

Cueni, Reto; Frey, Bruno S.

Working Paper

Forecasts and Reactivity

CREMA Working Paper, No. 2014-10

Provided in Cooperation with:

CREMA - Center for Research in Economics, Management and the Arts, Zürich

Suggested Citation: Cueni, Reto; Frey, Bruno S. (2014) : Forecasts and Reactivity, CREMA Working Paper, No. 2014-10, Center for Research in Economics, Management and the Arts (CREMA), Zürich

This Version is available at:

<https://hdl.handle.net/10419/214556>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Center for Research in Economics, Management and the Arts

Forecasts and Reactivity

Working Paper No. 2014-10

CREMA Südstrasse 11 CH - 8008 Zürich www.crema-research.ch

Forecasts and Reactivity

by Reto Cueni* and Bruno S. Frey**

Version: March 21, 2014

Abstract:

Forecasters' estimates influence peoples' expectations, their decisions and thus also actual market outcomes. Such reactivity to forecasts induces externalities which harm the ex-post assessment of the forecasters' accuracy and in turn the improvement of forecasting accuracy and market outcomes. To empirically analyze the impact of reactivity on forecast accuracy, we compare the development of the forecast error in a non-reactive system, such as the market for weather forecasts, to reactive systems, such as the stock and the art market. The evidence strongly supports our proposition that in reactive systems, despite the vast increase of data and processing techniques, forecast accuracy did not improve, while forecast accuracy in non-reactive systems did.

(110 words)

We thank Christoph Engel, Jana Gallus, Kai Konrad, Siegwart Lindenberg, Margit Osterloh, Jean-Charles Rochet, Lasse Steiner, Christian Ulbrich Carl-Christian von Weizsäcker and participants at the 2013 ISNIE Conference in Florence, the 2013 Max-Planck Institute Munich Workshop and the Seminar 2013 at the Max-Planck Institute Bonn. Additionally, we would like to express our gratitude to Ludwig Zraggen, Christophe Voisard and Alexander Giordano from the Federal Office of Meteorology for retrieving and helping us to understand the meteorological data and to Jia Guo for generously sharing her Chinese art market data with us.

Authors: ***Reto Cueni** (corresponding author), reto.cueni@econ.uzh.ch, Department of Economics, University of Zurich. ****Bruno S. Frey**, bruno.frey@econ.uzh.ch, Zeppelin University and CREMA – Center for Research in Economics, Management and the Arts.

I. INTRODUCTION

Forecasts are ubiquitous. There is virtually no place in society in which forecasts do not play a role. Before taking a train, an application on our mobile phone tells us which train to choose in order to avoid overcrowded carriages. Future freshmen try to forecast the best college based on the latest undergraduate school rankings. Shareholders try to predict the highest prices to sell their shares, and governments issue predictions about the general economic outlook and take measures based on those predictions.

In all these cases, fundamental questions arise: Which forecast should we follow? Perhaps we follow the forecast provider that issued the most accurate forecast in the past and assume that the same provider will also be successful in the future. But what if most of the train passengers follow the same forecast because it was previously the most accurate and therefore take the same train? The carriages will be overcrowded. Consequently, the formerly best forecast may no longer be the most accurate one, even though it had been so *ex-ante*, i.e. before the forecast was publicly issued. In contrast, if all talented students apply to the university that has mistakenly been placed highest in a recent college ranking, this university will receive the most talented applicants. It will therefore be ranked number one in future rankings, though the *ex-ante* ranking was wrong.

In a market where participants react to forecasts, an *ex-post* analysis of the forecast accuracy is difficult. Given a high degree of *reactivity* in a market, individuals' reliance on the previous period's best-performing forecasters can create a self-fulfilling prophecy. The accuracy of a forecaster's past predictions provides no guarantee that future predictions will be correct. However, if many individuals believe the prediction of the best former forecaster to be the most correct one and act upon it, this makes the market move in this prediction's direction. Under such circumstances, it becomes rational to follow the forecaster with the most accurate past prediction, because this will again emerge as the most correct one due to the self-fulfilling prophecy induced by reactivity. In the train example, in contrast, a self-defeating prophecy may arise if a large majority follows the previously most accurate forecast provider.

Reactivity is not only of concern for the forecast users (i.e. market participants). The forecasters themselves are equally affected by reactivity. The actual outcome is

influenced, and this hampers both the ex-post calibration of their models and their learning in general. Reactivity in a market makes the distinction between a good and a bad forecast difficult and sometimes impossible to make. Although the forecasters and the market participants may act in a fully rational way, both are deceived by reactivity.

This paper pursues three aims: First, we intend to direct attention to the forecasting problems evoked by reactivity which cannot be solved by using more data or faster information technology. Modern economic forecasting literature tends to be enthusiastic about the usefulness of ever more data (big data) accessible due to new information technologies. It points toward an increasing ability to provide more accurate and timely forecasts in the near future (see, e.g., Elliott and Timmermann, 2008; Clements and Hendry, 2011: part II and III). We intend to show that this is not necessarily the case. We support our proposition with a brief review of the literature on how reactivity has been approached in the social sciences and discuss recent empirical studies and new developments in forecasting based on web search engine data like Google Trends.

Second, we point out two areas where the neglect of reactivity by scientists may lead to biased results. The examples considered are financial analysts' forecasts (see, e.g., Kothari, 2001; Clement and Tse, 2005; Beyer, 2008; Liu and Natarajan, 2012) and art auctioneers' presale price forecasts (see, e.g., Ashenfelter and Graddy 2003, 2006; Mei and Moses, 2005; Beggs and Graddy, 2009).

Finally, we empirically test our proposition that, despite the vast increase of data and processing techniques, forecast accuracy has not improved in reactive systems, while in non-reactive systems forecast accuracy has. We therefore exploit the difference in reactivity between reactive forecasts in financial and art markets and non-reactive weather forecasts. Such analysis of reactivity has, to the best of our knowledge, not yet been applied.

In the following section, we discuss the literature and background of reactive behavior. Section III examines externalities in reactive markets. Section IV provides an empirical analysis of forecast accuracy by comparing reactive and non-reactive markets. Section V scrutinizes the robustness of the results. The final section concludes.

II. BACKGROUND AND LITERATURE

This literature review focuses on how reactivity has been approached in the social sciences and in particular in economics. As early as 1866, John Venn explained the basic problem of reactivity: "... when the inference is about the conduct of human beings it is often forgotten that in the inference itself, if published, we may have produced an unsuspected source of disturbance" (Venn, 1866: 345).

Morgenstern (1928) adapted the problem of reactivity to economic forecasting and stated his impossibility theorem of a correct prediction in an economic context, where individuals react to predictions. Merton (1936, 1948) coined the concept of a self-fulfilling prophecy in a sociological context, also referring to economic case studies such as bank runs. He emphasized a type of reactivity that has a positive impact on the forecasters' accuracy if the market participants believe strongly enough in the forecast. This stands in contrast to Morgenstern's claims. According to Grunberg and Modigliani (1954), Morgenstern's impossibility theorem does not hold in their particular theoretical model. However, their assumptions have been taken to be partially unrealistic, even by themselves (see, e.g. Henshel, 1995). Machlup (1954) discussed the problems arising from reactivity through the forecasters' point of view. He emphasizes that a verification problem hinders learning by forecasters. Whenever a model is tested with new data, it remains unclear whether the deviations stem from the faultiness of the model or from other unpredictable changes in the modeled system, making thorough model corrections impossible.

