Asymmetric Loss Functions and the Rationality of Expected Stock Returns

Aretz, Kevin and Bartram, Söhnke M. and Pope, Peter F.

Manchester University, Warwick University, City Univesity London

2011

Asymmetric Loss Functions and the Rationality of Expected Stock Returns

Kevin Aretz Söhnke M. Bartram Peter F. Pope*

Abstract

We combine the innovative approaches of Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007) with a block bootstrap to analyze whether asymmetric loss functions can rationalize the S&P 500 return expectations of individual forecasters from the Livingston Surveys. Although the rationality of these forecasts has often been rejected, earlier studies rely on the assumption that positive and negative forecast errors of identical magnitude are equally important to forecasters. Allowing for homogenous asymmetric loss, our evidence still strongly rejects forecast rationality. However, if we allow for variation in asymmetric loss functions across forecasters, we not only find significant differences in preferences, but we can also often no longer reject forecast rationality. Our conclusions raise serious doubts about the homogeneous expectations assumption often made in the theory of asset pricing, portfolio construction or corporate finance.

Keywords financial markets, general loss functions, GMM block bootstrapping, Livingston Survey, price forecasting

JEL Classification G11, G12, G15

This version April 29, 2009

*Address for correspondence: Kevin Aretz, Department of Accounting and Finance, Lancaster University, Lancaster LA1 4YX, UK, Tel: +44(0)1524 593 402, e-mail: <k.aretz@lancaster.ac.uk>. Söhnke M. Bartram's telephone number and e-mail address are +44(0)1524 592 083 and <s.m.bartram@lancaster.ac.uk>. Peter Pope's telephone number and e-mail address are +44(0)1524 593 978 and <p.pope@lancaster.ac.uk>. We would like to thank Mark Bradshaw, Gregory Connor, Petko Kalev, Stephen Taylor, Allan Timmermann, Joakim Westerholm and conference participants at the 2007 Annual Meeting of the Financial Management Association in Orlando, Florida for their insightful comments. We are indebted to David Peel for his assistance and advice.

1 Introduction

We combine two new approaches by Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007)[1] with a GMM block bootstrap to revisit the rationality of market return expectations obtained from survey data. Expectations of returns formed by rational agents play an important role in numerous studies in economics and finance, e.g., in studies in modern portfolio theory, asset pricing or corporate finance. However, many studies that check the general properties of return forecasts often raise doubts about whether these forecasts are rational (e.g., Pearce, 1984; Brown and Maital, 1981; Lakonishok, 1980).[2] If agents fail to form rational expectations, this could have important implications for financial decision-making. It would also leave academic studies, whose research methodologies implicitly or explicitly rely on rationality, misspecified.[3]

Existing tests of the rationality of stock return expectations are joint tests of rationality *and* preferences. More specifically, all existing studies implicitly base their tests on the assumption that preferences can be well described with mean-squared error (MSE) loss functions. Under MSE loss, the realized loss equals the squared deviation of the realized return from the prediction. Academic studies often assume MSE loss, as it is easy to study, and as its prediction errors have well-established properties. Notwithstanding these advantages, a number of studies question the assumption that MSE loss can generally be economically justified (Peel and Nobay, 2003; Granger and Newbold, 1986).

As Granger and Pesaran (2000) point out, the applicability of any loss function hinges on the purpose of the forecast. Return forecasts can be used by portfolio managers, e.g., when choosing optimal portfolios, security analysts, e.g., when making stock recommendations, or individual investors, e.g., when gauging the prospects of a stock market investment meant to cover a future capital expenditure, such as retirement or children's education. In each case, there could be reasons for why agents might be less averse to positive or negative forecast errors and might thus base their decisions on especially aggressive or conservative forecasts (Granger, 1999).[4] Hence, tests of rationality need to allow for the potential impact of asymmetries in agents' loss functions.

We base our rationality tests on the studies of Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007), which allow us to relax the commonly-made assumption of homogeneous MSE loss. The research design of Elliott, Komunjer, and Timmermann (2005) starts from a general parametric loss function, L, which can reduce to a number of well-known loss functions, including mean-absolute error (MAE) loss, MSE loss, lin-lin loss and quad-quad loss. The parametric loss function depends on two parameters, α and p, controlling the asymmetry of the function and the power to which the forecast error is raised, respectively. In contrast, the method proposed by Patton and Timmermann (2007) is based on an estimate of the first derivative of the loss function through a linear spline model. Using a sufficient number of nodes, their approach can capture a wide variety of loss functions including the class of loss functions used by Elliott, Komunjer, and Timmermann (2005). Moreover, the loss functions nested by the approach of Patton and Timmermann (2007) can also depend on state variables other than the forecast error (e.g., the state of the economy).

Under both methodologies, rational forecasters minimize expected loss with respect to the parameters underlying the forecast by setting a system of partial derivatives equal to zero. Conditional on some pre-estimation choices, e.g., the value of p or the number and values of the nodes, we choose the free parameters so as to make the system of partial derivatives hold as closely as possible. If agents are rational, the system of partial derivatives must hold exactly, which can be evaluated with a test of the overidentifying restrictions (J-test).

Although the approach of Patton and Timmermann (2007) nests a broader range of loss functions than that of Elliott, Komunjer, and Timmermann (2005), it requires large numbers of observations to yield good size and power properties. This justifies our implementation of both approaches.

We examine the rationality of individual market return forecasts obtained from survey data. When we pool data across forecasters assuming identical loss, we have to address the issue of autocorrelation and cross-sectional correlation in forecast errors. To this end, we base our inferences both on asymptotic theory and on the Hall and Horowitz (1996) GMM bootstrap with non-overlapping blocks.[5] When we examine rationality at the level of individual survey participants, inferences are based on asymptotic theory alone. The analysis of the individual survey responses has the advantage that we can not only test for rationality, but also for differences in loss functions across agents.

Our data are from the Livingston Surveys of Professional Forecasters. Although data on S&P 500 market index forecasts are available from 1952, we only analyze forecasts made between 1992 and 2008 for the following reasons. First, in 1992 the Livingston Surveys of Professional Forecasters were taken over by the Federal Reserve Bank of Philadelphia. The Federal Reserve Bank of Philadelphia introduced important changes to the conduct of the survey and the storage of survey responses.

Second, the analysis of long sample periods might suffer from changes in agents' loss functions over time which are due to idiosyncratic reasons (e.g., the age of the survey participant) and are therefore hard to model. Finally, empirical evidence reported in Goyal and Welch (2006, 2003) and Pesaran and Timmermann (1995) suggests that the ability to forecast market returns is low in our sample period. A constant forecast, which according to theory would be the unconditional expectation of the variable to forecast plus an optimal bias, should hence be fairly close to an apparently rational forecast. As a result, our outcomes should be biased towards the non-rejection of rationality.

Our empirical evidence based on the assumption of homogeneous preferences and constant (i.e., time-invariant) asymmetric loss is not consistent with forecast rationality. Although we obtain the intuitive finding that estimated loss functions emphasize negative forecast errors more strongly than positive forecast errors, there are large and persistent deviations from the system of optimality conditions. These deviations are highly significant according to both asymptotic theory and the bootstrap. The rejection of rationality can be attributed to two phenomena. First, a time-invariant asymmetric loss function cannot explain the initially low average return forecasts, which suggest an extreme aversion to losses, jointly with the large increase in average forecasts in the 2000-2004 period, especially not since most instruments at this time suggested that the stock market was overvalued (e.g., Welch, 1999). Second, a homogenous loss function cannot account for the large variation in return forecasts at each forecast production date.

