribute to explaining error magnitude in either case and is omitted from Table 10. The second measure of instability—the effect of a transition to or from positive growth, Z_2—significantly increases error magnitude for both Nowcast and Expected. For Nowcast, this effect persists over two

TABLE 10 Results of the linear regressions on forecast error magnitude

	Nowcast		Expected	
Variable	**Estimate**	**p-value**	**Estimate**	**p-value**
Intercept	0.94	0.74	5.33	0.06
Z_2	3.07	0.00	2.66	0.00
$Z_{2(t-1)}$	2.98	0.00		
M_2	0.13	0.19		
D_1	−5.39	0.05	−6.82	0.16
D_2			10.16	0.04
D_3	10.58	0.00		
Adjusted R^2		0.52		0.18

quarters. Although the magnitude of the actual growth figures has no significant impact on error magnitude, the direction of growth does have an effect. If D_1—the most recent known growth—is positive then error magnitude tends to be reduced; however, this reduction is counteracted if previous quarters also showed positive growth.

7 | SUMMARY AND CONCLUSIONS

This paper looks in detail at responses to four questions from the South African Bureau of Research manufacturing survey, containing information on future capital investment. We carefully map the timing of the release of both official and survey data so as to mirror real-time forecasting by using only those data available contemporaneously. This approach requires that we perform accuracy comparisons for each of three phases within the forecasting quarter to ensure that the forecasts compared were compiled using the same information set.

Our analyses of both the directional accuracy and point-forecast accuracy of the survey-based forecasts follow the same pattern. The accuracy is investigated over horizons ranging from 1 quarter to a year or more; the sensitivity of forecasting accuracy to the effects of the financial crisis is investigated; and the effect of the economic environment at the survey completion date is investigated in respect of forecast accuracy.

Regarding the directional accuracy of predictions of official investment data, we confirmed the value of the survey-based forecasts up to 1 quarter ahead, for all three phases within the forecasting quarter. However, we show that the chosen time series benchmark, an AR(8) model, tends to have greater directional accuracy, measured by the Kuiper score, than the survey predictions. For a 1-year horizon, we show that the survey question on business conditions has value and is directionally more accurate than the time series benchmark. In regard to the effect of the financial crisis, the most noticeable effect on directional accuracy was to destabilize the shorter term survey variables in the post-crisis period. We show that the probability of directionally accurate forecasts from the survey Nowcast, and from the 1-year-ahead prediction of business conditions, can be usefully modeled by a small set of variables including the state and stability of current investment.

In relation to the point forecasts, we considered the use of scaled survey forecasts, time series forecasts, and forecasts using information from both survey and official data, where the survey series were either augmented by a time series model, or the survey forecasts were combined with a time series forecast. Having checked that our findings were consistent across the range of our data which included the effects of 2008 financial crisis, we found that survey forecasts enhanced by time series data were significantly more accurate than the pure survey forecasts, but only for shorter horizons of 1 or 2 quarters. The latter finding is consistent across the three subperiods that encompass the financial crisis. For longer horizons of 3 or more quarters, the benefit of enhancing survey-based forecasts with time series data disappeared. This finding is consistent with the significantly greater directional accuracy of the 1-year-ahead forecast of the business conditions survey variable compared to a time series forecast.

We investigated also whether it was possible to predict the probability of an accurate forecast for both the directional and point forecast models. We constructed two measures of stability that might affect the survey correspondent's predictive accuracy: Z_1, their perception of growth at the time of prediction, |Expected – Nowcast|; and Z_2, the occurrence and magnitude of a switch into, or out of, positive growth. We found that while Z_1 had a significant influence on the probability of a correct directional forecast for the Nowcast and the prediction of Business Conditions, Z_2 had a significant effect on the magnitude of the Nowcast error. Both these probability models are potentially useful to forecasters. Supplementing the survey-based directional forecast with a probability of its directional accuracy is a valuable addition to investment decision making. The main usefulness of the predicted magnitude of the survey-based point forecast error is as a warning of changes in the uncertainty of the decision-making process.

ORCID

Ciaran Driver https://orcid.org/0000-0002-9297-806X
Nigel Meade https://orcid.org/0000-0001-6689-5052

REFERENCES

Aron, J., & Muellbauer, J. (2002). Interest rates effect on output: Evidence from a GDP forecasting model for South Africa. *IMF Staff Papers*, *49*(Special issue), 185–213.

Arora, V., & Bhundia, A. (2003). *Potential output and total factor productivity growth in post-Apartheid South Africa* (Working Paper WP/03/178). Washington, DC: International Monetary Fund.

Bai, P., & Perron, P. (2003). Critical values for multiple structural change tests. *The Econometrics Journal*, *6*(1), 72–78.

Biau, O., Erkel-Rousse, H., & Ferrari, N. (2006). Individual responses to business tendency surveys and the forecasting of manufactured production: An assessment of the Mitchell, Smith and Weale dis-aggregate indicators on French data. *Economie et Statistique*, December, *2006*, No., 395–396.

Bond, S., Leblebicioğlu, A., & Schiantarelli, F. (2010). Capital accumulation and growth: A new look at the empirical evidence. *Journal of Applied Econometrics*, *25*(7), 1073–1099.

Bond, S., Soderbom, M., & Wu, G. (2011). Pursuing the wrong options? Adjustment costs and the relationship between uncertainty and capital accumulation. *Economics Letters*, *111*(3), 249–251.

Breitung, J., & Schmeling, M. (2013). Quantifying survey expectations: What's wrong with the probability approach? *International Journal of Forecasting*, *29*, 142–154.

