Using forecasts of forecasters to forecast

Ingmar Nolte *, Winfried Pohlmeier [1]

Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany

Abstract

Quantification techniques are popular methods in empirical research for aggregating the qualitative predictions at the micro-level into a single figure. In this paper, we analyze the forecasting performance of various methods that are based on the qualitative predictions of financial experts for major financial variables and macroeconomic aggregates. Based on the Centre of European Economic Research's Financial Markets Survey, a monthly qualitative survey of around 330 financial experts, we analyze the out-of-sample predictive quality of probability methods and regression methods.

Using the modified Diebold–Mariano test of Harvey, Leybourne and Newbold (Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*, 281–291), we compare the forecasts based on survey methods with the forecasting performance of standard linear time series approaches and simple random walk forecasts.

© 2006 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

JEL classification: G10; E30; E31; E37; C10; C42

Keywords: Forecasting quality; Qualitative survey data; Quantification methods; Linear time series models; Turning points

1. Introduction

Using the assessments of experts when predicting the future development of major economic aggregates and financial variables has a long tradition in empirical economics. It is often regarded as an alternative to the standard time series framework. The notion behind this approach is that the individual assessments of experts gained through surveys, or an aggregation of these assessments obtained by a quantification technique, contains valuable information about the future which may be superior to the pure time series information. Even if one is willing to accept that experts' forecasts are based on superior information sets, and that problems such as strategic forecasting or herd behavior among the experts, which systematically bias their survey responses, are negligible, there is still the risk of loosing information due to the

* Corresponding author.Tel.: +49 7531 88 3753; fax: +49 7531 88 4450.

E-mail addresses: Ingmar.Nolte@uni-konstanz.de (I. Nolte),
Winfried.Pohlmeier@uni-konstanz.de (W. Pohlmeier).

[1] Tel.: +49 7531 88 2660; fax: +49 7531 88 4450.

qualitative nature of the information at the micro-level, combined with the use of a (potentially) misspecified quantification method. Moreover, the use of information owned by individuals with superior information sets through quantification methods, if not used as a stand alone forecasting model, may be useful in combination with pure time series approaches.

In this paper, we analyze the performance of both traditional time series methods and quantification methods with respect to their forecasting performance. Since sufficiently long time series of aggregated qualitative forecasts at the micro-level are available nowadays, we are able to use out-of-sample forecasting performance measures to compare the information content of aggregated micro-forecasts with the pure time series information. This allows us to undertake a rigorous comparison of the forecasting quality of survey methods and time series methods by means of statistical inference for 13 major economic and financial time series. In particular, we compare forecasts obtained by survey methods, such as the popular probability method of Carlson and Parkin (1975) and Pesaran's (1987) regression methods, with standard AR(p) forecasts, as well as with random walk forecasts as the ultimate benchmark for a simple forecasting rule. We compare our forecasts based on the Root Mean Squared Prediction Errors (RMSPE), which we find to be significantly different from each other using the modified Diebold and Mariano (1995) test of Harvey, Leybourne, and Newbold (1997). Furthermore, we compare our forecasts by their ability to predict the correct sign. Here we test for significance using a χ^2-independence test.

Our empirical analysis is based on data from the Financial Markets Survey (Finanzmarkttest) collected by the Centre of European Economic Research (ZEW) in Mannheim, Germany. The Financial Markets Survey is a monthly panel survey of around 330 experts in the financial sector which has been collected since 1991, and contains qualitative six-month forecasts of major macroeconomic aggregates and financial variables. For several reasons, experts' forecasts are usually being aggregated into a few numbers or even a single forecast, neglecting potentially valuable cross-sectional information. As a survey of qualitative responses, the ZEW Financial Markets Survey is representative of a number of

similar designed surveys such as the UK Gallup Poll, the German IFO Business Survey, the Confederation of Australian Industries (CAI)/Westpac Survey, and the Dunn & Bradstreet Business Expectation Survey. Our data allows us to distinguish between different groups of forecasters and to check for differences in the forecasting performance of these groups. Moreover, the fairly large time series dimension of our panel dataset allows us to identify systematic prediction biases of the survey methods in the neighborhood of turning points. We construct a simple turning point indicator based on the information in the corresponding series sampled at a higher (daily) than monthly frequency, which we incorporate into our forecasting framework as additional information, especially for the pure financial series.

The rest of this paper is structured as follows: In Section 2, we briefly introduce the two main approaches to exploiting qualitative forecasts at the micro-level and present our unifying approach. Relevant information on the data is summarized in Section 3. Section 4 contains our empirical findings, while Section 5 concludes and gives an outlook on future research.

2. Forecasting methodology

A major purpose of qualitative survey data is to provide the basis for a k-step ahead forecast (usually some months) for a variable of interest P_t, where $t = 1, \ldots, T$ denotes a time index. The variables of interest differ depending on the research focus of the survey-collecting institute; for example, macroeconomic series like growth domestic/net products, consumer price indices, unemployment rates, or even financial time series like stock indices or foreign exchange rates. In most of the surveys, the individual respondents are experts who give their assessments of the future development of the variables of interest. The information contained in such a survey can therefore be considered as an addition to public information stemming from the progression of the time series. We therefore distinguish between two different types of information filtrations. Let $\mathfrak{F}_t^S$ be the information contained in the survey, and $\mathfrak{F}_t^T$ be the information stemming from the time series at time t. Both filtrations are summarized in $\mathfrak{F}_t = \sigma(\mathfrak{F}_t^S, \mathfrak{F}_t^T)$.

Then the best k-step ahead forecast of P_t in terms of the mean squared error criterion is given by the conditional expectation $E[P_{t+k}|\mathfrak{F}_t]$. Our strategy is to exploit the available information in an optimal way to obtain an estimable conditional mean function. In this set-up, a pure time series model and a survey method define the two extreme benchmarks. A time series approach ignores the survey information completely, while a survey approach neglects the information on the dynamics of the variable of interest. Surprisingly, pure survey methods are used in practice quite frequently,[2] and can be associated with the literature of probability methods which dates back to Carlson and Parkin (1975) or even Anderson (1951). Probability methods typically quantify (aggregate) qualitative survey information to the k-step ahead forecast. An alternative method that exploits both sources of information, and which is usually referred to as the second strand of quantification methods, is the regression method of Pesaran (1987). The idea is to consider the overall expected six month forecast to consist of the weighted sum of individual six month forecasts, where these individual six month forecasts are based on simple time series models. In the following we will briefly review the probability and regression methods in a unifying notation.