To address the first problem, we rerun our tests on two subsamples formed according to whether prior year's S&P 500 return was above or below the sample median. We also allow the loss function to explicitly depend on the level of the S&P 500 market index at the forecast production date. Consistent with our expectations, survey participants are more loss averse at the beginning of the sample period than in the 2000-2004 period. However, although these remedies substantially decrease the magnitude of the J-tests, our outcomes still strongly reject forecast rationality. As a second step, we also allow loss functions to vary across forecasters. Our evidence from these tests suggests that there are significant cross-sectional differences in the skewness of loss functions. Adjusting for differences in preferences, the approach of Elliott, Komunjer, and Timmermann (2005) almost always fails to reject forecast rationality. Perhaps surprisingly, this is less often the case for the approach of Patton and Timmermann (2007). This is probably due to well-known size problems of this method. Overall, it seems safe to conclude that homogeneous return expectations are a myth in the real world.

Our findings have several interesting implications. From a theoretical point of view, they suggest that it might be important to incorporate heterogenous expected returns into studies in economics and finance. Although a limited number of studies have followed up on this idea (e.g., Grossman and Stiglitz, 1980; Admati, 1986), in their case heterogenous expectations are induced through different access to information. In contrast, in our case heterogeneous expected returns can arise through the different processes through which agents transform (potentially identical) information into forecasts. We are not aware of any prior studies which have examined the implications of cross-sectional differences in information processing approaches. From a practical point of view, our findings suggest that the return forecasts of different survey participants might be more or less useful to different forecast users, depending on the preferences of the forecaster and those of the user.

Our study is organized as follows. In Section 2, we describe the survey data and the instruments. Section 3 reviews existing empirical tests of rationality and then introduces the approaches of Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007). In Section 4, we present our empirical findings. Section 5 concludes. Appendix A and B contain details about a Monte Carlo analysis of the size of two hypothesis tests and the GMM block bootstrap, respectively.

2 Data

2.1 Livingston Survey Stock Return Expectations

We obtain data on S&P 500 price forecasts from the Livingston Surveys of Professional Forecasters. The Livingston Surveys were initiated in 1946 by Joseph Livingston, who used them as a basis for semi-annual newspaper articles on the state of the U.S. economy. After his death in 1990, the Livingston Surveys were taken over by the Federal Reserve Bank of Philadelphia, which introduced important changes to the survey design and the handling of data. Many of these changes were made to overcome previous well-known obstacles to the academic analysis of the survey responses. For our purpose, the three most important modifications are the following. First, the Federal Reserve Bank of Philadelphia abolished Joseph Livingston's practice of altering the survey responses on a subjective basis, e.g., for journalistic purposes. Second, when asking about market return expectations the Federal Reserve

Bank decided to consistently focus on the S&P 500 as a proxy for the U.S. stock market. Previously, the stock market proxy frequently changed, as can be seen on the Federal Reserve Bank's website.

Arguably the most important change the Federal Reserve Bank initiated was to ask survey participants about their expectations of the variables to forecast on the publication date of the survey. Although the Livingston Surveys are always published in June and December, survey questionnaires are mailed to participants in May and November, respectively. As a result, existing studies had to make an adhoc assumption about the date on which survey participants filled in the questionnaires, in order to construct expected growth rates from the level data (see, e.g., Pearce, 1984; Lakonishok, 1980). This approach could be problematic, especially when forecasting highly volatile variables. In fact, specifying one single date on which, by assumption, all agents made their predictions overstates some return expectations, while it understates others. Using the new forecasts, this complication can in principle be avoided by computing expected growth rates starting from the survey publication date. In contrast to other studies, we thus compute annualized expected market returns through:

$$SPIER_{i,t}^{f} = ((SPIF_{i,t}^{f}/SPIF_{i,t}^{pub})^{12/f} - 1.0)100, \tag{1}$$

where $SPIER_{i,t}^{f}$ is the annualized S&P 500 return forecast of individual i at time t for the period starting from the survey publication date ($> t$) to f months ahead in the future. Similarly, $SPIF_{i,t}^{f}$ is the S&P 500 stock price forecast for f months ahead in the future and $SPIF_{i,t}^{pub}$ is the stock price forecast for the survey's publication date. We use the 6-month, 12-month, 1-$\frac{1}{2}$ year, or 2-year ahead stock price forecasts to construct $SPIER_{i,t}^{f}$. Forecast errors are defined as the difference between the annualized realized and the forecasted S&P 500 return.

As we analyze individual survey responses, we have to deal with the problem that only a minority of forecasters participate in all surveys on a year-by-year basis, which could undermine inferences in the presence of dependent error terms. Time-series correlation in the error term necessarily arises from the fact that forecasting periods overlap by some months. More precisely, when survey participants form market return expectations in May or November, they are still unaware of the accuracy of their prior forecasts. Cross-sectional correlation in the error term could stem from the use of similar information or from a similar susceptibility to trends in stock market beliefs. As a result, we follow two approaches. First, we form an unbalanced panel out of the available forecasts through always choosing the longest uninterrupted sequence of forecasts by an individual survey participant. Second, we form an unbalanced panel through choosing all available forecasts. As both approaches lead to similar conclusions, we later only report the empirical findings based on the second approach.

It might be interesting to benchmark our S&P 500 return forecasts against those used in other empirical studies. To this end, we compare our consensus forecasts with those of Brav, Lehavy, and Michaely (2005), who imply expected market returns from Value Line target prices. Other studies, e.g., Michaely and Womack (1999) and Rajan and Servaes (1997), offer evidence that Value Line target prices are less subject to over-optimism or conflicts of interest and might therefore accurately reflect expectations. The finding that our consensus expected returns are similar in nature to those in Brav, Lehavy, and Michaely (2005) is thus encouraging news for the quality of our data.

2.2 Information Variables

We use a comprehensive set of instrumental variables to approximate all information available to survey participants at the time they made their forecasts. When we rely on the approach of Elliott, Komunjer, and Timmermann (2005), this set consists of the return spread between a 3-month government bill and a 1-month government bill portfolio, the dividend yield of the S&P 500 stock index, the yield spread between an Aaa- and a Baa-rated corporate bond portfolio, the yield spread between a 10-year and a 1-year government bond portfolio and the yield on a 3-month government bill. Together with a constant, we therefore obtain a total of six instruments. As the approach of Patton and Timmermann (2007) often requires more instruments, we separately include the returns of the 1-month and 3-month government bill portfolios and the yields of the Aaa- and Baa-rated corporate bond portfolios and the 10-year and 1-year government bond portfolios in addition to the other three instruments. Together with a constant, this choice gives us a total of nine instruments. A large collection of studies reveals the ability of these instruments to model variation in market returns.[6]

The earliest possible time at which survey participants can fill in the survey questionnaires is at the beginning of May or November. As a result, we lag the instruments by two months to make sure that they were known to survey participants. Although this implies that some forecasters might have used more up-to-date information, most instruments move only slowly over time, which should alleviate this problem. Lagging the instruments by only one month does not change our findings.

We obtain the 1-month and 3-month U.S. Treasury Bill return from CRSP. The S&P 500 dividend yield can be found on Robert Shiller's website.[7] The yields on Aaa- and Baa-rated corporate bond portfolios, on 10-year and 1-year government bond portfolios and on 3-month government bills are from the website of the Board of Governors of the Federal Reserve System.