Simon (1954) and Baumol (1957) showed that polls and other pre-election predictions may lead to both self-fulfilling prophecies and self-defeating prophecies. Both scholars were aware that it is questionable to assume that the reaction function of market participants (i.e. the voters) is known to the forecaster and is continuous. Kemp (1962) derived theoretically that correct public forecasting in economics (and other social applications) is impossible. However, Grunberg and Modigliani (1963) criticized his assumptions. Rothschild (1964) showed that forecasts, although imperfect, can have positive effects in dampening the fluctuations and thus improving the stability of economic systems.

Hensel (1995) and Hands (1990) claim that the idea of (fully) rational expectations (Muth 1961) almost ousted the idea of reactivity. However, even under the assumption of fully rational expectations, reactivity remained an issue. Expectations can be seen as forecasts; it is therefore important to understand to which market outcome fully rational forecasts would lead. Samuelson (1965) explored this problem, arguing that properly anticipated prices fluctuate randomly (i.e. follow a martingale process). His random walk hypothesis became influential especially in relation to financial markets (see also Malkiel, 1973; Samuelson, 1974). The hypothesis states that the reactivity between expectations (i.e. forecasts) and actions in a market leads to a randomly oscillating market outcome. Grossman (1977), among others, suggests that reactivity in markets produces noisy rational expectations, leading to informational externalities. The impossibility of informationally efficient markets can lead to markets which “rest in an equilibrium degree of disequilibrium” (Grossman and Stiglitz, 1980: 393). Lucas’s critique (1976), which is that it is naïve to try to forecast the outcome of a change in policy with a macroeconomic model based on aggregate data, also affects the relationship between reactivity and forecasting. A policy change leads to a change in the micro-foundation of the forecasting model, resulting in wrong policy advice. Campbell (1957) and Goodhart (1975) previously made a similar point, approaching reactivity from the market participants’ point of view, who are seen to produce the forecasted outcome: “A reactive measure is one which modifies the phenomenon under study, which changes the very thing that one is trying to measure” (Campbell, 1957:298).¹ As a consequence, the forecaster has to deal with a reactive outcome when trying to predict a new target.

Callon (1998), and later MacKenzie (2006) and Esposito (2013), applied the term “performativity” to the problem, which leads to the unpredictability of anthropogenic outcomes. MacKenzie and Esposito consider the effect on financial markets. They seek to show how the problem of performativity, especially for the field of risk management, has become unmanageable during the recent financial crisis. It is exactly the most basic assumption of risk management, “that risk can be predicted with probabilistic models” (Esposito 2013: 121), that is cast into doubt when considering the most recent financial crises (MacKenzie, 2006 or Millo and MacKenzie 2009). In particular, the risk models

¹ The idea is identical to the observer effect in physics (and indirectly related to Heisenberg’s (1927) uncertainty principle), stating that by measuring a certain object it is inevitable that the object itself will be affected and, hence, the result will deviate from the actual value

lead to a form of reactive behavior among the market participants (traders), who try to exploit the models (Beunza and Stark, 2012). Hence, the models become wrong at the very moment when at least some of their parameters become known in the market. We call such a mechanism *reactivity*.

The problem of reactivity involved in forecasts for anthropogenic systems has been largely neglected in the recent literature on economic forecasting (see, e.g., Lamont, 2002; Elliott and Timmermann 2008; Clement and Hendry 2011: part II and III).² Reactivity could be incorporated in the structural parameters of forecast models, which are of course discussed in the literature. However, neither the reactivity to forecasts nor possible ways of proxying for the degree of reactivity are dealt with (see also Hamilton, 2009; Vahey and Walkerley, 2013; Doovern et al., 2013). Instead, it is argued that real-time feedback will help the forecasters to learn more quickly and to adapt their models more appropriately (Elliott and Timmermann, 2008:4, Clements and Hendry, 2011: part II). In contrast, we argue that this is unlikely to be the case, because the market participants and their access to new technology (just as the forecasters' access) will enable them to react as fast as the forecasters can recalibrate their predictions.

When focusing on new trends in forecasting, we see a similar pattern predominating. A new forecast technique is introduced, and the enthusiasm for it induces people to neglect the imminent reactivity of the market participants, for example in the case of the proposal that web queries could be used to forecast human behavior (Goel et al., 2010; McLaren and Shanbhogue, 2011; Choi and Varian, 2012).³ A prominent example is the Google Flu trend forecast during the last winter season (2012/13), when Google's forecasts substantially overestimated the flu situation in the US (Butler 2013). The new forecasting technique that uses the flu-related Internet searches to locate flu pandemics (Eysenbach, 2006; Ginsberg et al., 2009) is certainly promising. However, the problem of reactivity has to be kept in mind. As argued in Butler (2013), the Google Flu forecast might have suffered from Internet searches looking for news about the flu situation. In addition, the flu hit early and extraordinarily severely last season and,

² Many studies have been conducted on learning and the generation of expectation in economic systems (see, e.g., Evans and Honkapohja 2001, for an overview concerning macroeconomics and Wenzelburger 2006, for an overview on mathematical economics). However, they do not study the impact of reactivity to the full extent or from the two sides of forecaster, and market participants, and their learning and adaptation.

³ See also Helbing and Balmietti (2011: 13) for a more critical evaluation of modern forecasting methods, new access to big data and how the "illusion of control over complex systems" may arise.

hence, seems to have excessively attracted the public's attention. This might have amplified the searches for preventive measures via the Internet from people who were not ill.

Such changes in people's search behavior might have led to a substantial forecast error in Google's Flu forecast (Butler 2013: 156). We assign this effect to the reactive behavior of individuals. But there might be another way in which reactivity can distort such a forecast. If the forecast predicts a severe flu, people could tend to mismatch the symptoms of a cold with that of a flu. This would lead to a biased result even if the individuals were not to change their searching behavior but, instead, just started an Internet search based on their wrong perception of having flu. Google Flu would be powerless to adapt to this direct form of reactivity, because the people are actually thinking that they caught the flu and interpret their symptoms this way, induced by the news about the now-casting based on the search queries. Such behavior is self-reinforcing and starkly weakens the forecasting accuracy, even though the company intends to adapt the search-based model to the changes in peoples' search behavior. Only a medical test of their symptoms can uncover and differentiate between actual flu symptoms and symptoms only perceived as evoked by a flu.

Another large strand of the literature dealing with a special kind of forecasters, stock analysts, has also neglected the problem of reactivity. Stock analysts try to forecast a firm's earnings per share over a certain period. Because these forecasts shape expectations on the stock market, the managers have an incentive to meet the earnings forecasted by the analysts. This leads to a problem of reactivity between the forecast and the actual outcome. Since the turn of the millennium, the accounting literature has identified the problem of a possible interaction between analysts' earnings per share forecasts and the company managers' incentive to converge their earnings towards the analysts' forecasts (see, e.g. Abarbanell and Lehavy, 2003; Burgstahler and Eames, 2006). Beyer (2008), to the best of our knowledge, is the first study theoretically analyzing reactivity between analysts' forecasts and managers' earnings management. Yet, the newer literature on analyst forecasts only partly considers the effects of reactivity when calculating analysts' forecast accuracy (e.g. Liu and Natarajan, 2012; Bissessur and Veenman, 2013).