Chirinko, R. S., & Schaller, H. (2009). The irreversibility premium. *Journal of Monetary Economics*, *56*, 390–408.

Clar, M., Duque, J.-C., & Moreno, R. (2007). Forecasting business and consumer surveys indicators: A time-series models competition. *Applied Economics*, *39*(20), 2565–2580.

Claveria, O., Pons, E., & Ramos, R. (2007). Business and consumer expectations and macroeconomic forecasts. *International Journal of Forecasting*, *23*, 47–69.

Conover, W. J. (1999). *Practical nonparametric statistics*. New York, NY: Wiley.

Cumby, R. E., & Modest, D. M. (1987). Testing for market timing ability: A framework for forecast evaluation. *Journal of Financial Economics*, *19*, 169–189.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

Doswell, C. A., Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, *5*, 576–585.

Driver, C., & Urga, G. (2004). Transforming qualitative survey data: Performance comparisons for the UK. *Oxford Bulletin of Economics and Statistics*, *66*(1), 71–90.

Eyraud, L. (2009). *Why isn't South Africa growing faster? A comparative approach* (IMF WP/09/25). Washington, DC: International Monetary Fund.

Fedderke, J. (2004). Investment in fixed capital stock: Testing for the impact of sectoral and systemic uncertainty. *Oxford Bulletin of Economics and Statistics*, *66*(2), 165–187.

Fedderke, J. (2009). Capital formation in South Africa. In J. Aron, B. Kahn, & G. Kingdon (Eds.), *South African economic policy under democracy* (pp. 182–207). Oxford, UK: Oxford University Press.

Fedderke, J., & Simkins, C. (2012). Economic growth in South Africa. *Economic History of Developing Regions*, *27*(1), 176–208.

Frankel, J., Smit, B., & Sturzenegger, F. (2008). South Africa: Macroeconomic challenges after a decade of success. *The Economics of Transition*, *16*(4), 639–677.

Granger, C. W. J. (1994). Forecasting in economics. In A. S. Weigend, & N. A. Gershenfeld (Eds.), *Time series prediction: forecasting the future and understanding the past* (pp. 529–538). Wokingham, UK: Addison-Wesley.

Granger, C. W. J. (1996). Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics*, *11*(5), 455–473.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, *54*, 513–533.

Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*, 120–127.

Hurvich, C. M., & Tsai, C. L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, *78*, 499–509.

Lui, S., Mitchell, J., & Weale, M. (2011). Qualitative business surveys: Signal or noise? *Journal of the Royal Statistical Society, Series a*, *174*(2), 327–348.

Merton, R. C. (1981). On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business*, *54*(3), 363–406.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. A. Mincer (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). Cambridge, MA: National Bureau of Economic Research.

Mitchell, J., Smith, R. J., & Weale, M. R. (2005). Forecasting manufacturing output growth using firm-level survey data. *The Manchester School*, *73*(4), 479–499.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*(3), 315–335.

OECD (2017). *Economic surveys: South Africa*. Brussels, Belgium: OECD Economics Department.

OECD. (2018). *OECD Economic Outlook 2018(1): Preliminary version*. Retrieved from http://www.oecd.org/eco/outlook/economic-forecast-summary-south-africa-oecd-economic-outlook.pdf

Pesaran, M. H. (1984). Formation of inflation expectations in British manufacturing industries. *Economic Journal*, *95*, 948–975.

Pesaran, M. H. (1987). *The limits to rational expectations*. New York, NY: Blackwell.

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, *10*(4), 461–465.

Smith, J., & McAleer, M. (1995). Alternative procedures for converting qualitative response data to quantitative expectations: An application to Australian manufacturing. *Journal of Applied Econometrics*, *10*, 165–185.

South African Reserve Bank (2018). Quarterly Bulletin, June.

Van Walbeek, C. (2013). How accurate are south African macroeconomic forecasts? *Journal for Studies in Economics and Econometrics*, *37*(3), 1–23.

Van Walbeek, C., & Sessions, M. (2007). A directional analysis of the Bureau for Economic Research's quarterly forecasts. *Studies in Economics and Econometrics*, *31*(3), 119–138.

Viegi, N. (2014). (Dis) saving in South Africa. In H. Bhorat, A. Hirsch, R. Kanbur, & M. Ncube (Eds.), *The Oxford Companion to the Economics of South Africa* (pp. 134–139). Oxford, UK: Oxford University Press.

von Kalckreuth, U. (2006). Financial constraints and capacity adjustment: Evidence from a large panel of survey data. *Economica*, *73*, 691–724.

World Bank (2010). *South Africa: Second investment climate assessment: Improving the business environment for job creation and growth* (Report 55699-ZA). Washington, DC: World Bank.

World Bank (2018). *Republic of South Africa systematic country diagnostic: An incomplete transition overcoming the legacy of exclusion in South Africa* (Report No. 125838-ZA). Washington, DC: World Bank.

AUTHOR BIOGRAPHIES

Ciaran Driver is Professor of Economics in the School of Finance and Management, SOAS University of London and a Fellow of the Academy of Social Sciences. He has published numerous articles on capital investment under uncertainty. Details of his books, articles and ongoing research projects can be found on https://www.researchgate.net/profile/Ciaran_Driver.

Nigel Meade is Emeritus Professor of Quantitative Finance, Imperial College, Business School, London, UK. He is an experienced statistical modeler with a background in operational research and statistics applied to innovation diffusion and finance. He has published over 50 papers and 8 book chapters, and successfully supervised more than 20 Ph.D. students.