3 Basic Concepts and Methodology

3.1 Traditional Rationality Tests and their Weaknesses

Earlier studies test the rationality of market return forecasts almost exclusively based on OLS regression methods. In this section, we shortly review why OLS methods, which most often take the form of an unbiasedness test and a more powerful orthogonality test, implicitly rely on the assumption of MSE loss. Moreover, we will also see that the orthogonality test is a restricted version of the more general methodologies proposed by Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007). We use the following notation: A_{t+f} is the realized value of the object to forecast at time $t+f$, P_t^{t+f} is the prediction of the object to forecast made at time t for time $t+f$, where f is the forecast horizon, and Ω_t is the information set available at time t.

Existing studies conclude that the expectation P_t^{t+f} is optimal, if

$$P_t^{t+f} = \mathrm{E}[A_{t+f}|\Omega_t], \tag{2}$$

since "no other unbiased predictor has smaller variance" (Brown and Maital, 1981). In other words, the estimator in equation [2] minimizes:

$$\mathrm{var}[A_{t+f} - P_t^{t+f}|\Omega_t], \qquad \text{s.t. } \mathrm{E}[A_{t+f} - P_t^{t+f}|\Omega_t] = 0. \tag{3}$$

Substituting the constraint into the objective function, the objective function can be rewritten as $\mathrm{E}[(A_{t+f} - P_t^{t+f})^2|\Omega_t]$, which is equivalent to expected MSE loss. As a result, the predictor shown in equation [2] is optimal under MSE loss. However, it is not necessarily optimal under alternative loss functions. For example, if agents' preferences follow an MAE loss function, the optimal predictor becomes the conditional median of the variable to forecast (see Appendix A in Gu and Wu, 2003).

Similarly, if agents' preferences are skewed towards positive or negative forecast errors, the optimal forecast can be biased (Elliott and Timmermann, 2004).

Under MSE loss, we can perform a weak rationality test through an OLS regression which checks the unbiasedness of the forecast:

$$A_{t+f} = \alpha + \beta P_t^{t+f} + u_{t+f}. \tag{4}$$

If the forecast is unbiased, we should be unable to reject the joint hypothesis of $\alpha = 0$ and $\beta = 1$. To see that an unbiasedness test constitutes a weak test of rationality, it should be noted that P_t^{t+f} can always be written as a combination of variables in the available information set. As a result, P_t^{t+f} must also be in the information set. In the absence of other available information, we can thus rewrite equation [2] as $P_t^{t+f} = \mathrm{E}[A_{t+f}|P_t^{t+f}]$, which holds if $\alpha = 0$ and $\beta = 1$.

A stronger test of rationality can be constructed through decomposing the forecast error into its conditional expectation and an orthogonal error term:

$$(A_{t+f} - P_t^{t+f}) = \mathrm{E}[(A_{t+f} - P_t^{t+f})|\Omega_t] + u_{t+f}. \tag{5}$$

Under equation [2], the conditional expectation of $A_{t+f} - P_t^{t+f}$ must be zero. As the conditional expectation can be any combination of variables in the information set, OLS regressions of the forecast error on all variables known at the forecast production date should have no explanatory power. Hence, the adjusted R^2 of this regression should be close to zero. However, as with the unbiasedness test, we can only reach this conclusion if we assume that optimal forecasts are equal to the conditional expectation of the variable to forecast. Under non-MSE preferences, this assumption might not hold and test outcomes might therefore be unreliable.[8]

3.2 More General Tests of Rationality

In this study, we use the approaches of Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2005) to test for forecast rationality. Both approaches allow for a wider range of preferences than traditional OLS regression tests. More specifically, the methodology of Elliott, Komunjer, and Timmermann (2005) assumes that agents' preferences can be described with the following general loss function, which we denote by L and which depends on two parameters:

$$L(p, \alpha) = [\alpha + (1 - 2\alpha) \cdot 1(A_{t+f} - P_t^{t+f} < 0)]|A_{t+f} - P_t^{t+f}|^p, \tag{6}$$

where α, which must lie between zero and one, controls the asymmetry of the loss function and p is the power to which the absolute value of the forecast error is raised. If $\alpha = 0.5$ and $p = 1$ or 2, we obtain MAE or MSE loss, respectively. In contrast, if $\alpha \neq 0.5$ and $p = 1$ or 2, then we obtain the asymmetric counterparts of the MAE and the MSE loss functions, which are the lin-lin and the quad-quad loss functions, respectively.

In this setting, rational forecasters minimize the conditional expectation of the loss function with respect to the parameters underlying the forecast (P_t^{t+f}), which yields a system of optimality conditions. The rationality test proposed by Elliott, Komunjer, and Timmermann (2005) simply suggests to evaluate whether the optimality conditions are jointly equal to zero. For this test to be feasible, optimal forecasts must be derived from a forecasting model which is at least partially linear. Formally speaking, $P_t^{t+f} = \theta' W_t + [\ldots]$, where vector θ contains the choice parameters, vector W_t contains the realizations of the information variables included in the information set Ω_t and $[\ldots]$ is the non-linear part of the forecasting model. In this case, the optimality conditions associated with the linear part of the forecast do not depend on the unknown choice parameters and can be written as:

$$E[W_t(1(A_{t+f} - P_t^{t+f} < 0) - \alpha)|A_{t+f} - P_t^{t+f}|^{p-1}] = 0. \tag{7}$$

Conditional on knowledge of p and P_t^{t+f}, Elliott, Komunjer, and Timmermann (2005) show that α can be estimated from the sample counterpart of equation [7] under fairly mild conditions. Perhaps surprisingly, we can obtain a consistent estimate of α even without identifying all information variables used by forecasters. This is a powerful result, as it is probably the case that forecasters employ both private and public information sources when forming their expectations. Not only is data on public information sufficient to consistently estimate α, their theoretical outcome also implies that we can estimate loss functions even if the forecasting model is partially non-linear.

The population value of α is the unique minimum of the quadratic form:

$$E[g(\alpha, p, \text{data})]' S^{-1} E[g(\alpha, p, \text{data})], \tag{8}$$

where the vector $E[g(\alpha, p, \text{data})]$ contains the moment conditions shown in equation [7] and S is any positive definite weighting matrix. Taking the derivative and solving for α, we obtain:

$$\alpha = \frac{E[V_t|A_{t+f} - P_t^{t+f}|^{p-1}]' S^{-1} E[V_t 1(A_{t+f} - P_t^{t+f} < 0)|A_{t+f} - P_t^{t+f}|^{p-1}]}{E[V_t|A_{t+f} - P_t^{t+f}|^{p-1}]' S^{-1} E[V_t|A_{t+f} - P_t^{t+f}|^{p-1}]}. \tag{9}$$

We can minimize the asymptotic standard error of α through setting S equal to the covariance matrix of the moment conditions. However, since this covariance matrix depends on α, estimation is done iteratively in practice. We start by setting S equal to an identity matrix to compute an initial estimate of α. The initial α estimate is subsequently used to construct an initial approximation of the covariance matrix of the moment conditions. We compute S through the positive definite matrix of Newey and West (1987), with l, the number of lags, set equal to four and w, the weighting, set equal to one. We then employ our estimate of S to update α, which is afterwards used to update S. We stop this procedure when the deviation between two consecutive estimates is smaller than 0.001.

Under stronger mixing conditions, it can be shown that α follows a normal distribution in the asymptotic limit. If α is in the interior of the parameter space (i.e., between zero and one), we can write the asymptotic variance of $\sqrt{T}(\alpha - \alpha_0)$ as:

$$[E[V_t|A_{t+f} - P_t^{t+f}|^{p-1}]' S^{-1} E[V_t|A_{t+f} - P_t^{t+f}|^{p-1}]]^{-1}, \tag{10}$$

where α_0 is the true value of α, T is the number of observations and the term in the expectation operator is the vector of the derivatives of the moment conditions with respect to α.