2.1. Probability methods

The probability method was proposed by Anderson (1951, 1952), Theil (1952, 1958) and Pflanzagl (1952). It was refined by Knöbl (1974), and Carlson and Parkin (1975), and was criticized and extended inter alia by Foster and Gregory (1977), Fishe and Lahiri (1981), Batchelor (1982), Batchelor and Orr (1988), Seitz (1988), Fishe and Idson (1990), Dasgupta and Lahiri (1992), Smith and McAleer (1995), Cunnigham, Smith, and Weale (1998), and Mitchell, Smith, and Weale (2002). There are alternative ways to derive the probability methods. Here, we review them in terms of ordered response approaches, where the conditional mean function of the latent variable is only a constant.[3] For notational simplicity, we assume

a forecasting horizon of $k=6$, which corresponds to the forecasting horizon asked for in the ZEW Financial Market Survey. Let N_t be the number of participants in the survey taken at time t; $i=1,\ldots,N_t$ denotes the ith participant, and $t=1,\ldots,T$ the time index. In order to circumvent stationarity problems we do not focus on the level P_t as the variable of interest, but on the six month growth rate (return rate) $r_{t,t+6} \equiv \frac{P_{t+6}-P_t}{P_t}$, except for inflation rates, where we consider six month price changes, $d_{t,t+6} \equiv P_{t+6} - P_t$. This means that the forecast $E[P_{t+6}|\mathfrak{F}_t]$ will be obtained by using the decomposition $E[P_{t+6}|\mathfrak{F}_t] = P_t \left(1 + E[r_{t,t+6}|\mathfrak{F}_t]\right)$ or $E[P_{t+6}|\mathfrak{F}_t] = P_t + E[d_{t,t+6}|\mathfrak{F}_t]$ respectively.[4] Let $\mathfrak{F}_{it}$ be the information σ-Algebra of participant i at time t. We assume that the participants can give an independent assessment for the expected conditional six months return with probability $(1-\alpha_{it})$ and that they cannot give an assessment with probability α_{it}. Thus we observe with probability $(1-\alpha_{it})$ his up/same/down-response, conditional on his information $\mathfrak{F}_{it}$, i.e. $Y_{it}|\mathfrak{F}_{it}$, as

$$
Y_{it}|\mathfrak{F}_{it} = \begin{cases} \text{up} & (+1) & \tau_{it}^u < Y_{it}^*|\mathfrak{F}_{it} \\ \text{same} & (0) & \text{if } \tau_{it}^d < Y_{it}^*|\mathfrak{F}_{it} \le \tau_{it}^u \\ \text{down} & (-1) & Y_{it}^*|\mathfrak{F}_{it} \le \tau_{it}^d \end{cases}
$$

where $Y_{it}^*|\mathfrak{F}_{it}$ is the latent variable and τ_{it}^d, τ_{it}^u with $\tau_{it}^d < \tau_{it}^u$ are the corresponding down- and up- thresholds for participant i at time t. Assuming that at time t each participant knows the current base level value for which he is going to give a six-month direction-forecast, one can interpret the expectation of his conditional latent variable $Y_{it}^*|\mathfrak{F}_{it}$ as his conditional expected six month return $E[r_{t,t+6}|\mathfrak{F}_{it}]$. For estimation purposes we specify a model for $Y_{it}^*|\mathfrak{F}_{it}$, assuming that

$$
Y_{it}^*|\mathfrak{F}_{it} = \mu(x_{it}) + \varepsilon_{it}, \tag{1}
$$

where $E[\varepsilon_{it}|\mathfrak{F}_{it}] = 0, E[\varepsilon_{is}\varepsilon_{it}|\mathfrak{F}_{it}] = 0$ for all $s \neq t$, $E[\varepsilon_{it}\varepsilon_{jt}|\mathfrak{F}_{it}] = 0$ for all $j \neq i, E[\varepsilon_{it}^2|\mathfrak{F}_{it}] < \infty$, and where the conditional mean $\mu(x_{it})$ is a scalar mean function which is allowed to depend on k individual

---

[2] An example is the ZEW: their forecasts are based on the Carlson and Parkin (1975) method.

[3] See Fishe and Lahiri (1981), Fishe and Idson (1990) or Kaiser and Spitz (2000).

[4] In the sequel we consider only the return case, the price change case is analogous.

characteristics $x_{it} \in \mathbb{R}^k$. Given $\varepsilon_{it}|\mathfrak{F}_{it} \sim N\left(0, \ E\left[\varepsilon_{it}^2|\mathfrak{F}_{it}\right]\right)$ the model is equivalent to an ordered probit model with three categories:

$$
\begin{aligned}
P_{it}^- &\equiv P(Y_{it} = -1|\mathfrak{F}_{it}) = (1-\alpha_{it})\Phi\left(\frac{\tau_{it}^d - \mu(x_{it})}{V[\varepsilon_{it}|\mathfrak{F}_{it}]^{1/2}}\right), \\
P_{it}^= &\equiv P(Y_{it} = 0|\mathfrak{F}_{it}) \\
&= (1-\alpha_{it})\left(\Phi\left(\frac{\tau_{it}^u - \mu(x_{it})}{V[\varepsilon_{it}|\mathfrak{F}_{it}]^{1/2}}\right) - \Phi\left(\frac{\tau_{it}^d - \mu(x_{it})}{V[\varepsilon_{it}|\mathfrak{F}_{it}]^{1/2}}\right)\right), \\
P_{it}^+ &\equiv P(Y_{it} = -1|\mathfrak{F}_{it}) = (1-\alpha_{it})\Phi\left(-\frac{\tau_{it}^u - \mu(x_{it})}{V[\varepsilon_{it}|\mathfrak{F}_{it}]^{1/2}}\right).
\end{aligned}
\tag{2}
$$

Carlson and Parkin (1975) set $\mu(x_{it}) \equiv \text{const}_t$ and interpret const_t as the consensus conditional expectation of all survey participants, i.e. $\text{const}_t \equiv E\left[r_{t,t+6}|\mathfrak{F}_t^S\right]$. In addition, they assume for each t that $\tau_{it}^d = \tau_t^d, \tau_{it}^u = \tau_t^u, V[\varepsilon_{it}|\mathfrak{F}_{it}]^{1/2} = \sigma_t$ and $\alpha_{it} = \alpha_t$. One equation in system (2) is redundant since the probabilities have to sum up to $(1 - \alpha_{it})$. The "no assessment" state is assumed to be independent of the other outcomes. Because there are now two equations with four unknown parameters ($\text{const}_t, \sigma_t, \tau_t^d, \tau_t^u$), additional identification restrictions have to be imposed. Carlson and Parkin (1975) assume the thresholds to be symmetric and identical across time ($-\tau_t^d = \tau_t^u = \tau$) to obtain τ as an estimate from the following time series regression:

$$
\begin{aligned}
r_{t-6,t} = &-\tau \frac{\Phi^{-1}\left(\frac{\hat{P}_{t-6}^+}{(1-\hat{\alpha}_{t-6})}\right) - \Phi^{-1}\left(\frac{\hat{P}_{t-6}^-}{(1-\hat{\alpha}_{t-6})}\right)}{\Phi^{-1}\left(\frac{\hat{P}_{t-6}^+}{(1-\hat{\alpha}_{t-6})}\right) + \Phi^{-1}\left(\frac{\hat{P}_{t-6}^-}{(1-\hat{\alpha}_{t-6})}\right)} \\
&+ \varepsilon_t,
\end{aligned}
\tag{3}
$$

where $\hat{\ }$ denotes the sample probability estimator of the corresponding probability. Batchelor (1982) suggests using the OLS estimator instead. Another possibility for identifying the system is to assume a one-argument functional relationship between the thresholds and to provide an estimate of the conditional standard deviation σ_t. This can be done by using an estimator which exploits the underlying time series, such as the sample standard deviation, a GARCH-model standard deviation, or a realized volatility[5] standard deviation. However, solving the