Similarly, the literature on art auctions does generally not consider the effects of reactivity when treating experts' forecasts about the value of an auctioned artwork

(Ekelund et al., 1998 being a notable exception; for an overview of art auctions, see Ashenfelter and Graddy, 2003, 2006). Some individual studies recently analyzed auctioneers' presale price estimates (i.e., art experts' forecasts). These surveys tackle the problem of studying forecast accuracy when the forecast itself can influence the actual outcome but do not fully take into consideration the problem of reactivity (Mei and Moses, 2005; McAndrew et al., 2012, and partially also Beggs and Graddy, 2009, Ekelund et al. 2013).

III. POSITIVE AND NEGATIVE EXTERNALITIES OF REACTIVITY

Externalities produced by reactivity

From an economic and social point of view, it is important to analyze the possible positive and negative externalities of reactivity. Reactivity can take four possible courses, and the different outcomes these produce can be identified thus:

- (1) The forecast was ex-ante wrong, but the market participants reacted in such a way that the predicted value occurred. This leads ex-post to the judgment of a correct forecast – the typical case of a self-fulfilling prophecy.
- (2) The forecast was ex-ante correct, but the market participants reacted in such a way that the predicted outcome did not occur. This leads ex-post to the judgment of a wrong forecast – the typical case of a self-defeating prophecy.
- (3) The forecast was ex-ante correct, and the market participants reacted in such a way that the predicted value occurred. This leads ex-post to the judgment of a correct forecast.
- (4) The forecast was ex-ante wrong and the market participants reacted in such a way that the predicted outcome did not occur. This leads ex-post to the judgment of a wrong forecast.

In cases 1 and 2, reactivity exerts negative externalities on forecasters and market participants. First, we focus on the *market participants* who are misguided and who allocated resources wrongly. They detect their misbehavior at best when the reactive market outcome is corrected. An example for the first case is a bank run. A wrong forecast that a bank is not solvent can lead to the insolvency of a bank as it causes people to withdraw money from the bank. Any reactive behavior in accordance with the wrong forecast exerts a negative externality on the other market participants.

During a market bubble that emerges due to a wrongly exuberant forecast, the self-fulfilling prophecy can have positive externalities in the short-run. Temin and Voth (2004) provide a convincing example. During the South Sea Bubble, even fully rational investors informed about the existence of a bubble in the market and with full knowledge of the pessimistic forecasts still promoted the price exuberance. As long as such rational investors assume that the market sentiment is still increasing, they support the further inflation of the bubble, encouraging a self-fulfilling prophecy for a certain period. During this period, the self-fulfilling prophecy exerts a positive externality on the forecaster, who issues the wrong prediction, and the market participants, who believe in this wrong forecast, as long as there are enough other participants believing in it. However, in the long run, when the bubble bursts, the reactive behavior of market participants is unveiled. The negative externalities and the resulting inefficient resource allocation during the bubble become visible.

An example for case 2, a self-defeating prophecy, is a warning against an upcoming threat. Imagine the case of an unemployment forecast for the next year. The market participants (employees) are assumed to correctly infer from the forecast that they might lose their jobs and, hence, start to work harder and do not ask for higher wages. This reaction results in a self-defeating prophecy for the forecast of higher unemployment. The reactivity might even lead to an increasing pressure to perform among employees.⁴ The outcome, though, is desired, and the menace of unemployment is averted. Yet, ex-post the forecasters' prediction is judged negatively, providing a negative incentive for future forecasts of the same type although they are socially and economically desirable.

⁴ Lazear et al. (2013) provide evidence that during recessions the productivity of employees increases due to the "work-harder" effect and not due to other effects such as a better selection of workers.

Second, we focus on the *forecasters'* point of view. In case 1, the forecasters that initially issued a wrong forecast but benefit from the reactive outcome that meets their forecast are imposing negative externalities on the forecasters that issued initially correct forecasts. This is captured in case 2, which is the mirrored outcome of case 1, but from the point of view of the ex-ante correct forecasters. Both case 3 and case 4 involve no negative externalities. At first sight, the forecasters face only positive externalities, fostering correct forecasts and revealing the wrong ones.

However, cases 3 and 4 still pose a problem. As indicated above, case 3 is ex-post indistinguishable from case 1, and case 4 cannot be differentiated from case 2. This leads to negative externalities imposed by the reaction of market participants. For the market participants, both situations with indistinguishable cases make it impossible to select the truly able forecasters, as ex-ante wrong forecasts can be judged ex-post as correct and vice versa. Therefore, the market participants exert a negative externality on their own future behavior and prevent themselves from learning. The same applies for the forecasters. They will never reliably know if their own forecast was initially correct in cases 1 and 3 or was initially wrong in cases 2 and 4. Hence, the learning of market participants and forecasters in reactive markets is substantially restrained.

There is evidence that economic forecasts indeed did not improve for different periods and countries, which supports our analysis. Zarnowitz (1992), for example, shows that forecasts for the US GDP did not improve between 1953 and 1989. In addition, Heilemann and Stekler (2013) show unsteadiness in the accuracy of forecasts for German GDP growth between 1967 and 2001 and no overall trend of improvement. Despite the massive increase in available data and continuously improving information technologies, there is no discernible trend towards an improvement in economic forecasts (see also Vogel, 2007 or Oeller and Barot, 2000).

We are not aware of studies directly addressing the empirical question of the development of forecast accuracy of art experts or stock analysts over time. Consequently, we will focus on these fields in our empirical analysis.

IV. FORECAST ACCURACY OF MARKETS WITH AND WITHOUT REACTIVITY

To improve our knowledge about reactivity, we compare forecast markets for reactive and non-reactive systems. A prominent example of forecasts in a non-reactive system is that of *weather forecasts*, which we daily make use of. The weather does not react to the forecasts being issued. Self-fulfilling prophecy and self-defeating prophecy are equally barred in such systems. The absence of reactivity is expected to lead to a continuous improvement of forecast accuracy due to the learning of forecasters. The forecasters can, to a specific measurement error, fully rely, on the ex-post measured outcomes, adjust their models, and calibrate their calculations accordingly. The learning of forecasters will establish reliable practices and rules on how to produce correct forecasts.

Technological progress leads to an improvement of forecast accuracy in such non-reactive systems. We hypothesize that more data about the initial conditions and the dynamics of the weather and advances in the transmission, processing and storage of data lead to a steady increase of weather forecast accuracy over time. Several meteorological studies claim that weather forecasts did indeed improve over time (see Katz and Laso, 2011; American Meteorological Society, 2008; Buizza et al., 2010).

In contrast, we hypothesize that, in reactive markets, neither the advances in information technology and data processing nor the learning of forecasters have a positive long-term effect on the predictability of outcomes. Market participants in anthropogenic systems adapt to the new technology or the new data processing techniques. Such adaptation and learning of market participants impedes forecasters' learning and its positive effect on forecast accuracy. Our proposition is that, despite the positive technological development and learning of forecasters in all three markets, only the error of weather forecasts permanently and constantly decreases over time, while the forecast error in financial and art markets does not continuously improve.