In contrast, if the true value of α is on the boundary of the parameter space, then we simulate the asymptotic distribution of the α estimate following an approach proposed by Andrews (2002).[9] In his study, Andrews (2002) shows that, in the asymptotic limit, the GMM criterion function is equal to a quadratic function (plus a term unrelated to parameters). Using a concept developed by Chernoff (1954), he then approximates the parameter space through a convex cone. Simulating the quadratic function and minimizing its value over the convex cone allows us to construct the asymptotic distribution of the parameter estimate. More details and a brief Monte-Carlo analysis of the size properties of this approach are provided in Appendix A.

We can test the hypothesis that the optimality conditions are jointly equal to zero with:

$$J(\hat{\alpha}) = T \cdot E[V_t(1(A_{t+f} - P_t^{t+f} < 0) - \hat{\alpha})|A_{t+f} - P_t^{t+f}|^{p-1}]' S^{-1} \\ E[V_t(1(A_{t+f} - P_t^{t+f} < 0) - \hat{\alpha})|A_{t+f} - P_t^{t+f}|^{p-1}], \tag{11}$$

which is often referred to as a J-test. Under the null hypothesis, the test statistic follows a chi-square distribution with degrees of freedom equal to the number of instruments.

Consistent with the former approach, the methodology of Patton and Timmermann (2007) also relies on a joint test of whether the partial derivatives of expected loss with respect to the parameters underlying the forecasting model are equal to zero, i.e., whether $E[W_t \cdot \partial L/\partial P_t^{t+f}] = 0$. However, in their study they approximate the partial derivative based on a linear spline model. To see how this works, define $\lambda(A_{t+f} - P_t^{t+f}, \text{other}; \theta) = \partial L/\partial P_t^{t+f}$, where $\theta = [\gamma_1, \gamma_2, \ldots, \gamma_{K+1}]'$. Let $\zeta_1, \ldots, \zeta_K$ be the nodes of the spline, of which one equals zero. Our notation reveals that, while the loss function depends explicitly on the forecast error, it can also depend on other state variables.

We use exactly the same empirical specification as Patton and Timmermann (2007) to model the derivative of λ, which is:

$$\frac{\partial \lambda(e, \text{other}; \theta)}{\partial e} = \begin{cases} \gamma_1 \equiv \Gamma(\varphi_{01} + \varphi_{11}\text{other} - \ln K) & \text{for } e \leq \zeta_1 \\ \gamma_i \equiv (1 - \sum_{j=1}^{i-1} \gamma_j)\Gamma(\varphi_{0i} + \varphi_{1i}\text{other} - \ln K) & \text{for } \zeta_{i-1} < e \leq \zeta_i, i = 2, \ldots, K \\ \gamma_{K+1} = 1 - \sum_{j=1}^{K} \gamma_j & \text{for } e > \zeta_K \end{cases} \tag{12}$$

where, for notational convenience, $e = A_{t+f} - P_t^{t+f}$ and $\Gamma(x) = (1 + \exp(-x))^{-1}$. Our specification has several advantages. First, it ensures that the value of the loss function is weakly increasing in the absolute value of the forecast error. Second, it imposes the restriction that the sum of the slopes is equal to unity, which is important, as $\lambda(e, \text{other}; \theta)$ is only identified up to a multiplicative constant. Under the assumptions that $\lambda(0, \text{other}; \theta) = L(0, \text{other}; \theta) = 0$, we can construct both the derivative of the loss function with respect to the forecast error and the loss function from the vector θ.

The values of the free parameters can be estimated with the GMM. In particular, the estimate of θ minimizes the value of $g_T(\theta)'Wg_T(\theta)$, where $g_T(\theta)$ is the sample counterpart of the optimality conditions, i.e., $E[W_t \cdot \partial L/\partial P_t^{t+f}] = 0$, and W is any consistent weighting matrix. We set W equal to the inverse of the covariance matrix of the moment conditions, which we compute from a first-stage estimation with $W = I$ using the approach of Newey and West (1987). If the number of moment conditions is larger than the number of free parameters, we can use a J-test to evaluate whether the optimality conditions are jointly equal to zero. Formally:

$$Tg_T(\theta)'Wg_T(\theta) \rightarrow \chi^2_{\text{\# moment conditions - \# parameters}}, \tag{13}$$

where W must be a consistent estimate of the optimal weighting matrix, i.e., the inverse of the covariance matrix of the moment conditions.

In our implementation of this approach, we choose a total of three nodes, whose values are -20, 0 and 20. The values of the nodes correspond very roughly to the 20th, 50th and 80th percentile of the distribution of the forecast error. We have checked that other choices for the number of nodes and their values do not materially affect our findings. In our initial tests, we specify that the loss function only depends on the forecast error. In this case, we estimate three parameters with nine moment conditions. We later also allow the loss function to depend on the level of the S&P 500 market index at the forecast production date, i.e., 2 months before the publication of the Livingston Surveys. At the peak of an economic expansion, survey participants might believe that the economy has 'overheated' and could therefore fall into a recession. If loss aversion is negatively related to the state of the economy, survey participants might produce especially conservative forecasts in such a situation. In the second setting, we estimate six parameters with nine moment conditions.

3.3 Bootstrap Inferences

In this study, we focus on the forecasts of individual survey participants. As the pooled forecast data should therefore be highly dependent, the application of asymptotic theory will most probably distort inferences. To alleviate these concerns, we hence also report inferences based on the Hall and Horowitz (1996) GMM block bootstrap. As Hall and Horowitz (1996) illustrate, the GMM block bootstrap can account for the dependence in survey responses.

We implement the GMM bootstrap as follows. We treat the sample data as the population. We then draw random bootstrap samples with replacement from this population. For the short-term forecasts (i.e., the bi-annual 6-month and 12-month survey responses), the length of each block equals 20 observations. For the long-term forecasts (i.e., the annual 18-month and 24-month survey responses), the length of each block equals 12 observations.[10] Using the bootstrap samples, we construct empirical distributions for α and the J-test. As a final step, we use the empirical distributions to compute the probabilities that the realizations of α and the J-test are more extreme than the whole-sample values of α and the J-test. Although the idea of the bootstrap method is straightforward, certain technicalities have to be taken into account. Appendix B gives more details about the Hall and Horowitz (1996) GMM block bootstrap and these technicalities. Evidence in the literature, e.g., Inkmann (2005) or Anatolyev (2002), suggests that, although bootstrapping cannot completely eliminate finite sample distortions resulting from the usage of asymptotic critical values, bootstrap critical values are usually more accurate than those obtained from asymptotic theory.

4 Empirical Findings

4.1 Summary Statistics and Graphical Analysis

Table 1 shows the realized S&P 500 return, the number of forecasters in each Livingston Survey, and the mean and standard deviation of the S&P 500 return forecasts over the sample period. Interestingly, the number of survey participants decreases markedly during the economic downturn in the period from 2001 to 2004, i.e., while the total number of forecasters is usually clearly above 20, it decreases to around 16-17 in the 2001-2004 period. Although it is hard to determine the exact reason for the decline in responses, one could conjecture that some survey participants felt unable to predict S&P 500 prices during the 'burst of the Internet bubble' and therefore did not provide a forecast.

Consensus (average) return expectations are surprisingly low during the earlier decade, e.g., in 10 out of 13 cases they are below the riskfree rate for the 6-month forecast horizon. A market return expectation below the riskfree rate implies a negative market risk premium which violates basic finance principles.[11] At the start of the new millennium, consensus return forecasts increase to more realistic levels, but then level off again after 2005. Overall, the time-series average of the consensus return forecasts over the next six months (5.5%) is somewhat lower than that of the realized returns (7.2%) over the same period. Similar observations can be made for the other forecast horizons.