Carlson and Parkin (1975) equation system[6] for $E\left[r_{t,t+6}|\mathfrak{F}_t^S\right]$ and σ_t yields

$$
E\left[r_{t,t+6}|\mathfrak{F}_t^S\right] = -\tau \frac{\Phi^{-1}\left(\frac{P_{it}^+}{(1-\alpha_t)}\right) - \Phi^{-1}\left(\frac{P_{it}^-}{(1-\alpha_t)}\right)}{\Phi^{-1}\left(\frac{P_{it}^+}{(1-\alpha_t)}\right) + \Phi^{-1}\left(\frac{P_{it}^+}{(1-\alpha_t)}\right)},
\tag{4}
$$

$$
\sigma_t = \frac{-2\tau}{\Phi^{-1}\left(\frac{P_{it}^+}{(1-\alpha_t)}\right) + \Phi^{-1}\left(\frac{P_{it}^-}{(1-\alpha_t)}\right)}.
\tag{5}
$$

Estimates of $E\left[r_{t,t+6}|\mathfrak{F}_t^S\right]$ and σ_t are obtained by replacing P_{it}^+, P_{it}^- and α_t with their sample estimators. An alternative method for solving equation system (2) is to perform Maximum Likelihood estimation, which becomes indispensable if $\mu(x_{it})$ is parameterized not only by a constant.

Several modifications of the original Carlson and Parkin method have been proposed, see e.g. Wren-Lewis (1986), Pesaran (1987), Batchelor and Orr (1988), and Maddala (1990). Dasgupta and Lahiri (1992) use the logistic probability density function instead of the normal because it allows for leptokurtosis. For the same reason Carlson (1977) and Lahiri and Teigland (1987) use a non-central scaled t-distribution and found that it fits the examined inflation rates remarkably well, whereas Dasgupta and Lahiri (1992), who compared the normal-, logistic- and non-central scaled t-distributions, found that there is no significant difference when using the more flexible distributions, at least in the U.S. National Association of Purchasing Management-data. Despite the generalizations, the probability method still has a number of drawbacks. The most serious one is that only the information contained in the survey is used to derive the forecast of interest, i.e. either one neglects nearly all time series information or one has to assume that the survey participants account for this a priori. Another drawback of the Carlson and Parkin methodology is that the conditional mean rises, assuming a fixed portion for the up- and not-assessable-categories, where the up-portion is greater than one half, and a shift from the same- to the down-category. Of course,

---

[5] See Andersen, Bollerslev, Diebold, and Labys (2001).

[6] This is system (2), imposing the Carlson and Parkin (1975) assumptions and considering the second equation as redundant for ease of computation.

the reverse of this argument also holds and can easily be proven. In this case, taking the derivative of the conditional mean in Eq. (4) with respect to the down-probability yields

$$\frac{\partial E\left[r_{t,t-6}^S|\mathfrak{F}_t\right]}{\partial P_{it}^-}>0. \tag{6}$$

For a more elaborate discussion of probability methods we refer to Pesaran (1987).

2.2. Regression methods

To overcome some of the drawbacks of the probability methods, Pesaran (1984, 1985, 1987) proposed an alternative strategy for quantifying survey data for forecasting purposes by means of a linear regression. These "regression methods" are based on the fundamental equation

$$E\left[r_{t,t+6}|\mathfrak{F}_t\right] \equiv \sum_{i=1}^{N_t^u} w_{it}^u E^u\left[r_{t,t+6}|\mathfrak{F}_{it}\right]$$
$$+ \sum_{i=1}^{N_t^d} w_{it}^d E^d\left[r_{t,t+6}|\mathfrak{F}_{it}\right],$$

where w_{it}^u and w_{it}^d are weights on the ith participant responding up and down, respectively, and $E^u\left[r_{t,t+6}|\mathfrak{F}_{it}\right]$ and $E^d\left[r_{t,t+6}|\mathfrak{F}_{it}\right]$ are the quantified conditional expectations under the condition that the ith survey participant responds up or down, respectively. The corresponding regression equation takes the following form

$$r_{t-6,t} = \sum_{i=1}^{N_t^u} w_{it-6}^u\left(\mu^u(\mathfrak{F}_{it-6}) + \varepsilon_{it}^u\right)$$
$$+ \sum_{i=1}^{N_t^d} w_{it-6}^d\left(\mu^d(\mathfrak{F}_{it-6}) + \varepsilon_{it}^d\right)$$

where $\mu^{u/d}(\mathfrak{F}_{it-6})$ denote parametric (dynamic) models for the conditional expectations. Pesaran (1987) and Smith and McAleer (1995) parameterize $\mu^{u/d}(\mathfrak{F}_{it-6})$ as problem related linear functions and assume equal weights for the participants. The Anderson (1952) model can be considered as a simple specification in this class assuming $\mu^u(\mathfrak{F}_{it-6}) = \alpha^u$, $\mu^d(\mathfrak{F}_{it-6}) = \alpha^d$

and equal weights. This yields the forecasting equation for the 6-months return in t

$$r_{t-6,t} = \alpha^u \hat{P}_{t-6}^+ + \alpha^d \hat{P}_{t-6}^- + \epsilon_t, \tag{P1}$$

where $\epsilon_t = \sum_{i=1}^{N_{t-6}^u} w_{it-6}^u \varepsilon_{it}^u + \sum_{i=1}^{N_{t-6}^d} w_{it-6}^d \varepsilon_{it}^d$. As before, $\hat{P}_t^{+/-}$ represents the sample portions of those participants who report up/down. In the following we refer to the Anderson model as model (P1).

An augmented specification is the general error learning model of Nerlove (1983) with $\mu^u(\mathfrak{F}_{it-6}) = \alpha^u + \beta_p(L)r_{t-12,t-6}$, and $\mu^d(\mathfrak{F}_{it-6}) = \alpha^d + \beta_p(L)r_{t-12,t-6}$ where $\beta_p(L)$ denotes the pth order lag polynomial. Retaining the equal weights assumption yields

$$r_{t-6,t} = \alpha^u \hat{P}_{t-6}^+ + \alpha^d \hat{P}_{t-6}^- + \hat{P}_{t-6}^+ \beta_p(L)r_{t-12,t-6}$$
$$+ \hat{P}_{t-6}^- \beta_p(L)r_{t-12,t-6} + \epsilon_t \tag{P2}$$

In the remainder of the paper we refer to this model as the augmented Pesaran-type model (P2), where the sample portions $\hat{P}_t^{+/-}$ enter the linear regression function directly or as interaction variables, which interact with those variables determining the dynamics.

2.3. Time series strategies

As an alternative to the probability and regression methods which both exploit information contained in the survey (though in different ways), we consider conventional time series models. These models neglect the information contained in the survey in the first step and rely only on the time series information $\mathfrak{F}_t^T$. Assuming $E\left[r_{t,t+6}|\mathfrak{F}_t^T\right] \equiv 0$ yields the random walk (RW) model for P_t. The RW model serves as the extreme benchmark in the class of pure time series models. A natural class of time series models to consider is the class of autoregressive moving average (ARMA) models. Here, we focus only on autoregressive AR(p) models for two reasons. Firstly, they can easily be used for forecasting. Secondly, we want to restrict our attention to parsimonious models, to avoid problems with data mining. In particular, we consider models of the form

$$\left(1 - \beta_p(L)\right)r_{t-6,t} = \alpha + \varepsilon_t \tag{T1}$$

where again $\beta_p(L)$ denotes the pth order lag polynomial, and, we assume for further simplification of the

computation of $E\left[r_{t,t+6}\mid\mathfrak{F}_t^T\right]$ that $\beta_1=\beta_2=\cdots=\beta_5=0$. We refer to this model as model (T1).