To test this proposition, we rely on three data samples from the markets for weather, financial and art forecasts. All of them include forecasts and actual measured outcomes to calculate the forecasting quality in the form of a forecast error.

Weather data

To analyze weather forecasts, we use data from the Swiss National Weather Service (the Federal Office of Meteorology and Climatology MeteoSwiss). Each day at noon, MeteoSwiss issues its main weather forecast for the next day. To keep our analysis as lucid as possible, we focus just on forecasts for the maximal temperature of the following day in degree Kelvin. In Switzerland, as in most other countries, this temperature forecast is economically important for the prediction of future energy consumption. Utility companies base their expected demand for electricity to heat and cool offices, apartments, storage rooms, hospitals, and factories on such estimates.

MeteoSwiss also provided us with the actual outcomes for each day from January 1, 1999 until the last revision of the data system on December 3, 2010. We obtained data for six different regions in Switzerland, four in the Swiss midlands (Zurich, Basel, Bern, Geneva), one in the southern part of Switzerland (Lugano), and one in the Swiss mountain region (Sion). This covers most of the climatic and topological environments for weather forecasting in Switzerland. Data from July 2000 to July 2001, October to November 2001, and November to December 2007 could not be retrieved due to wrongly coded database entries in the original data set. The data consists of 22,454 observations from MeteoSwiss. Each observation includes a forecast produced on day t for the temperature maximum on the following day ($t+1$), named here as (*TempForecast*), as well as the actual measured maximum temperature of the following day (*TempOutcome*).

Our calculations focus on the forecast error that is most comparable among the different markets, measured as the absolute forecast error in percent of the actual outcome (*AFEP*). The *AFEP* of the temperature forecast is calculated as $Temp_AFEP = \text{abs}((TempForecast - TempOutcome)/(TempOutcome))$. The forecast errors for art data and financial data are calculated analogously.

In the case of weather forecasts, the error of the temperature forecast (*Temp_AFEP*) shows a mean error of 0.58% and a median of 0.45%. The minimum, naturally, lies at 0.00% and the maximum at 15.2%, with an interquartile range from

0.21% to 0.79%. The 99th percentile stands low at 2.21%, and the 1st percentile is still at 0.00%.⁵

Financial data

For the sample of forecasts and the actual outcomes in the financial market, we rely on the earnings per share forecasts of sell-side analysts for the companies of the US stock market (NASDAQ and NYSE). The data were retrieved from the Institutional Brokerage Estimation System (I/B/E/S) Detailed Earnings History Files. The database was accessed through the Wharton Research Data Services (WRDS) interface. This database provides the yearly earnings per share estimates of almost all analysts covering US stocks and working for a brokerage house in the US. The houses include both large financial services firms as Citigroup or Morgan Stanley which employ more than 100 analysts and small investment advisers with only one or two analysts. The earnings per share forecasts are estimated for a specific financial year of each company. We only retrieved observations for companies that have their financial year ending at December 31, which is the vast majority of US firms, and we only consider forecasts that were issued less than 12 months before the financial period ends. We focused on the same period as with the other data samples, namely on the years between 1999 and 2010. We discarded 32,265 (2.9%) observations that either have a missing forecast or a missing actual earnings per share entry in the database. The final data sample consists of 1,046,115 observations with earnings per share forecasts (*EPSForecast*) and actual earnings per share (*EPSOutcome*), including 763 brokerage houses employing more than 11,036 analysts covering 7,831 individual stocks.

We calculate the forecast error in the financial market identically to the temperature forecast as the absolute forecast error in percent of the actual earnings per share $EPS_AFEP = \text{abs}((EPSForecast - EPSOutcome) / (EPS Outcome))$. The *EPS_AFEP* has a corresponding median of 9.71%, a mean of 55.7%, and an interquartile range from 3.1% to 30%. It is important to bear in mind that the skewed distribution of the forecast errors strongly suggests the use of the median whenever the errors are later aggregated.

⁵ Please find a full set of descriptive statistics illustrated in Table A1 in the appendix of the paper.

Art data

For the data on art forecasts and the actual prices fetched at the auctions, we use data covering the vast majority of Chinese art auction sales between 1994 and 2011 and examines information on more than 735,593 individual lots from 7,143 auctions that took place in 426 auction houses mostly in China (including Hong Kong, Macau and Taiwan) but also in the UK and the USA. The data were obtained from www.artron.net, one of the largest databases covering Chinese artworks (Bai et al., 2013).⁶

The data cover classical Chinese paintings, calligraphy, and oil paintings. In the sample, we only retain lots which have an auctioneer presale price estimate (the forecast) as well as a hammer price (the actual outcome) for the artwork in the database. In order to arrive at the same period as for the weather or financial data, we focus on auctions between 1999 and 2010. The art data sample consists of 323,478 individual lots (i.e. observations) that took place in 262 auction houses, at 7,143 auctions and aggregate to a total sales turnover of more than RMB 57 billion.

The Chinese art market experienced a boom during our sample period and has developed into one of the economically most important art marketplaces in the world. In mainland China, the sales revenue rose from RMB 97 million in 2000 to RMB 32 billion in 2010 (equals more than USD 5 billion at 2013 values of RMB of 6.20 per USD). Art Market Trends, the annual art report 2011 from Artprice, identified Beijing as the global top marketplace for art in terms of revenue, accounting for more than 27% of the global art auction revenues (see also, Bai et al., 2013; Horowitz, 2011).

The auctioneers usually provide an upper and a lower bound estimate to give unskilled and inexperienced retail clients an estimated price range rather than a precise point estimate. We use the average estimate between the upper and the lower bounds as the point forecast (*PriceForecast*). This procedure is standard in the literature on art economics (see, e.g., Moses and Mei 2005, Ashenfelter and Graddy 2006, Ekelund et al. 2013). Each of the observations also include the actual price the art object fetched at the auction (*PriceOutcome*).

We calculate the forecast error in the art market identically to the temperature forecast and the earnings per share forecast as the absolute forecast error in percent of the actual price fetched ($Price_AFEP = \frac{abs((PriceForecast - PriceOutcome))}{PriceOutcome}$).

⁶ We thank Jia Guo for generously sharing her Chinese art data set with us.

Outcome)). The *Price_AFEF* is right-skewed, with a mean of 34.8% and a median of 25.8%. These numbers strongly suggest that we have to use the median to aggregate the forecast errors to receive a comparable measure of forecast accuracy for the different markets.

Empirical Findings

Figure 1 illustrates that the difference between the weather forecasts and actual weather, i.e. the forecast error, has significantly improved over the last years. The details of the fitted regression lines pictured in Figures 1 -3 are provided in Table 1 below.

[Figure 1 around here]

This result is contrasted with the forecasts in two reactive markets. Figure 2 shows the forecast error of earnings per share estimates for US stock markets. The figure clearly indicates that the forecasting accuracy did not improve over time despite the large increase in data availability and more efficient information processing.