At each forecast production date, we find substantial variation in the forecasts made by survey participants, e.g., the mean and median standard deviation of the 6-month forecast is 8.9% and 8.0%, respectively. Variation in forecasts is especially large in the Internet bubble period which stretches from 1996 to 2004. However, even in more normal periods, e.g., in June of 1992, return forecasts within one standard deviation of the consensus range from -3.5% to 9.2%. If survey participants have identical access to information and properly process this information, then the large variation in forecasts is a sign of differences in preferences across survey participants.

Table 2 reports summary statistics on the forecast error. We offer summary statistics for the whole sample period from 1992 to 2008 and for two subperiods which are formed according to whether the realized return of the S&P 500 during the prior year has been above (expansion) or below (recession) the sample median return. Our analysis of the subsamples is motivated by the hypothesis that survey participants' loss functions could vary with the economic climate. Similar to the evidence in Pearce (1984) and Lakonishok (1980), we find that survey participants on average underpredict market returns by close to 5%. The tendency to underpredict is more pronounced during expansions, i.e., realized returns are on average around 10% higher than return forecasts in these periods. In contrast, survey participants slightly overshoot S&P 500 returns for some forecast horizons during recessions. A potential explanation for our findings from the subsamples could be that forecasters expect a reversal in economic conditions over the forecast horizon. Assuming forecasters are more loss averse in recessions, they would then issue especially conservative predictions in an economic expansion.

In Figure 1, we plot the evolution of the S&P 500 price at the end of the forecast horizon against the consensus price forecast over all forecast production dates. Consistent with our earlier discussion, consensus price forecasts are initially below realizations. For the first four years of the new millennium, this pattern reverses. Starting from 2005, average price forecasts seem on target, with only the average price forecasts during the 2008-2009 credit crunch being significantly higher than realized S&P 500 prices. Another interesting question is whether forecasters from the four dominant affiliations participating in the Livingston Surveys show similar propensities to over- or underpredict market returns. Hence, we offer forecast errors averaged by affiliation and forecast production date in Figure 2.

Average forecast errors at each forecast production date are highly positively correlated across the four affiliations, implying that differences in preferences are not driven by affiliation. Applying the methods of Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007) to subsamples based on affiliation (not reported) confirms these preliminary conclusions.

4.2 Findings from General Rationality Tests

In this section, we report our empirical findings from the application of the methodologies of Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007). Our α estimates from the method of Elliott, Komunjer, and Timmermann (2005) with p=2 are shown in Table 3.[12] In Tables 4 and 5, we report the outcomes from the rationality tests of the two methods. The structure of the three tables is similar. In Panel A, we report our findings from tests which assume that forecasters have identical preferences. In Panel B, we allow loss functions to vary across survey participants. Subpanels I-IV contain the analysis of the four forecast horizons. For each forecast horizon, we examine the whole sample and two subsamples based on whether prior year's S&P 500 return has been above (expansion) or below (recession) the sample median. In Panel B, the low numbers of observations (obs) make the analysis of the long-term forecasts and the subsamples infeasible.[13]

Panel A of Table 3 shows the α estimate (in bold) and three statistical tests (in parentheses). The numbers reported under 'asy. p-value α=0.50' and 'boot. p-value α=0.50' are the asymptotic and bootstrap p-values of a test whether the α estimate is different from 0.50 (symmetry). The numbers reported under 'asy. p-value α=0 or 1' are the p-values from an Andrews (2002) test of whether the α estimate is on the closer boundary (zero or one). With the exception of the bootstrap p-values, we show the same information in compressed form in Panel B. In particular, we offer the proportion of α estimates within one specific range, i.e., below 0.10, between 0.10 and 0.50, between 0.50 and 0.90 and above 0.90 (% in group). We then also report the proportion of α estimates within each range which are different from 0.50 according to an asymptotic t-test (% $\neq$ 0.5) and the proportion of α estimates which are different from the closer boundary according to a test proposed by Andrews (2002) (% $\neq$ boundary). We use the 95% confidence level for the two hypothesis tests.

In Panel A of Tables 4 and 5, we show the J-test statistic (J) together with the p-values from an asymptotic test ('asy. p-value') and a bootstrap test ('boot. p-value') of whether the J-test is different from zero. We show the same information in compressed form in Panel B, i.e., we again offer the proportion of J-tests within one specific range (% in group). We also report the proportion of J-tests within one specific range which are different from zero according to an asymptotic t-test (% $\neq$ 0). To be more conservative, we use the 90% confidence level for this classification. In Table 5, the loss function can either depend on the forecast error (static loss) or the forecast error and the S&P 500 index two months prior to the survey's publication (flexible loss).

For robustness reasons, we have used three different sets of instruments in our application of the method of Elliott, Komunjer, and Timmermann (2005). However, as our findings are similar across the three specifications, we focus on the set featuring all instruments in our discussion.

Assuming that survey participants have identical preferences, the empirical findings obtained from the method of Elliott, Komunjer, and Timmermann (2005) suggest that the Livingston Survey forecasters are extremely averse to losses (Panel A, Table 3). Considering the whole sample, the α estimates range from 0.03 to 0.09, with the estimates from the long-term forecasts being smaller than those from the short-term forecasts. With the exception of the 24-month forecast horizon, the asymptotic t-test and the bootstrap reject the hypothesis of symmetric loss at the 95% confidence level. Similarly, the simulation test proposed by Andrews (2002) consistently rejects the hypothesis that α is on the lower boundary of its parameter space. Notwithstanding the finding that forecasters' preferences are apparently not characterized by MSE loss, the J-tests in Panel A of Table 4 always reject rationality at high confidence levels according to both asymptotic theory and the bootstrap.

One potential reason for these findings is that the general loss function specified by Elliott, Komunjer, and Timmermann (2005) is not flexible enough to capture agents' true loss function. However, if we base our tests on the less restrictive approach of Patton and Timmermann (2007), our findings still strongly reject forecast rationality (Panel A, Table 5). Next, the loss functions estimated from the two methods, shown in Figures 3 and 4, seem similar, with one difference being that the loss functions implied from the Patton and Timmermann (2007) approach only weakly penalize forecast errors close to zero. Overall, it appears unlikely that flexibility issues drive the rejection of rationality.

Our summary statistics suggest two other reasons for the rejection of forecast rationality. First, the change in the value and sign of the average forecast error across economic states of nature suggests that forecasters' loss functions depend on the economic state. Second, the large variation in return forecasts could indicate that preferences vary across survey participants.

A simple test of the first hypothesis is to split the sample into subperiods based on the macroeconomic state, e.g., based on whether prior year's market return has been above or below the sample median, and then to rerun the tests of Elliott, Komunjer, and Timmermann (2005). A more sophisticated method is to explicitly link the loss function of Patton and Timmermann (2007) to the level of the S&P 500 index 2 months prior to the survey's publication date. Consistent with our intuition, our findings from these tests reveal that forecasters are more loss averse during expansions than during downturns. For example, the α estimates in Panel A of Table 3 are smaller in expansions than in recessions, with the difference decreasing in the forecast horizon. However, most α estimates are not estimated with great precision, as can be seen from the often high bootstrap p-values. Next, in two cases we can no longer reject the hypothesis that α is on the boundary of its parameter space. In Figure 4, we plot the loss functions implied from the Patton and Timmermann (2007) approach using three different values for the level of the S&P 500 index (800, 1000 and 1200). Again, we find strong evidence that loss aversion increases with the level of the market index.