The first and third graphs of Fig. 2 below depict the plot of the true series (DAX 30, USD/EUR foreign exchange rate) against the forecasts from RW, the Carlson–Parkin (CP) and AR(p) models. We observe that the RW forecast (since it is only a 6 month shift of the underlying series), and more surprisingly, the CP forecasts, tend to over- (under-) predict the true price level in upward (downward) sloping time periods. One possible interpretation of this effect is the existence of a systematic misprediction of the CP forecasts. We try to exploit this systematic deviation by the following kind of error correction procedure

$$r_{t-6,t} = \alpha + \beta_p(L)\eta_{t-6,t} + \varepsilon_t \tag{T2}$$

where $\beta_p(L)$ is defined as above, and $\eta_{t-6,t}$ denotes the systematic misprediction of the CP forecast defined as $\eta_{t-6,t}\equiv\hat{E}\left[r_{t-6,t}\mid\mathfrak{F}_{t-6}^S\right] - r_{t-6,t}$, where $\hat{E}\left[r_{t-6,t}\mid\mathfrak{F}_{t-6}^S\right]$ is the CP estimate of $r_{t-6,t}$. Model (T2) can be interpreted as a random walk with drift extended by the forecasting errors of the CP method; i.e. if $\beta_p=0$, $\forall p$, then the CP forecasting errors do not contain any systematic information that can be exploited to improve the RW forecasts. This model is not a pure time series information based model, since it now incorporates the survey information via the CP forecasts.

A different strategy is to capture the behavior described in Fig. 2 with a model that directly accounts for the downward- and upward-sloping trend periods by including information on the turning points of the underlying series. A reasonable strategy is to use the information contained in the history of the process itself. Unlike macroeconomic time series, the financial series such as interest rates, exchange rates, and stock indices, can be sampled on much higher than monthly frequencies. For these series it could be appropriate to exploit the information contained in daily frequency data to construct a turning point indicator. Therefore, we construct dummy variables as a first step, which can easily be included in the model specifications suggested above. Let D_t^u and D_t^d denote dummy variables indicating an upward or downward sloping trend period at time t. The

AR(p) model (T1), augmented with trend dummies, is given by

$$\left(1 - \beta_p(L)\right)r_{t-6,t} = \alpha + \beta^u D_{t-6}^u + \beta^d D_{t-6}^d + \varepsilon_t, \tag{T3}$$

and we refer to it as model (T3). Finally we consider the possibility that experts may react differently to downward and upward sloping trend periods, e.g. they may be overconfident in upward sloping trend periods and therefore expect and predict higher values. An extension of model (T3) which accounts for this in a simple way is given by

$$\begin{aligned}
\left(1 - \beta_p(L)\right)r_{t-6,t} = {} & \alpha + \beta^u D_{t-6}^u + \beta^d D_{t-6}^d \\
& + \theta_1 D_{t-6}^u \hat{P}_{t-6}^+ + \theta_2 D_{t-6}^d \hat{P}_{t-6}^+ \\
& + \theta_3 D_{t-6}^u \hat{P}_{t-6}^= + \theta_4 D_{t-6}^d \hat{P}_{t-6}^= \\
& + \theta_5 D_{t-6}^u \hat{P}_{t-6}^- + \theta_6 D_{t-6}^d \hat{P}_{t-6}^- + \varepsilon_t,
\end{aligned} \tag{T4}$$

where we include the interactions between the sample probabilities and the trend indicators as regressors. We augment the simple Pesaran model (P1) in a similar way to the augmentation of model (T1) to models (T3) and (T4), and we refer to these augmented versions as models (P3) and (P4).

The trend dummies are computed as a function of two moving averages. The idea is to measure the difference between a long term (≈ 1 year: 250 trading days) and a short term (≈ 1 month: 20 trading days) moving average, and according to the value of this difference, to switch between three regimes: upward trend, no trend, and downward trend. To formalize the discussion, $c_u\in\mathbb{R}^+, c_d\in\mathbb{R}^-$, $s_{d_1}\in\mathbb{R}^+, s_{d_2}\in\mathbb{R}^+, s_{u_1}\in\mathbb{R}^-, s_{u_2}\in\mathbb{R}^-$ where $s_{d_1}<s_{d_2}$ and $s_{u_1}<s_{u_2}$ denote corridor- and support-levels and $h_1,h_2\in\mathbb{N}$ are time horizons in days, with $h_1<h_2$. We summarize all these parameters in the parameter vector κ. Let P_τ denote the time series P_t sampled a daily (not at a monthly) frequency and let τ_t denote that day where the questionnaire was to be returned in month t. The corresponding price is $P_{\tau t}$. Let $\bar{P}_t^j$ denote the h_j-day average of P_τ, $j=1,2$, computed as $\bar{P}_t^j = \frac{1}{h_j}\sum_i P_{\tau_t-h_j+i}$. The trend

dummies $D_t^d = D_t^d(\boldsymbol{\kappa})$ and $D_t^u = D_t^u(\boldsymbol{\kappa})$ are then defined as

$$D_t^d \equiv \begin{cases} 1 & \text{if} \quad D_t(\boldsymbol{\kappa}) = -1 \\ 0 & \text{otherwise} \end{cases} \qquad D_r^u \equiv \begin{cases} 1 & \text{if} \quad D_t(\boldsymbol{\kappa}) = -1 \\ 0 & \text{otherwise} \end{cases}$$

with

$$D_t(\boldsymbol{\kappa}) \equiv \begin{cases} -1 & \text{if} \quad D_{t-1}(\boldsymbol{\kappa}) = 0 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in (-\infty, c_d] \\ & \text{or if} \quad D_{t-1}(\boldsymbol{\kappa}) = -1 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in (-\infty, s_{d_1}] \\ & \text{or if} \quad D_{t-1}(\boldsymbol{\kappa}) = 1 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in (-\infty, s_{u_1}] \\ 0 & \text{if} \quad D_{t-1}(\boldsymbol{\kappa}) = 0 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in (c_d, c_u) \\ & \text{or if} \quad D_{t-1}(\boldsymbol{\kappa}) = -1 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in (s_{d_1}, s_{d_2}) \\ & \text{or if} \quad D_{t-1}(\boldsymbol{\kappa}) = 1 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in (s_{u_1}, s_{u_2}) \\ 1 & \text{if} \quad D_{t-1}(\boldsymbol{\kappa}) = 0 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in [c_u, \infty) \\ & \text{or if} \quad D_{t-1}(\boldsymbol{\kappa}) = -1 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in [s_{d_2}, \infty) \\ & \text{or if} \quad D_{t-1}(\boldsymbol{\kappa}) = 1 \wedge \left(\overline{P_t^1} - \overline{P_t^2}\right) \in [s_{u_2}, \infty) \end{cases}$$

where we have to specify an initial trend direction $D_0(\boldsymbol{\kappa})$. The models introduced above incorporate different kinds of information sets. The pure AR(p) model (T1) is related to an information filtration sampled at a monthly frequency. The AR(p) model extended with the trend dummies (T3) additionally incorporates information from the daily filtration. The error correction model (T2) uses aggregated information stemming from the survey and time series information based on a monthly sampling frequency. Model (T4) is the richest specification in terms of the information set, since it incorporates, like the Pesaran-type models, survey information and time series information, where the time series information is sampled at both a daily and a monthly frequency.