[Figure 2 around here]

Figure 3 reports auctioneers' sale price estimates for the art market in China. The forecast illustrates that the forecasting accuracy did not improve between 1999 and 2010. The same is visible in the regression details in Table 1. The development of the yearly median *AFEF* is captured in the coefficient of the variable *Year* for each of the three forecast systems. Only in the weather system is the coefficient for *Year* negative (-0.000058) and highly significant at the 1% significance level, indicating a steady decrease in the forecast error over time. The coefficients for the anthropogenic systems of art (0.0061) and finance (0.0047) are both positive and not significant or only significant at the 10% level. These results show that the forecast accuracy in reactive

systems did not improve over time despite the massive increase in available data and processing techniques.

[Figure 3 around here]

[Table 1 around here]

It must be concluded that it is erroneous to believe that better data and processing techniques always lead to better forecasts. In systems with human actors, it is necessary to analyze their reactivity to the forecasts. In the extreme, such reactivity may even vitiate forecasts that were ex ante correct or confirm forecasts that were ex ante wrong. However, this insight has been obscured by our vastly improved technical capacities.

Comparing the production of a forecast for future weather conditions with that for future economic conditions, both seem to be of similar complexity. Naturally, the steadily improving weather forecast accuracy has drawn the attention of economic forecasters. Vahey and Wakerly (2013), in a working paper of the Bank for International Settlement (BIS), praise the probability forecasting in the meteorological sciences for attaining a similar forecast accuracy for macroeconomic forecasting.⁷ They propose that this technique be copied. However, the market participants in anthropogenic systems not only react to a specific forecast but also adapt to any new forecasting technique. The reactivity absorbs the improvement of forecast accuracy in anthropogenic systems such as financial or art price forecasting.

V. ROBUSTNESS

To provide additional evidence supporting the robustness of our empirical findings, a further time series analysis is used for each of the three markets. This procedure checks whether our analysis of the temporary changes in the aggregated forecast error of weather, financial and art forecasts has not fallen prey to a “spurious regression” problem (Granger and Newbold, 1974). Such a problem would result in a wrong measurement of the statistical significance in the OLS regressions due to a unit root in

⁷ Other examples exist. A recent paper by the Bank of England proposes the use of web queries in order to forecast upcoming unemployment and other economic trends (McLaren and Shanbhogue, 2011).

the time series. In such circumstances, the yearly median error of the weather forecasts (*TEMP_AFEP*) would not continuously improve as suggested by the OLS regression. Instead, it would show a spurious correlation between time and the decline of the forecast error.

A unit root test (Augmented Dickey-Fuller (ADF) test) is used for all three time series. Our proposition states that the error for weather forecasts (*TEMP_AFEP*) declines over time, while the errors for art (*PRICE_AFEP*) and financial forecasts (*EPS_AFEP*) do not. Accordingly, we expect the time series of the *TEMP_AFEP* to be negatively trend stationary, i.e. to have no unit root and to exhibit a clear negative trend, even though there might be periodical oscillations around the trend line. In contrast, the other two markets for art and financial forecasts are expected to show no trend stationarity.

The improved Akaike Information Criterion (AIC, see Ng and Perron, 2001) is employed to identify the number of lags for the test. For all three time series, the AIC suggests the use of 1 lag in the ADF test. The ADF test supports our first result. The weather forecast error (*TEMP_AFEP*) exhibits a highly significant trend-stationary decline of the error over time (ADF test statistic: -4.67; critical value on the 1% significance level: -4.380). For the time series of the forecast errors in the financial (*EPS_AFEP*) and the art markets (*PRICE_AFEP*), the null hypothesis of unit root cannot be rejected: the *EPS_AFEP* yields a test statistic of 0.063 and a critical value at the 10% significance level of -3.24; the *PRICE_AFEP* yields a test statistic of -2.23 and a critical value at the 10% level of significance of -3.26. Our finding of a significant decline in weather forecast errors and the findings of an absence of trend in the temporary change of the forecast errors in the art and financial markets are robust.

VII. CONCLUSION

This paper sheds light on the problems that publicly issued forecasts can create in reactive systems. Reactivity exists in every market, or, more broadly, in every anthropogenic system. While this issue has been recognized for a considerable time, the vast increase in available data and in the capacity to generate, transmit, store and process these data has led many scholars and laypeople to believe that forecasting in reactive market systems will become more accurate over time. We argue that this is

unlikely to materialize, because market participants will react to the forecasts and to the forecasters' new data collecting and processing techniques. Advances in forecast accuracy will thereby be absorbed.

Any forecast in anthropogenic systems induces positive and negative externalities of reactivity. Forecasting accuracy in non-reactive systems, such as weather forecasts, profits steadily from the new achievements in information technology. However, the use of meteorological models in reactive market systems does not provide a solution to achieve a higher forecast accuracy. We argue that forecasting for non-reactive markets is basically different from forecasting in reactive markets. We show that the forecast error in markets such as finance and art auctions is not reduced by new and more sophisticated information technology and more data. The difference in reactivity should also be taken into account when considering the use of forecasting tools and methods from the natural sciences in the social sciences.

While reactivity is a fundamental feature in anthropogenic systems, institutions can be created to mitigate its effect in those cases in which the aggregate effect is damaging. We are thinking of institutional measures to reduce reactivity of market participants. The ban on forecasts based on polls during and shortly before elections, which is existing in several developed countries like Canada or Italy, constitutes a simple example. On the other hand, one might think about the installation of a publicly funded forecast agency that would take the burden to issue potentially self-defeating prophecies or counteract potentially self-fulfilling prophecies in regard to socially important topics. The question how such institutions are to be designed is left to future research.

References

- Abarbanell, J., Lehavy, R. (2003). Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. *Journal of Accounting and Economics* 36(1), 105-146.
- American, Meteorological, Society, 2008. Enhancing Weather Information with Probability Forecasts. *Bulletin of the American Meteorological Society* 89: 1-12.
- Ashenfelter, O., 1989. How auctions work for wine and art. *The Journal of Economic Perspectives* 3: 23-36.
- Ashenfelter, O., Graddy, K., 2003. Auctions and the Price of Art. *Journal of Economic literature* 41: 763-787.
- Ashenfelter, O., Graddy, K., 2006. *Art auctions*. London: Elsevier.
- Bai, J., Guo, J., Mandel, B., 2013. Going global: markups and product quality in the Chinese art market. *FRB of New York Staff Report*.
- Bai, J., Philippon, T., Savov, A., 2012. Have financial markets become more informative? *FRB of New York Staff Report*.
- Baumol, W.J., 1957. Interactions between successive polling results and voting intentions. *Public Opinion Quarterly* 21: 318-323.
- Beggs, A., Graddy, K., 2009. Anchoring effects: Evidence from art auctions. *The American Economic Review* 99: 1027-1039.
- Beunza, D., Stark, D., 2012. From dissonance to resonance: cognitive interdependence in quantitative finance. *Economy and Society* 41: 383-417.
- Beyer, A., 2008. Financial analysts, forecast revisions and managers reporting behavior. *Journal of Accounting and Economics* 46: 334-348.
- Bissessur, S., Veenman, D., 2013. The Role of Ex-Ante Uncertainty in Explaining Why Firms Meet or Just Beat Analysts' Earnings Forecasts. *Working Paper Series INSEAD*
- Buizza, R., Hagedorn, R., Isaksen, L., 2010. *Recent changes of the ECMWF Ensemble Prediction System*. In: European Center for Medium-Range Weather Forecasts Conference Vitart F (Ed.). City.
- Burgstahler, D., Eames, M., 2006. Management of earnings and analysts' forecasts to achieve zero and small positive earnings surprises. *Journal of Business Finance & Accounting* 33: 633-652.
- Butler, D., 2013. When Google got flu wrong. *Nature* 494: 155.
- Callon, M., 1998. *The laws of the markets*. Oxford: Blackwell.
- Campbell, D.T., 1957. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54: 297-319.
- Choi, H., Varian, H., 2012. Predicting the present with google trends. *Economic Record* 88: 2-9.
- Clement, M.B., Tse, S.Y., 2005. Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance* 60: 307-341.