Unfortunately, our empirical evidence in Panel A of Tables 4 and 5 reveals that, although it helps to condition on the state of the economy, the J-tests still strongly reject forecast rationality according to both asymptotic theory and the bootstrap. As a result, time variation in agents' loss functions does not seem to be the dominant driver of the rejection of rationality.

In Panel B of Tables 3-5, we analyze whether heterogeneity in agents' preferences can rationalize the S&P 500 return forecasts. To this end, we repeat our rationality tests separately by survey participant. Outcomes from these tests are more encouraging. We find that α estimates vary substantially across individual forecasters (Panel B, Table 3), e.g., while 50% of all estimates obtained from the 6-month forecasts are below 0.10, 20% are between 0.10 and 0.50, 15% are between 0.50 and 0.90, and 15% are above 0.90. Moreover, it seems unlikely that the variation in α estimates is entirely driven by sampling error. Most estimates below 0.10 and above 0.90 are significantly different from 0.50, while most estimates between 0.10 and 0.90 are significantly different from their closer boundaries. Although not shown, we also find large variation in the loss functions of the individual survey participants when we use the alternative method of Patton and Timmermann (2007).

Under heterogenous asymmetric loss, the rationality tests in Panel B of Tables 4 and 5 no longer consistently reject forecast rationality. In fact, when we base our rationality tests on the method of Elliott, Komunjer, and Timmermann (2005) the J-tests only reject the rationality of the 6-month forecasts in 10% of all cases, which is identical to the specified error level. Rejection rates for the other specifications are similar to the one for the 6-month forecast horizon. Using the method of Patton and Timmermann (2007), the J-tests reject rationality slightly more often. When we allow the loss functions of the individual survey participants to depend on the level of the S&P 500, the rejection rates increase even further. The difference in findings is probably due to the size properties of the method of Patton and Timmermann (2007), which are especially poor when the number of observations is low (see Table 2 in Patton and Timmermann, 2007).

In conclusion, our empirical evidence could be interpreted as showing that differences in preferences across forecasts can rationalize market return expectations. Differences in preferences have been shown to be important in other settings, e.g., in Capistrán and Timmermann (2009).

4.3 Discussion

When we allow for variation in preferences across survey participants but not across economic states, our statistical tests fail to reject the rationality of the individual stock market expectations in the vast majority of cases. Given that the magnitude and sign of the average forecast error clearly depend on the economic state, it might at first be surprising that we do not need to link survey participants' loss functions to the macroeconomic climate to restore rationality. However, the reason for this finding is rather trivial. Most survey participants only provide forecasts for a limited range of usually consecutive dates, which often fall either into the early or the late sample period. As a result, when we condition loss functions on survey participants, we very roughly also condition on the economic state. Our implicit conditioning of agents' loss functions on the economic state appears to be sufficient to account for the switch in preferences at the start of the new millennium.

Our failure to reject forecast rationality does not necessarily imply that market return forecasts are rational. Another interpretation of our findings could be that our tests have simply introduced enough additional degrees of freedom to sufficiently erode the power of the J-tests. Although we offer some evidence that this is not the case, e.g., even when allowing for variation in loss functions the J-tests still reject forecast rationality for some survey participants, we cannot rule out this possibility.

The fact that the majority of market return forecasts were below the riskfree rate of return in the earlier sample period does certainly leave a bad feeling. Unfortunately, our two methods cannot take this additional piece of information into account.

Even if we accept the hypothesis that heterogenous asymmetric loss can restore the rationality of market return expectations, the re-establishment of rationality comes at a substantial cost. The vast majority of theoretical models in economics and finance assumes that economic agents have homogenous expectations. Although there are a limited number of studies which explicitly rely on differences in opinion (e.g., see Grossman and Stiglitz, 1980; Admati, 1986), in their settings heterogenous expectations arise through different access to information. In contrast, in our setting, even if agents had identical information, variation in preferences would transform this information into different expectations. As we are not aware of studies analyzing the implications of variation in preferences for asset pricing, this could be an interesting area for future research.

5 Summary and Conclusion

We use two new methods proposed by Elliott, Komunjer, and Timmermann (2005) and Patton and Timmermann (2007) to analyze the market return expectations of individual participants from the Livingston Surveys of Professional Forecasters. Although other studies have often rejected the rationality of these forecasts, it is important to understand that their rationality tests assume that agents' preferences can be described with an MSE loss function. Our tests allow for a much wider range of symmetric and asymmetric loss functions and are therefore less likely to reject forecast rationality due to an erroneous assumption on forecasters' preferences. To control for dependence in our data, we combine the two new rationality tests with the Hall and Horowitz (1996) GMM block bootstrap.

Our empirical findings suggest that allowing for homogenous asymmetric loss does not restore forecast rationality. Although forecasters' loss functions penalize negative forecast errors more strongly than positive forecast errors, all J-test statistics are still strongly significant. We believe that two reasons drive this finding. First, the magnitude and sign of the average forecast error depend on the state of the economy, suggesting that agents' loss functions might explicitly depend on variables other than the forecast error. Second, we document a large variation in return forecasts at each forecast production date, which could indicate that loss functions vary across survey participants. Allowing loss functions to depend on the price level of the stock market index helps to reduce the magnitude of the J-tests, but fails to render them insignificant. However, when we allow for variation in loss functions across forecasters, our empirical evidence reveals significant differences in preferences. Moreover, under heterogenous asymmetric loss we can often no longer reject rationality.

References

Admati, A., 1986, "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets," *Econometrica*, 53, 629–657.

Amromin, G., and S. Sharpe, 2006, "From the Horse's Mouth: Gauging Conditional Expected Stock Returns from Investor Surveys," Federal Reserve Board of Governors Working Paper.

Anatolyev, S., 2002, "Bootstrap Inference in Multi–Period Prediction Problems: A Simulation Study," New Economic School of Moscow Working Paper.

Andrews, D. W. K., 2002, "Generalized Method of Moments Estimation When a Parameter Is on a Boundary," *Journal of Business and Economic Statistics*, 20, 530–544.

Bergström, P., M. Dahlberg, and E. Johansson, 1997, "GMM Bootstrapping and Testing in Dynamic Panels," Uppsala University Working Paper.

Brav, A., R. Lehavy, and R. Michaely, 2005, "Using Expectations to Test Asset Pricing Models," *Financial Management*, 34, 31–64.

Breen, W., L. R. Glosten, and R. Jagannathan, 1989, "Economic Significance of Predictable Variation in Stock Index Returns," *The Journal of Finance*, 44, 1177–1190.

Brown, B. W., and S. Maital, 1981, "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica*, 49, 491–504.

Campbell, J. Y., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Capistrán, C., and A. Timmermann, 2009, "Disagreement and Biases in Inflation Expectations," Forthcoming in the *Journal of Money, Credit and Banking.*

Chernoff, H., 1954, "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573–578.

Dokko, Y., and R. H. Edelstein, 1989, "How Well Do Economists Forecast Stock Market Prices? A Study of the Livingston Surveys," *American Economic Review*, 79, 865–871.

Elliott, G., I. Komunjer, and A. G. Timmermann, 2005, "Estimation and Testing of Forecast Rationality under Flexible Loss," *Review of Economic Studies*, 72, 1107–1125.

Elliott, G., and A. G. Timmermann, 2004, "Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions," *Journal of Econometrics*, 122, 47–79.

Fama, E. F., and K. R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

Fama, E. F., and K. R. French, 1989, "Business Conditions and Expected Stock Returns," *Journal of Financial Economics*, 25, 23–50.