Fig. 1. Participation in the ZEW Financial Markets Survey from December 1991 to April 2004.

3. Data

The ZEW Financial Markets Survey has been collected since December 1991 on a monthly basis. It focuses on international macroeconomic time series, and in particular, on financial market series. The current questionnaire consists of nine standard questions, which can be further classified as one retrospective and eight prospective questions. The participants are asked to give their assessment of the current economic situation (retrospective) and to make six month predictions for the overall economic situation, the inflation rate, short and long term interest rates, stock-indices, exchange rates, and profits in different sectors of the German economy, as well as the oil price. With the exception of the last three, all questions are asked with respect to Germany, USA, Japan, Great Britain, France, Italy, since January 1999, with respect to the Euro-Area. Up to November 1998, the exchange rate question relates to US-Dollar, Yen, UK-Pound and Swiss Franc per DM and since December 1998 per euro. The ZEW Financial Market Survey is similar to the CAI/Westpac survey to the CAI/Westpac survey (Smith & McAleer, 1995), the Confederation of British Industry Industrial Trends Survey (Mitchell et al., 2002), and the 1961 United Kingdom Gallup Poll (Carlson & Parkin, 1975), in that it is a purely qualitative survey, where the respondents are expected to choose between four categories: up, same, down and no assessment.[7] In the early years, the questionnaire was to be returned by the third Friday of the month in question, but this deadline changed to the second Friday in December 1999. Responses obtained over the weekend, or on the Monday or Tuesday after the Friday of submission are also utilized in the computations. The ZEW Financial Markets Survey is claimed to be an expert survey because the majority of the respondents are employees of financial markets or economic research divisions at (i) banks or investment firms, (ii) insurance companies and (iii) industry firms. In August 1999 the group of respondents who do not belong to (i), (ii) or (iii) became large enough to be referred to as group (iv),
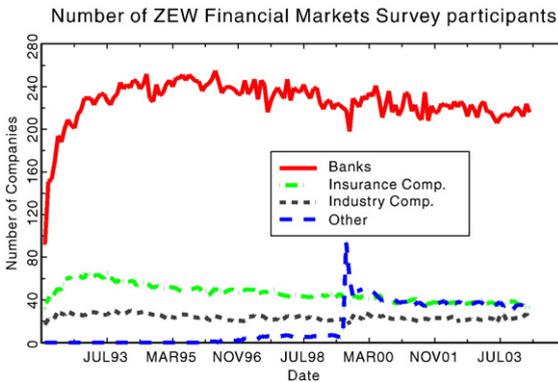
---

[7] For the retrospective question 'current economic situation' the categories are designated: good, normal, bad and no assessment.

other companies. From December 1991 to April 2004 (149 months) 918 participants took part at least once in the ZEW Financial Markets Survey, whereas the number of companies responding per month has levelled off at around 330. For the progression of the participation figures split up into the four groups see Fig. 1. Every month the ZEW distributes the main findings of the survey among the participants and others via the Internet[8] in the form of a five page ZEW Financial Markets Report (Finanzmarktreport). This report includes a detailed listing of the changes in the percentages of the different response categories, a point estimation, as well as its standard deviation, for the inflation rate, the short and long term interest rates, the stock indices, the exchange rates, and the oil price. The ZEW quantification method is a variant of the Carlson and Parkin (1975) quantification procedure where identification is obtained through estimates of the threshold values obtained from the survey respondents.

4. Empirical findings

Our empirical analysis is carried out using the following variables: German, US and Japanese inflation rates; German, US and Japanese short term interest rates; Deutscher Aktienindex 30 (DAX 30), Dow Jones Industrial Index, Nikkei 225 Index, FTSE 100 Index, CAC 40 Index; the USD/EUR foreign exchange (FX) rate, and the GBP/EUR FX rate. We estimate the models which include lagged variables recursively. We start making 6 month forecasts in September 2000 using the first 100 available observations since June 1992. The forecasts in October 2000 are then based on 101 observations, the forecasts in November 2000 on 102 observations, and so on. Thus, our prediction period ranges from March 2001 to May 2004 and contains 39 observations. For deriving the trend indicators, we use a 250 day (long term) moving average and a 20 day (short term) moving average. The thresholds are set to the following values: $\frac{c_d}{P_t} = -1\%, \frac{c_u}{P_t} = 4\%, \frac{s_{d_1}}{P_t} = -3\%, \frac{s_{d_2}}{P_t} = -6\%, \frac{s_{u_1}}{P_t} = -3\%, \frac{s_{u_1}}{P_t} = -6\%$. We are able to compute the trend indicators only for those variables which can be observed at a

daily frequency, i.e. DAX 30, Dow Jones, Nikkei 225, FTSE 100 Index, CAC 40 Index, the USD/EUR FX rate, and the GBP/EUR FX rate. Fig. 2 shows the forecast of the Carlson–Parkin (CP), the random walk (RW), and AR(p) models (T1) in comparison to the true DAX and USD/EUR FX series. Since the CP and RW forecasting techniques do not require an estimation period, the forecasts are available from June 1992 to May 2004 (which is the last date where the true underlying value was available). The forecasts from the AR(p) models are available from March 2001 onwards. The identification in the CP method is achieved by the information on thresholds obtained by the ZEW in a specific question to respondents. The plots, as well as the quality of our findings, are robust to the identification method chosen. We also identified the parameters by assuming symmetric thresholds and by using the sample standard deviation, a GARCH, and a Realized Volatility based standard deviation estimator. Including further explanatory variables (like responses to other questions) in the conditional mean function of the latent variable does not change the results qualitatively. Therefore, we restrict our attention to the parsimonious CP specification and refer to CP as the specification identified by the ZEW-thresholds without any further explanatory variables in the conditional mean function but only the constant. Both the RW and the CP forecasts reveal a clear tendency to over- (under-) predict the true price level in upward (downward) sloping time periods. Given that the CP method exploits superior information, it is somewhat surprising that the mispredictions of the CP method are mostly even more pronounced than those of the RW model. We also observe under- and over-prediction for the AR(p) model forecasts, but it is less severe than the CP and RW forecasts. These results confirm that the CP method does not quantify the superior micro-information appropriately, since the AR(p) forecasts are based on pure time series information.