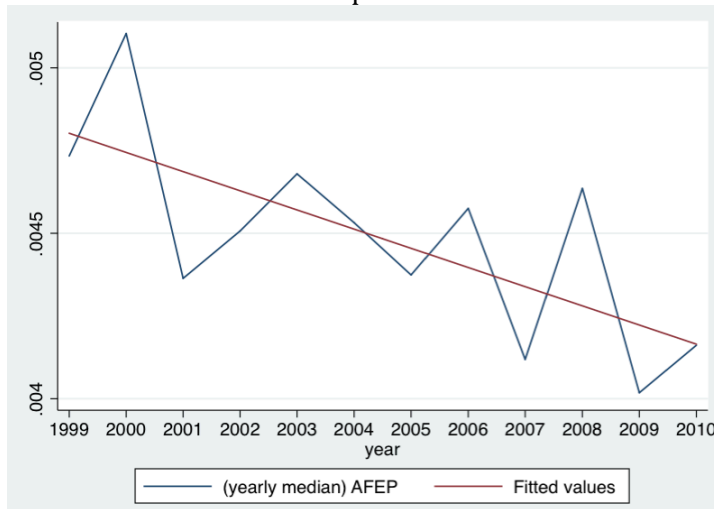
- Clements, M.P., Hendry, D.F., 2011. *The Oxford handbook of economic forecasting*:
- Dovern, J., Fritsche, U., Loungani, P., Tamirisa, N.T., 2013. Information Rigidities in Economic Growth Forecasts: Evidence from a Large International Panel. *IMF Working Paper* No.56.
- Ekelund Jr, R.B., Jackson, J. D., & Tollison, R.D., 2013. Are Art Auction Estimates Biased?. *Southern Economic Journal*, 80: 454-465.
- Ekelund Jr, R.B., Ressler, R.W., Watson, J.K., 1998. Estimates, bias and sales in Latin-American art auctions, 1977-1996. *Journal of Cultural Economics* 22: 33-42.
- Elliott, G., Timmermann, A., 2008. Economic forecasting. *Journal of Economic literature*: 3-56.
- Esposito, E., 2013. The structures of uncertainty: performativity and unpredictability in economic operations. *Economy and Society* 42: 102-129.
- Evans, G.W., Honkapohja, S., 2001. *Learning and expectations in macroeconomics*. Princeton: Princeton University Press.
- Eysenbach, G., 2006. *Infodemiology: tracking flu-related searches on the web for syndromic surveillance*. In: AMIA Annual Symposium Proceedings: 244. City: American Medical Informatics Association.
- Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. *Nature* 457: 1012-1014.
- Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, 107: 17486-17490.
- Goodhart, C., 1975. Problems of monetary management: the UK experience. *Papers in Monetary Economics (Reserve Bank of Australia)* 26: 219-245.
- Grossman, S.J., 1977. The existence of futures markets, noisy rational expectations and informational externalities. *The Review of Economic Studies* 44: 431-449.
- Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70: 393-408.
- Grunberg, E., Modigliani, F., 1954. The predictability of social events. *The Journal of Political Economy* 62: 465-478.
- Grunberg, E., Modigliani, F., 1963. Economic forecasting when the subject of the forecast is influenced by the forecast: comment. *The American Economic Review* 53: 734-737.
- Hamilton, James D., 2008. *Regime-switching models*. In: Durlauf, S.N., Blume, L.E. (eds) *The new palgrave dictionary of economics*. 2nd Ed. New York: Palgrave Macmillan.
- Hands, D.W., 1990. Grunberg and Modigliani, Public Predictions, and the New Classical Macroeconomics. *Research in the history of economic thought and methodology* 7: 207-223.
- Heilemann, U., Stekler, H.O., 2012. Has The Accuracy of Macroeconomic Forecasts for Germany Improved? *German Economic Review*
- Heisenberg, W., 1927. Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift fuer Physik* 43: 172-198.

- Helbing, D., Balmietti, S. 2011. From Social Data Mining To Forecasting Socio-Economic Crises. *European Physical Journal*, Special Topics 195:3-68.
- Henshel, R.L., 1995. The Grunberg/Modigliani and Simon possibility theorem: A social psychological critique. *Journal of Socio-Economics* 24: 501-520.
- Horowitz, N., 2011. *Art of the Deal: contemporary art in a global financial market*. Princeton: Princeton University Press.
- Katz, R., Lazo, J., 2011. *The economic value of weather and climate forecasts*. In: Clements, M.P. & Hendry, D.F. (eds.). *The Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.
- Kemp, M.C., 1962. Economic forecasting when the subject of the forecast is influenced by the forecast. *The American Economic Review* 52: 492-496.
- Kothari, S., 2001. Capital markets research in accounting. *Journal of Accounting and Economics* 31: 105-231.
- Lazear, E.P., Shaw, K., Stanton, C., 2013. Making Do With Less: Why Productivity Rises During Recessions. *NBER Working paper*
- Lamont, O., 2002. Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior and Organization* 48: 265-280.
- Liu, X.G., Natarajan, R., 2012. The Effect of Financial Analysts' Strategic Behavior on Analysts' Forecast Dispersion. *The Accounting Review* 87: 2123-2149.
- Lucas Jr, R.E., 1976. *Econometric policy evaluation: A critique*. In: Carnegie-Rochester conference series on public policy: 19-46. City: North-Holland.
- Machlup, F., 1955. The problem of verification in economics. *Southern Economic Journal* 1-21.
- MacKenzie, D.A., 2006. *An engine, not a camera: how financial models shape markets*. Boston: MIT Press.
- Malkiel, B.G., 1973. *A Random Walk Down Wall Street*. New York: Norton.
- McAndrew, C., Smith, J.L., Thompson, R., 2012. The impact of reserve prices on the perceived bias of expert appraisals of fine art. *Journal of Applied Econometrics* 27: 235-252.
- McLaren, N., Shanbhogue, R., 2011. Using internet search data as economic indicators. *Quarterly Bulletin of the Bank of England* (Q2): 134-140.
- Mei, J., Moses, M., 2005. Vested interest and biased price estimates: Evidence from an auction market. *The Journal of Finance* 60: 2409-2435.
- Merton, R.K., 1936. The unanticipated consequences of purposive social action. *American sociological review* 1: 894-904.
- Merton, R.K., 1948. The self-fulfilling prophecy. *The Antioch Review* 8: 193-210.
- Millo, Y., MacKenzie, D., 2009. The usefulness of inaccurate models: Towards an understanding of the emergence of financial risk management. *Accounting, Organization and Society* 34: 638-653.
- Morgenstern, O., 1928. *Wirtschaftsprognose*. Wien: Springer.