Ferson, W. E., 1989, "Changes in Expected Security Returns, Risk, and the Level of Interest Rates," *The Journal of Finance*, 44, 1191–1218.

Ferson, W. E., and C. R. Harvey, 1991, "The Variation of Economic Risk Premia," *Journal of Political Economy*, 99, 385–415.

Goyal, A., and I. Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49, 639–654.

Goyal, A., and I. Welch, 2006, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," Brown University Working Paper Series.

Granger, C. W. J., 1999, "Outline of Forecast Theory using Generalized Cost Functions," *Spanish Economic Review*, 1, 161–173.

Granger, C. W. J., and P. Newbold, 1986, *Forecasting Economic Time Series*, Academic Press.

Granger, C. W. J., and H. Pesaran, 2000, "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19, 537–560.

Grossman, S., and J. Stiglitz, 1980, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70, 393–408.

Gu, Z., and J. S. Wu, 2003, "Earnings Skewness and Analyst Forecast Bias," *Journal of Accounting and Economics*, 35, 5–29.

Hall, P., and J. L. Horowitz, 1996, "Bootstrap Critical Values for Tests Based on Generalized–Method–of–Moments Estimators," *Econometrica*, 64, 891–916.

Harvey, C. R., 1989, "Time-varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24, 289–318.

Horowitz, J. L., 1995, "Boostrap Methods in Econometrics: Theory and Numerical Performance," University of Iowa Working Paper.

Horowitz, J. L., 1998, "Bootstrap Methods for Covariance Structures," *Journal of Human Resources*, 33, 38–61.

Horowitz, J. L., 2003, "The Bootstrap in Econometrics," *Statistical Science*, 18, 211–218.

Inkmann, J., 2005, "Finite Sample Properties of One–step, Two–step and Bootstrap Empirical Likelihood Approaches to Efficient GMM Estimation," University of Konstanz Working Paper.

Ke, B., and Y. Yu, 2006, "The Effect of Issuing Biased Earnings Forecasts on Analysts' Access to Management and Survival," *Journal of Accounting Research*, 44, 965–999.

Keim, D. B., and R. F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics*, 17, 357–390.

Lakonishok, J., 1980, "Stock Market Return Expectations: Some General Properties," *The Journal of Finance*, 35, 921–931.

Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Michaely, R., and K. L. Womack, 1999, "Conflict of Interest and the Credibility of Underwriter Analyst Recommendation," *Review of Financial Studies*, 12, 653–686.

Newey, W. K., and K. D. West, 1987, "A Simple Positive–Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Patton, A. J., and A. Timmermann, 2005, "Testable Implications of Forecast Optimality," London School of Economics Working Paper.

Patton, A. J., and A. G. Timmermann, 2007, "Testing Forecast Optimality under Unknown Loss," *Journal of the American Statistical Association*, 102, 1172–1184.

Pearce, D. K., 1984, "An Empirical Analysis of Expected Stock Price Movements," *Journal of Money, Credit, and Banking*, 16, 317–327.

Peel, D. A., and A. R. Nobay, 2003, "Optimal Monetary Policy in a Model of Asymmetric Central Bank Preferences," *Economic Journal*, 113, 657–665.

Pesaran, M. H., and A. Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *The Journal of Finance*, 50, 1201–1229.

Rajan, R., and H. Servaes, 1997, "Analysts Following of Initial Public Offerings," *The Journal of Finance*, 52, 507–529.

Welch, I., 1999, "The Equity Size Puzzle," Brown University Working Paper Series.

Notes

[1]We are indebted to one of our referees for suggesting this method as an alternative to the one proposed by Elliott, Komunjer, and Timmermann (2005).

[2]An exception is the study of Dokko and Edelstein (1989), whose findings do not reject unbiasedness. On the other hand, however, their study also offers no convincing evidence of rationality and their findings on unbiasedness depend on Merton's (1980) version of the capital asset pricing model – a model which is itself not beyond doubt.

[3]As an example, asset pricing studies often use the average realized equity return to proxy for the expected equity return. However, a necessary condition for this approximation to be valid is that economic agents are rational and have quadratic loss functions (see, e.g., Amromin and Sharpe, 2006; Brav, Lehavy, and Michaely, 2005).

[4]In this context, we find it interesting to note that Ke and Yu (2006) report evidence which suggests that security analysts who make conservative earnings forecasts increase their chances of keeping their jobs.

[5]Horowitz (2003, 1995) offers additional information on this and other bootstrap methods used in econometrics. Inkmann (2005), Anatolyev (2002), Horowitz (1998) and Bergström, Dahlberg, and Johansson (1997) show that bootstrap critical values are usually more accurate reflections of the true values than the asymptotic critical values.

[6]See, e.g., Keim and Stambaugh (1986), Campbell (1987), Fama and French (1988), Breen, Glosten, and Jagannathan (1989), Fama and French (1989), Ferson (1989), Harvey (1989) and Ferson and Harvey (1991). It should be noted that all cited studies had been published before the start of our sample period. As a result, well-informed survey participants should have known about the predictive ability of these instruments.

[7]We thank Robert Shiller for making this variable available on his website.

[8]The special case of MSE loss is embedded in the method of Elliott, Komunjer, and Timmermann (2005). As such, there might not be a need to perform tests based on OLS regressions on our survey data. However, as earlier empirical outcomes could be driven by data quality issues (see Section [2]) and not the assumption of MSE loss, we found it interesting to perform these tests. We have estimated equations [4] and [5] with both maximum likelihood methods controlling for time-series and cross-sectional correlation, and OLS methods in combination with the GMM block bootstrap. Similar to prior studies, these tests strongly reject rationality.

[9]We are indebted to one of our referees for pointing out that, if the true value of α is on the boundary of its parameter space, the α estimate can no longer be normally distributed in the asymptotic limit.

[10]Other block lengths, e.g., 16 or 24, did not materially change our inferences.

[11]We acknowledge that the S&P 500 returns are excluding dividend payments, which complicates a comparison with the riskfree rate. However, data on Robert Shiller's website suggests that dividend payments on the S&P 500 ranged from 1% to 3% during our sample period. As a result, even if we added (correctly anticipated) dividends to return forecasts, this would not change the conclusion that expected market returns are extremely low during the earlier decade.

[12]We also applied the method of Elliott, Komunjer, and Timmermann (2005) with p=1. As our empirical findings from these tests are similar to the ones with p=2, we do not report them in this study. However, they are available upon request from the first author.

[13]An individual survey participant must produce at least 14 (16) forecasts to be included in the tests in Panel B of Tables 3 and 4 (5). Changes in these requirements do not materially impact our empirical findings.