The quality of the forecasts for the 39 months under consideration is analyzed using the Root Mean Squared Prediction Error (RMSPE) and the number of correct direction forecasts. We test for significant differences in the RMSPE using the modified Diebold–Mariano test of Harvey et al. (1997). The independence of the true and forecasted directions is tested by a χ^2-independence test. Table
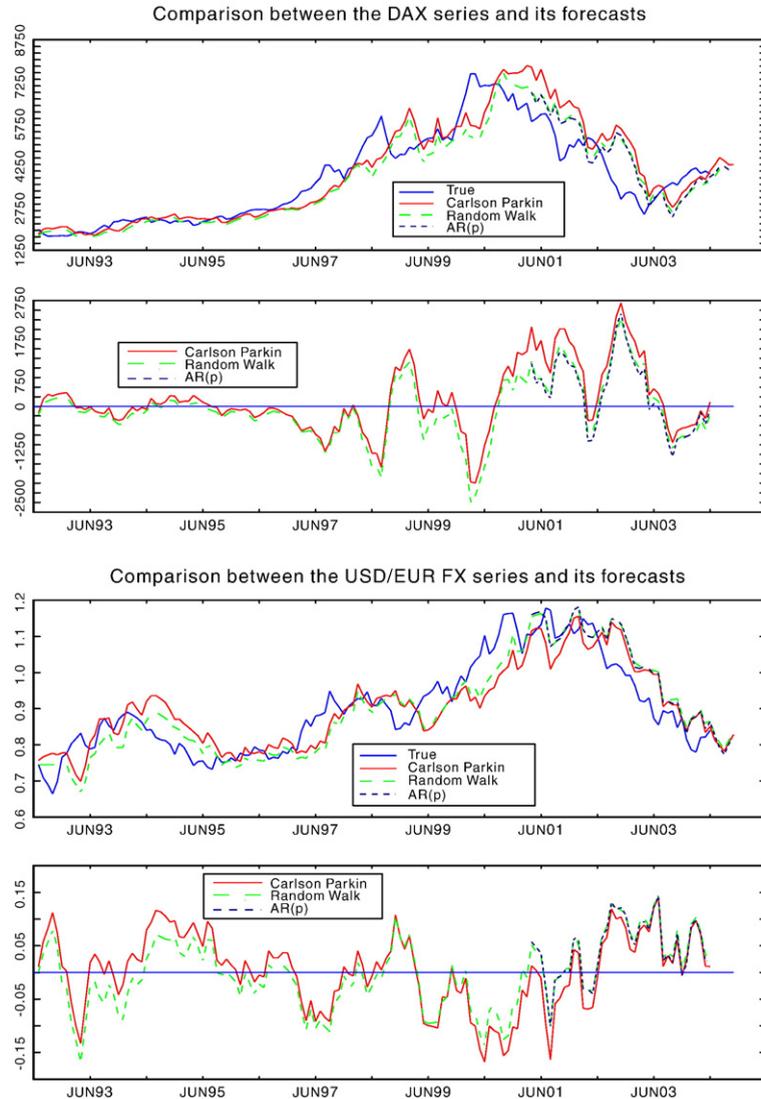
---

[8] http://www.zew.de.

23

Fig. 2. The first panel shows the DAX series in comparison to the series forecasted using the random walk, Carlson–Parkin, and AR(p) models. The values of the differences between the forecasted series and the DAX series are plotted in the second panel. Panels three and four contain the same information for the USD/EUR FX series.

1 reports on the RMSPEs and on the results of the Diebold–Mariano tests for all 13 different macroeconomic aggregates and financial variables. The two figures in parentheses below the value of the RMSPE show the p-values of the Diebold–Mariano test. The first (second) figure reports on the p-value of the Diebold–Mariano test with the null hypothesis that the RMSPE of the corresponding model is significantly smaller than the RMSPE of the random

walk (CP) forecast. The first two columns of Table 1 contain the results of the forecasting performance of the two benchmark models.

Obviously, using the RW forecast instead of the CP method forecast does not appear to be a bad strategy. The RMSPE is smaller for the RW than for the CP method for 7 out of the 13 time series. However (Japanese Inflation, DAX 30, FTSE 100 and CAC 40), the forecasting quality turns out to be signifi-

ge 24

Table 1
Root mean squared prediction errors and prediction test

Variable description	Benchmarks		Time series models				Pesaran type models			
	Rand. Walk	Carl.–Park.	(T1)	(T2)	(T3)	(T4)	(P1)	(P2)	(P3)	(P4)
Inflation rate (Germany)[†]	4135	4132	3029	3030			4140	3105		
	(•/87.8)	(12.2/•)	(3.1/3.1)	(3.1/3.1)			(50.3/50.6)	(0.7/0.7)		
Inflation rate (USA)[†]	3702	3701	3050	3051			4333	5229		
	(•/83.8)	(16.2/•)	(3.1/3.1)	(3.1/3.1)			(93.0/93.0)	(54.4/54.4)		
Inflation rate (Japan)[†]	3911	3912	2760	2760			3946	3282		
	(•/4.3)	(95.7/•)	(0.1/0.1)	(0.1/0.1)			(99.4/99.4)	(0.3/0.3)		
Short term interest rate (Germany)	0.574	0.565	0.539	0.535			0.412	0.406		
	(•/82.2)	(17.8/•)	(28.9/34.2)	(26.5/31.7)			(11.3/11.6)	(10.0/10.1)		
Short term interest rate (USA)	1.071	1.061	0.952	0.944			0.836	0.920		
	(•/86.0)	(14.0/•)	(25.6/26.4)	(24.6/25.3)			(12.5/12.5)	(5.4/4.3)		
Short term interest rate (Japan)	0.150	0.151	0.293	0.293			0.150	0.149		
	(•/18.3)	(81.7/•)	(86.0/86.0)	(86.0/86.0)			(49.4/49.2)	(0.1/0.6)		
DAX 30 index	1006	1297	985.7	986.2	1241	1124	1264	1151	1092	1144
	(•/4.8)	(95.2/•)	(18.7/5.4)	(15.1/5.6)	(87.1/32.6)	(72.5/13.6)	(90.4/34.1)	(95.2/12.1)	(78.1/8.7)	(84.7/16.2)
Dow Jones industrial index	1016	1204	1077	1159	1322	1470	1222	1625	1613	1787
	(•/21.8)	(78.2/•)	(67.8/35.2)	(70.1/46.1)	(98.8/79.0)	(95.9/78.6)	(97.4/53.3)	(91.7/76.9)	(92.5/76.9)	(95.2/84.1)
Nikkei 225 index	1891	2098	1998	1950	2089	2153	1869	2127	2199	2269
	(•/11.4)	(88.6/•)	(85.1/18.8)	(57.8/23.4)	(95.9/47.7)	(92.4/60.8)	(41.8/19.4)	(93.6/60.0)	(95.8/81.3)	(94.5/86.9)
FTSE 100 index	594.7	774.8	598.8	592.8	954.1	1028	624.9	677.2	757.0	916.5
	(•/4.3)	(95.7/•)	(52.2/14.6)	(49.4/18.3)	(92.8/83.4)	(83.5/74.0)	(61.9/17.1)	(64.8/36.7)	(77.8/47.4)	(97.3/76.0)
CAC 40 index	749.9	1003	834.9	731.4	1009	881.8	897.6	838.7	847.1	867.2
	(•/4.8)	(95.2/•)	(77.1/5.9)	(15.0/4.9)	(86.5/51.9)	(79.9/14.6)	(89.1/2.8)	(76.0/10.1)	(74.7/9.3)	(79.1/16.4)
FX-rate ($/€)	0.073	0.070	0.071	0.072	0.084	0.128	0.097	0.104	0.131	0.167
	(•/59.6)	(40.4/•)	(15.2/52.5)	(23.1/58.7)	(92.9/82.3)	(93.1/91.3)	(99.9/99.9)	(96.0/98.3)	(96.9/96.8)	(96.9/96.5)
FX-rate (£/€)	0.062	0.061	0.063	0.066	0.071	0.066	0.061	0.063	0.065	0.059
	(•/55.9)	(44.1/•)	(60.0/63.3)	(99.9/74.7)	(99.9/90.0)	(59.6/69.0)	(47.6/49.7)	(52.1/56.2)	(54.6/59.2)	(41.4/40.6)

[†] Values of the RMSPE have to be multiplied by 10^{-6}.