- Muth, J.F., 1961. Rational expectations and the theory of price movements. *Econometrica* 315-335.
- Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69: 1519-1554.
- Oeller, L.-E., Barot, B., 2000. The accuracy of European growth and inflation forecasts. *International Journal of Forecasting* 16: 293-315.
- Rothschild, K.W., 1964. Cobweb cycles and partially correct forecasting. *The Journal of Political Economy* 72: 300-305.
- Samuelson, P.A., 1965. Proof That Properly Anticipated Prices Fluctuate Randomly. *Management Review* 6
- Samuelson, P.A., 1974. Challenge to judgment. *The Journal of Portfolio Management* 1: 17-19.
- Simon, H.A., 1954. Bandwagon and underdog effects and the possibility of election predictions. *Public Opinion Quarterly* 18: 245-253.
- Temin, P., Voth, H.-J., 2004. Riding the south sea bubble. *American Economic Review* 94
- Vahey, S., Wakerly, L., 2013. Moving towards probability forecasting. *BIS Working Papers* No. 70.
- Venn, J., 1866. *Logic of Chance*. Lond and Cambridge: MacMillan and Co.
- Vogel, L., 2007. *How do the OECD Growth Projections for the G7 Economies Perform?: A Post-Mortem*. Paris: OECD Publishing.
- Wenzelburger, J., 2006. *Learning in economic systems with expectations feedback* Berlin: Springer.
- Zarnowitz, V., 1992. *Facts and Factors in the Modern Evolution of US Economic Fluctuations*. In: Zarnowitz V (ed.). *Business Cycles: Theory, History, Indicators, and Forecasting*. University of Chicago Press. 77-124.

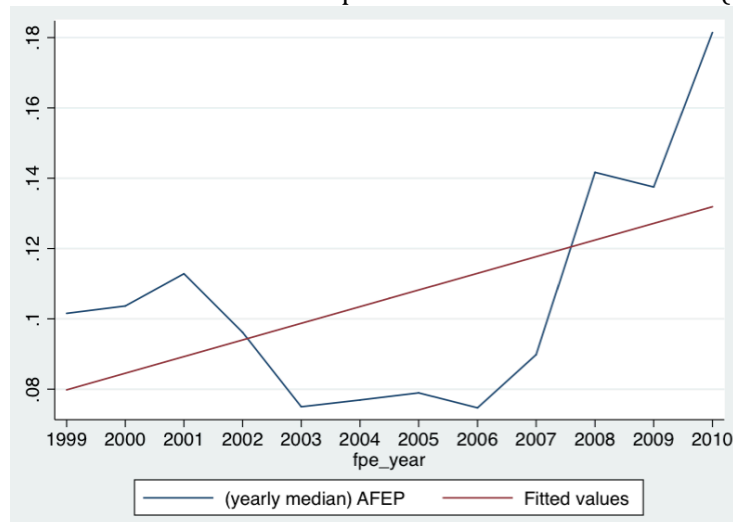
Figures

Figure 1: Weather forecast error (temperature forecasts), calculated as the yearly median of the absolute forecast error in percent of the actual outcome (AFEP)



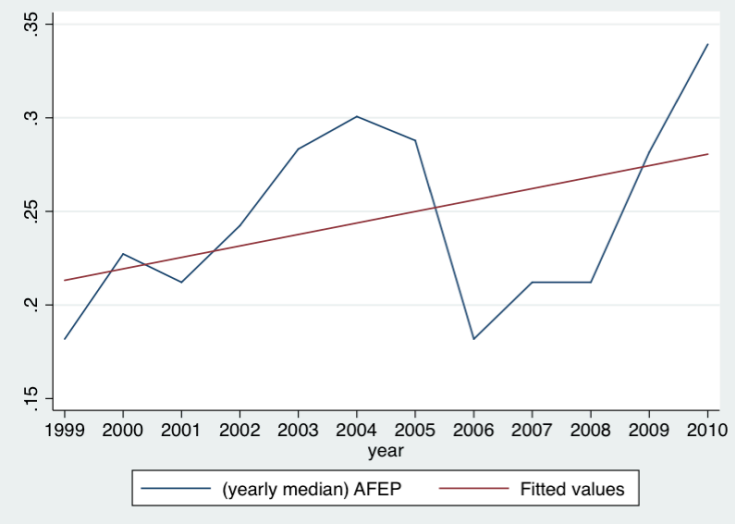
Note: The fitted line has a negative coefficient of -0.00006 , which is statistically significant at the 1% level (t-value= -3.04). Data Source: MeteoSwiss

Figure 2: Financial forecast error (earnings per share estimates), calculated as the yearly median of the absolute forecast error in percent of the actual outcome (AFEP)



Note: The fitted line has a statistically significant positive coefficient of 0.005 (t-value= 1.92)
Data Source: I/B/E/S

Figure 3: Art forecast error (sale price estimates), calculated as the yearly median of the absolute forecast error in percent of the actual outcome (AFEP)



Note: The fitted line has a statistically insignificant positive coefficient of 0.006 (t-value= 1.54) Data Source: www.artron.net

Tables

Table 1: Regression of Forecast Errors on Years

Dependent variable: <i>Yearly Median of AFEP</i>			
Variables	Weather	Art	Finance
<i>Year</i>	-0.000058*** (-3.04)	0.0061 (1.54)	0.0047* (1.92)
<i>Constant</i>	0.120645*** (23.82)	-12.038 (1.50)	-9.3873* (-1.90)
Observations	12	12	12
Adj. R2	0.43	0.11	0.19

Notes: The table reports OLS coefficient estimates, and t-values in parentheses. The models Weather, Art, Finance correspond to the respective data samples. The regression model is always the estimation of the *Yearly Median of AFEP* on *Year* and a *constant*. The dependent variable *Yearly Median of AFEP* is calculated as the median absolute forecast error in percentage of the actual outcome per calendar year of the respective sample for weather forecasts (temperature prediction), art forecasts (presale price estimate) or financial forecast (earnings per share estimate). Year is the 4-digit number of the respective year between 1999 and 2010. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Data sources: MeteoSwiss, I/B/E/S, www.artron.net

Appendix

Descriptive statistics

Table 1 provides the descriptive statistics of the forecasts, the actual outcomes and the forecast errors of the three forecast markets. Panel A shows the statistics for the weather data sample with 22,454 observations. On average, the daily maximum temperature measured in Switzerland is 288.7 in degree Kelvin (K), which corresponds to 15.6 degree Celsius (°C). The minimum of the daily maximum temperature between 1999 and 2010 in one of the six regions was 264.2K (-9.0°C) and the maximum was 314.2 K (41.0°C). The median maximum temperature was 289.2K (16.0°C).

[Table A1 about here]

The forecast error is measured in three different forms: (1) *Temp_FE* is the raw forecast error (*FE*) and is defined as $FE = \text{forecast} - \text{actual}$, i.e. $Temp_FE = TempForecast - TempOutcome$; (2) *Temp_AFE* is the absolute raw forecast error and is defined as $abs(Temp_FE)$; (3) *Temp_AFEP* is the absolute forecast error in percent of the actual outcome and is defined as $Temp_AFE/abs(TempOutcome)$. The forecast errors in panels B (Art data) and C (Financial data) are calculated in the same way.