$(\cdot/\cdot)$ Report on the p-values (in %) of the DieboldMariano tests with the null that the particular RMSPE is significantly smaller than that of the Random Walk model (1st entry) and/or that of the Carlson–Parkin model (2nd entry).

• Indicates that test statistic is not available.

cantly better at the 5% significance level for only four cases. On the other hand, even at the 10% significance level, the CP method never significantly outperforms the random walk forecast.

The forecasting performance of the AR(p) models are given in column 3 (T1). The forecasts of the AR(p) models are based on recursive estimations. The favorite model specification is selected for every month through the Schwarz Bayesian Information Criterion among all models with a maximum lag length of $p=17$. In the notation of Eq. (T1), a lag length of $p=17$ corresponds to using 12 months of lagged observations. Furthermore, the same model selection strategy is applied to all models that include lagged variables ((T2),(T3),(T4),(P2),(P3),(P4)). At the 5% (10%) significance level, the AR(p) forecasts outperform the RW forecasts in 3 only (3) cases and the CP forecasts in 3 (5) cases. The three series where the time series model beats the RW and CP forecasts are the inflation rates. These series are the only macroeconomic time series under investigation and do not reveal martingale properties. This is why the time series approaches are particularly attractive. We do not find convincing evidence that the inclusion of forecasting errors obtained from a survey method (model (T2)) generates additional information for the AR(p) model. However, this combined approach beats the RW forecasts and the CP forecasts in 3 and 4 cases respectively at the 5% significance level. These 3 cases are again the inflation series. Even less successful is the model (T3) which uses dummy indicators for trend changes as additional explanatory variables. This model never beats the simple forecasting models. Introducing interaction terms between trend dummies and the sample probabilities for the direction of the process (model (T4)) sometimes results in improvements over the trend dummy model, but is by no means convincing if we compare its RMSPEs with the corresponding figures for the simple AR(p) variant.

The simplest version of the Pesaran-type model (model (P1)) is not superior to a simple random walk model, and outperforms the CP forecasts in only one case at the 5% level. Augmentation of the Pesaran approach (model (P2)) might be a reasonable strategy: it yields that are superior to the RW model in 3 cases (2 of them are inflation series), and to the CP model in 4 cases (at the 5% level). Finally, the inclusion of trend dummies (P3) and interaction terms (P4) does not really improve the forecasts. All in all, we conclude that time series methods produce the smallest RMSPEs for six time series. The Pesaran approaches do best in five cases, while the RW only proved to be the best approach for the Dow Jones Industrial series, and the CP method only produces the smallest RMSPE for the USD/EUR FX rate.

For financial time series (interest rates, stock indices and exchange rates) there is no forecasting method which significantly outperforms the RW, even at the 10% level. Thus the RW remains the ultimate benchmark in the forecasting of financial time series, because no richer specification was able to outperform its forecasts. Note that the superior performance (at the 5% level) of the time series models and the Pesaran models mainly relates to the three inflation series. However, the most successful forecasting models for the inflation series rely solely on time series information, which renders the survey information meaningless, even there. Nevertheless, for the inflation series, it might be a reasonable strategy to start from a (probably more elaborate) time series or Pesaran-type model, and to search for additional information from the survey data rather than relying on the survey data solely.

A more fundamental way of looking at the forecasting performance is to check for coincidence between the direction of the forecasts and the true direction of the time series. The results of our analysis based on the percentage of correct direction forecasts are given in Table 2. Whether the true and forecasted directions are independent is tested with a χ^2-independence test. The cell entries in parentheses are the p-values of these tests. For a random walk forecast (without drift), this measure is not meaningful since the predicted change is zero by definition.

For the inflation series, a pattern similar to the case of the RMSPE comparison emerges: the time series models and the augmented Pesaran model (P2) do not yield an independent (of the true direction at the 5% level) direction forecast, with percentages of correct direction forecasts being between 69.2% and 87.2%. Among the three short term interest rates, the Japanese interest rate is the only one for which we do not obtain independent (5% level) direction forecasts in two cases (CP: 23.0%, Model (P1): 30.7%). However, both models indicate the opposite direction. Judging

Table 2
Percentage of correct direction forecasts

Variable description	Carl.–Park.	Time series models				Pesaran type models			
		(T1)	(T2)	(T3)	(T4)	(P1)	(P2)	(P3)	(P4)
Inflation rate (Germany)	53.8 (64.8)	87.2 (0.0)	87.2 (0.0)			56.4 (38.2)	87.2 (0.0)		
Inflation rate (USA)	56.4 (32.7)	69.2 (1.3)	69.2 (1.3)			35.9 (8.9)	69.2 (0.9)		
Inflation rate (Japan)	23.0 (0.0)	82.1 (0.0)	82.1 (0.0)			41.0 (30.9)	74.4 (0.2)		
Short term interest rate (Germany)	61.5 (18.5)	76.9 (41.8)	76.9 (41.8)			61.5 (18.6)	69.7 (24.8)		
Short term interest rate (USA)	66.6 (24.7)	66.7 (29.0)	66.7 (29.0)			74.3 (12.3)	76.9 (44.2)		
Short term interest rate (Japan)	23.0 (1.0)	38.5 (67.2)	38.5 (67.2)			30.7 (1.3)	56.4 (8.6)		
DAX 30 index	35.8 (•)	69.2 (7.9)	69.2 (8.8)	33.3 (19.6)	33.3 (12.8)	33.3 (12.8)	38.5 (34.5)	43.6 (97.0)	38.5 (34.5)
Dow Jones industrial index	48.7 (•)	56.4 (43.1)	53.8 (58.9)	35.9 (3.1)	41.0 (24.3)	20.5 (0.0)	33.3 (3.8)	33.3 (3.8)	33.3 (3.8)
Nikkei 225 index	41.0 (•)	43.6 (72.5)	41.0 (1.5)	17.9 (0.0)	20.5 (0.0)	53.8 (49.2)	30.8 (1.8)	38.5 (21.7)	28.2 (0.6)
FTSE 100 index	38.4 (•)	53.8 (71.8)	64.1 (26.2)	20.5 (0.0)	35.9 (1.0)	35.8 (0.4)	48.7 (5.8)	43.6 (2.1)	35.9 (0.4)
CAC 40 index	33.3 (•)	61.5 (46.2)	71.8 (6.0)	28.2 (5.0)	38.5 (46.2)	33.3 (5.3)	61.5 (25.0)	59.0 (48.5)	56.4 (64.5)
FX-rate ($/€)	82.0 (•)	71.8 (9.5)	56.4 (8.4)	28.2 (92.9)	43.6 (29.7)	17.9 (•)	48.7 (5.2)	46.2 (6.7)	38.5 (13.8)
FX-rate (£/€)	64.1 (•)	48.7 (34.4)	46.2 (71.8)	33.3 (12.8)	59.0 (39.4)	64.1 (35.1)	56.4 (97.0)	46.4 (65.6)	69.2 (6.5)