The meteorologists on average underestimate the maximum temperature but still score a low mean raw forecast error (*Temp_FE*) of -0.3K (or °C) and a median raw forecast error of -0.4K. The absolute raw forecast error (*TEMP_AFE*) has a mean of 1.7K, a median of 1.3K, a natural minimum at 0K and a maximum of 41.5K. We focus on the forecast error that is most comparable among the different markets, measured in percent of the actual outcome (*AFEP*). The *Temp_AFEP* shows a mean error of 0.58% and a median of 0.45%. The minimum, naturally, lies at 0.00% and the maximum at 15.2% with an interquartile range from 0.21% up to 0.79%. The 99th percentile stands at low 2.21% and the 1st percentile still at 0.00%.

Panel B of Table 1 shows the financial data sample, containing earnings per share (*EPS*) forecasts of the US stock market with more than a million observations. The average *EPSForecast* of USD -124.25 and a median of only USD 1.20 indicates that the distributions are highly skewed. Both the *EPSForecast* and the *EPSOutcome* show extreme outliers in their distributions, with minima at USD -3,767,500 and -2,275,000,

respectively, and maxima at USD 180,000 and 178,000, respectively. However, the interquartile ranges start at USD 0.41 and 0.32 and end at USD 2.35 and 2.30 for the forecasts and the actual outcomes, respectively. The absolute forecast error in percent of the actual earnings per share (*EPS_AFEF*) has a corresponding median of 9.71%, a mean of 55.7%, and an interquartile range from 3.1% to 30%. It is important to bear in mind that the skewed distribution of the forecast errors strongly suggests the use of the median whenever the errors are aggregated.

Finally, the art data sample (Panel C) includes more than 320,000 observations. On average, the auctioneer presale forecast of the sale price (*PriceForecast*) is RMB 113,564 (USD 18,317) with a minimum of RMB 6 and a maximum of RMB 403 million. The auctioneers usually provide an upper and a lower bound estimate to give the unskilled and inexperienced retail client an estimated price range rather than a precise point estimate. We use the average estimate between the upper and the lower bounds as the point forecast (*PriceForecast*). This procedure is standard in the literature on art economics (see, e.g., Moses and Mei 2005, Ashenfelter and Graddy 2006). The median lies at around RMB 20,000, much lower than the mean, which suggests a highly right-skewed distribution.

The actual prices fetched (*PriceOutcome*) at the auctions show a similar picture. The mean price is RMB 176,877, the minimum price is RMB 100, and the maximum is RMB 169 million. The auctioneers on average underestimate the sales price of artworks, which can be seen in the distribution of the raw forecast errors *Price_FE* that have a negative mean (-63,313) and a negative median of -3,440. Focusing on the absolute forecast error as a share of the actual outcome, the *Price_AFEF* is right-skewed with a mean of 34.8% and a median of 25.8%. These numbers strongly suggest the use of the median as the aggregation instrument of choice when comparing forecast errors in the different markets.

Appendix

Table A1: Descriptive Statistics

Panel A: Weather Data (N=22,454)										
Variables	Mean	Min	1th	5th	25th	Median	75th	95th	99th	Max
<i>TempForecast</i>	288.4	264.2	271.2	274.2	281.2	289.2	295.2	302.2	306.2	314.2
<i>TempOutcome</i>	288.7	263.1	270.7	274.3	281.7	289.3	295.8	302.4	305.8	311.8
<i>Temp_FE</i>	-0.3	-14.2	-5.6	-3.8	-1.7	-0.4	0.9	3.0	5.1	41.5
<i>Temp_AFE</i>	1.7	0.0	0.0	0.1	0.6	1.3	2.3	4.3	6.3	41.5
<i>Temp_AFEP</i>	0.0058	0.0000	0.0000	0.0004	0.0021	0.0045	0.0079	0.0150	0.0221	0.1522
Panel B: Financial Data (N= 1,044,288)										
Variables	Mean	Min	1th	5th	25th	Median	75th	95th	99th	Max
<i>EPSForecast</i>	-124.25	-3767500.00	-9.50	-1.50	0.41	1.20	2.35	6.25	53.40	180000.00
<i>EPSOutcome</i>	-141.37	-2275000.00	-14.16	-2.15	0.32	1.15	2.30	6.14	54.00	178000.00
<i>EPS_FE</i>	17.12	-1492500.00	-3.35	-0.75	-0.10	0.00	0.17	1.60	10.50	2117500.00
<i>EPS_AFE</i>	51.08	0.00	0.00	0.01	0.04	0.12	0.39	2.35	16.00	2117500.00
<i>EPS_AFEP</i>	0.5574	0.0000	0.0000	0.0039	0.0308	0.0971	0.3005	1.7500	7.4625	4359.6230
Panel C: Art Data (N= 323,478)										
Variables	Mean	Min	1th	5th	25th	Median	75th	95th	99th	Max
<i>PriceForecast</i>	113,564	6	464	1,500	7,000	20,000	62,500	400,000	1,650,000	403,000,000
<i>PriceOutcome</i>	176,877	100	880	2,300	9,900	30,800	89,600	599,040	2,576,000	169,000,000
<i>Price_FE</i>	-63,313	-144,000,000	-1,060,000	-210,000	-21,800	-3,440	380	10,000	57,515	375,000,000
<i>Price_AFE</i>	73,613	0	0	200	1,600	6,200	25,840	220,049	1,086,390	375,000,000
<i>Price_AFEP</i>	0.3478	0.0000	0.0000	0.0182	0.1071	0.2584	0.5058	0.8140	0.9495	908.0909

Notes: The table reports summary statistics for the weather data sample (Panel A), the financial data sample (Panel B), and the art data sample (Panel C). In all three samples we retain only observations with non-missing forecast and actual outcome. The variable *TempForecast* is the forecast of the maximum temperature during the following day $t+1$ issued on day t measured in degree Kelvin ($K=^{\circ}C+273.15$). *TempOutcome* is the actual measured maximum temperature on day $t+1$ for which the forecast was issue in t . *Temp_FE* is the raw forecast error defined as $(TempForecast - TempOutcome)$. *Temp_AFE* is the raw absolut forecast error defined as the absolute value of *Temp_FE*. *Temp_AFEP* is the absolut forecast error in percentage of the outcome, defined as $(Temp_AFE/abs(TempOutcome))$. *PriceForecast* is the auctioneer presale price estimate of an artwork. *PriceOutcome* is the actual price fetched of the artwork at the respective auction. *Price_FE*, *Price_AFE* and *Price_AFEP* are defined analogously to *Temp_FE*, *Temp_AFE* and *Temp_AFEP*. *EPSForecast* is the earnings per share (EPS) forecasts issued by a sell-side analyst during year t targeting the EPS of a specific company j at the end of the financial year t . *EPSOutcome* is the actual EPS figure made public by company j for the financial year t . *EPS_FE*, *EPS_AFE* and *EPS_AFEP* are defined analogously to *Temp_FE*, *Temp_AFE* and *Temp_AFEP*.

Data Sources: MeteoSwiss, I/B/E/S, www.artron.net