The second cell-entry reports on the p-value (in %) of a χ^2-independence test of the null that the direction forecast is not independent of the true direction change. • Indicates that test statistic is not defined due to zero cells.

the direction forecasting ability of the remaining financial series is more complicated. At the 5% (10%) significance level 16 (26) of the 55 models do not yield an independent direction forecast. Among these 26 models, 20 indicate the opposite direction. Given this crude forecasting performance measure, the Pesaran type models and the time series models are of equal quality. Model augmentation does not seem to be a fruitful strategy since the simple variants of the two classes of models produce the highest number of correct direction forecasts for many series. Moreover, the CP method is also unconvincing if applied to time series with many changing points such as financial prices. Here the problem lies in the lack of correct anticipation of the change points (see Fig. 2).

In a final step, we perform the same analysis for the different groups of experts separately.[9] The main findings for the overall sample are confirmed by this more detailed analysis. We do not find any evidence that experts of specific sectors have access to superior information, e.g. there is no evidence that the forecasters of the banking sector or forecasters of insurance companies are more reliable than their colleagues from industry firms.

5. Conclusion

In this paper we analyze the forecasting performance of survey methods using qualitative information from experts' forecasts and comparing them with the performance of standard linear time series methods, as well as with simple random walk forecasts. The comparison of the forecasts relies on the Root Mean Squared Prediction Errors, which we find to be significantly different across models using the modified Diebold–Mariano test of Harvey et al. (1997). We also compare the models with respect to their ability to predict directions correctly on the basis of a χ^2-independence test. We derive three striking results: First, the ultimate benchmark model for financial time series remains the random walk. Its forecasts cannot be outperformed by forecasts from either quantification techniques or time series approaches. Second, for the

---

[9] The results of our analysis based on the forecasts of the experts from specific sectors can be obtained from the authors on request.

three inflation series (as representatives of macroeconomic aggregates) the forecasts by pure quantification techniques are inferior to simple time series forecasts and very often do not even outperform random walk forecasts. Third, the forecasts cannot be improved by methods relying on both time series and survey information. Pesaran-type models turn out to be slightly more successful than the Carlson–Parkin method. Nevertheless, even richer specifications of Pesaran-type models are not able to outperform simple time series approaches.

We do not find convincing evidence that forecasts based on the predictions of experts from specific groups are systematically superior compared to the ones from other groups or the random walk forecasts. In the light of these findings, maintaining the assumption that survey information, even if it comes from experts, contains valuable information is at least questionable. Since generalizing the aggregation methods does not appear to be a promising strategy for improving the forecasts, future research should concentrate on extracting the survey information optimally. One strategy could be to optimally exploit the information on the individual level. Valuable information could be obtained from specific subsamples (best forecasters), from information on individual uncertainty (no-change interval), or the use of other covariates in the survey. In addition, the aggregated forecasts may be helpful for obtaining more precise predictions of the series for different forecasting horizons.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, *96*, 42–55.

Anderson, O. (1951). Konjunkturtest und Statistik, Möglichkeiten und Grenzen einer Quantifizierung von Testergebnissen. *Allgemeines Statistisches Archiv*, *35*, 209–220.

Anderson, O. (1952). The business test of the IFO-Institute for Economic Research, Munich, and its theoretical model. *Revue de l'Institut International de Statistique*, *20*, 1–17.

Batchelor, R. A. (1982). Expectations, output and inflation, the European experience. *European Economic Review*, *17*, 1–25.

Batchelor, R. A., & Orr, A. B. (1988). Inflation expectations revisited. *Economica*, *55*, 317–331.

Carlson, J. A. (1977). Price expectations normally distributed? *Journal of the American Statistical Association*, *70*, 749–754.

Carlson, J. A., & Parkin, M. (1975). Inflation expectations. *Economica*, *42*, 123–138.

Cunnigham, A. W. F., Smith, R. J., & Weale, M. R. (1998). Measurement errors and data estimation: The quantification of survey data. In I. Begg, & S. G. B. Herny (Eds.), *Applied Economics and Public Policy* (pp. 41–58). Cambridge: Cambridge University Press.

Dasgupta, S., & Lahiri, K. (1992). A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data. *Journal of Business and Economic Statistics*, *10*, 391–400.

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

Fishe, R. P. H., & Idson, T. L. (1990). Information-induced heteroscedasticity in price expectations data. *Review of Economics and Statistics*, *72*, 304–312.

Fishe, R. P. H., & Lahiri, K. (1981). On the estimation of inflationary expectations from qualitative responses. *Journal of Econometrics*, *16*, 89–102.

Foster, J., & Gregory, M. (1977). Inflation expectations: The use of qualitative survey data. *Applied Economics*, *9*, 319–329.

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

Kaiser, U., & Spitz, A. (2000). *Quantification of qualitative data using ordered probit models with an application to a business survey in the German service sector*. ZEW Discussion Paper No. 00-58.

Knöbl, A. (1974). Price expectations and actual prive behaviour in Germany. *International Monetary Fund Staff Papers*, *21*, 83–100.

Lahiri, K., & Teigland, C. (1987). On the normality of probability distributions of inflation and GNP forecasts. *International Journal of Forecasting*, *3*, 269–279.

Maddala, G. S. (1990). Survey data on expectations: What have we learnt? In M. Nerlove (Ed.), *Issues in Contemporary Economics*, vol. 2. Macmillan.

Mitchell, J., Smith, R. J., & Weale, M. R. (2002). Quantification of qualitative firm level survey data. *The Economic Journal*, *112*, C117–C135.

Nerlove, M. (1983). Expectations, plans, and realizations in theory and practice. *Econometrica*, *51*, 1251–1279.

Pesaran, M. H. (1984). Expectations formations and macroeconometric modelling. In P. Malgrange, & P. -A. Muet (Eds.), *Contemporary macroeconomic modelling*. Oxford: Basil Blackwell.

Pesaran, M. H. (1985). Formation of inflation expectations in British manufacturing industries. *The Economic Journal, 95*, 948–975.

Pesaran, M. H. (1987). *The limits to rational expectations*. Oxford: Basil Blackwell.

Pflanzagl, J. (1952). Zur Methodik des Konjunkturtestverfahrens. *Statistische Vierteljahresschrift, 5*, 161–173.

Seitz, H. (1988). The estimation of inflation forecasts from business survey data. *Applied Economics, 20*, 427–438.

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

Theil, H. (1952). On the time shape of economic microvariables and the Munich business test. *Revue de l'Institut International de Statistique, 20*, 105–120.

Theil, H. (1958). *Economic forecasts and policy*. Amsterdam: North Holland.

Wren-Lewis, S. (1986). An econometric model of UK manufacturing employment using survey data on expected output. *Journal of Applied Econometrics, 1*, 297–316.