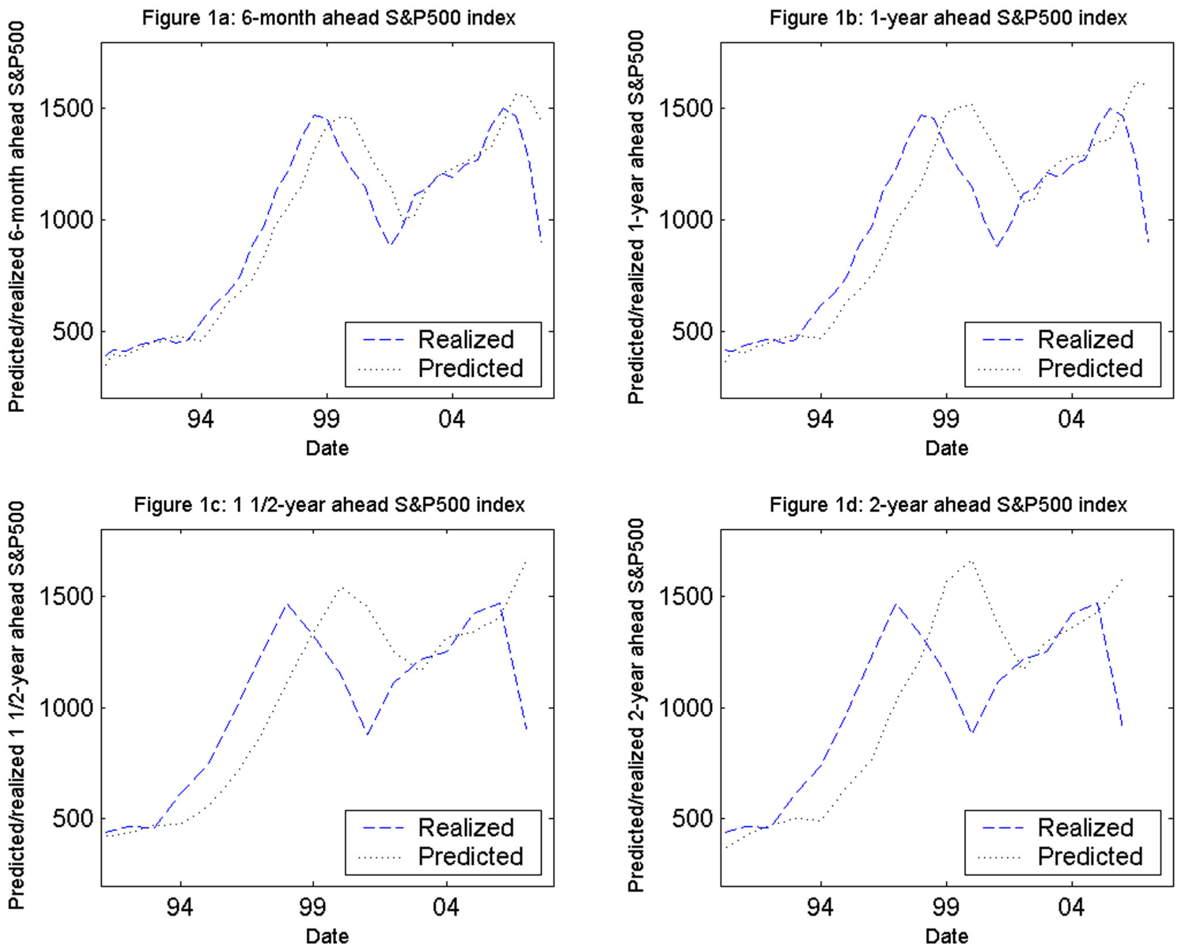

Figure 1: In this figure, we plot the consensus (i.e., the mean over all individual forecasters) S&P 500 price forecast against the subsequent realized price of the S&P 500 for each survey publication date. On the horizonal axis ("date"), we show the Livingston Surveys' publication dates. The vertical axis ("predicted/realized x-month ahead S&P 500") indicates the price of the S&P 500. The dashed line gives the price of the S&P 500 at the end of the forecast horizon, and the dotted line the consensus price forecast for the end of the forecast horizon. The forecast horizon is equal to 6 months (Panel A), 12 months (Panel B), 18 months (Panel C) and 24 months (Panel D). The 6-month and 12-month forecasts are produced at a biannual frequency, whereas the 18-month and 24-month forecasts are produced at an annual frequency.

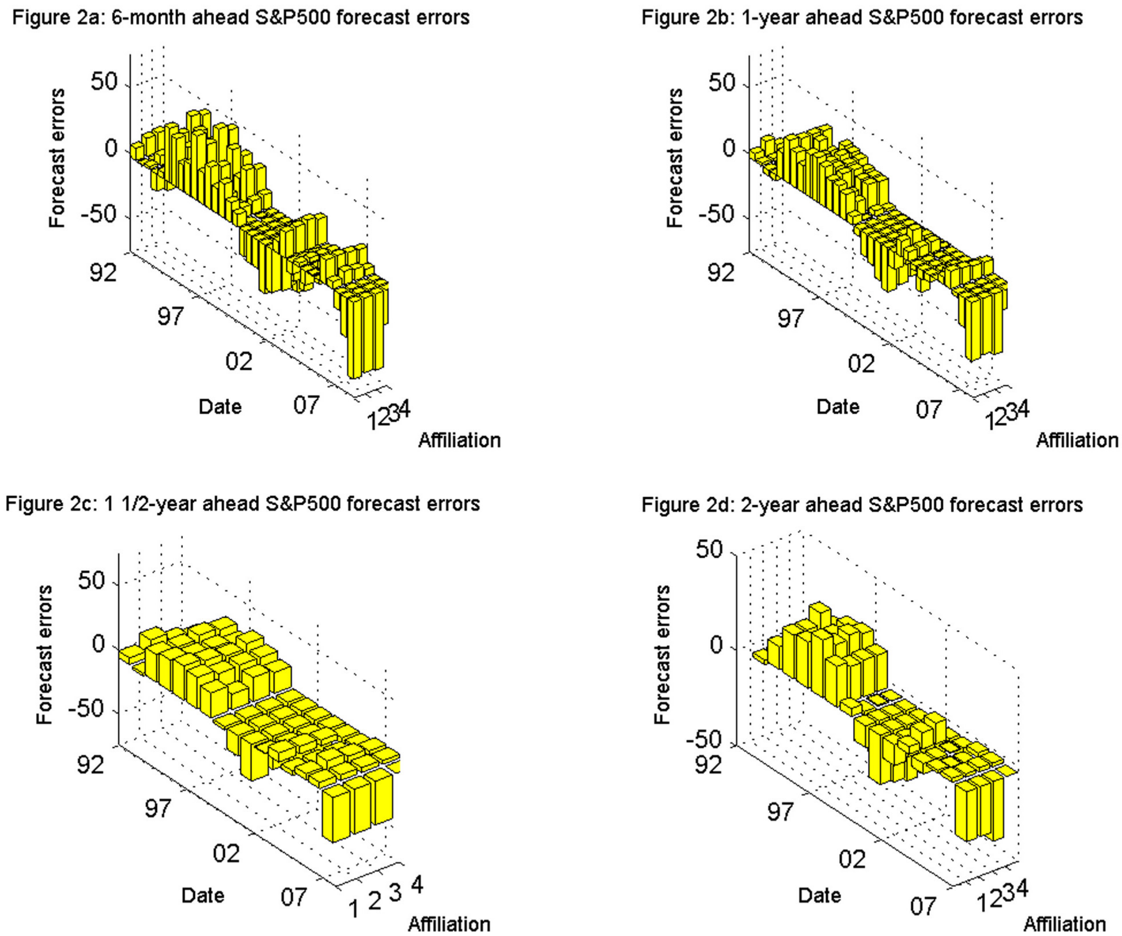

Figure 2: In this figure, we plot forecast errors averaged over survey participants from the four dominant affiliations participating in the Livingston Surveys over our sample period. The forecast error equals the realized return minus the expected return on the S&P 500 over the forecast horizon. On the axis labeled as "date", we plot the surveys' publication date. The axis labeled as "affiliation" shows the four most dominant affiliations participating in the Livingston Surveys, i.e., '1' = non-financial business, '2' = academic institution, '3' = commercial banking, '4' = investment banking. The final axis shows the mean forecast error. The forecast horizon is equal to 6 months (Panel A), 12 months (Panel B), 18 months (Panel C) and 24 months (Panel D). The 6-month and 12-month forecasts are produced at a biannual frequency, whereas the 18-month and 24-month forecasts are produced at an annual frequency.

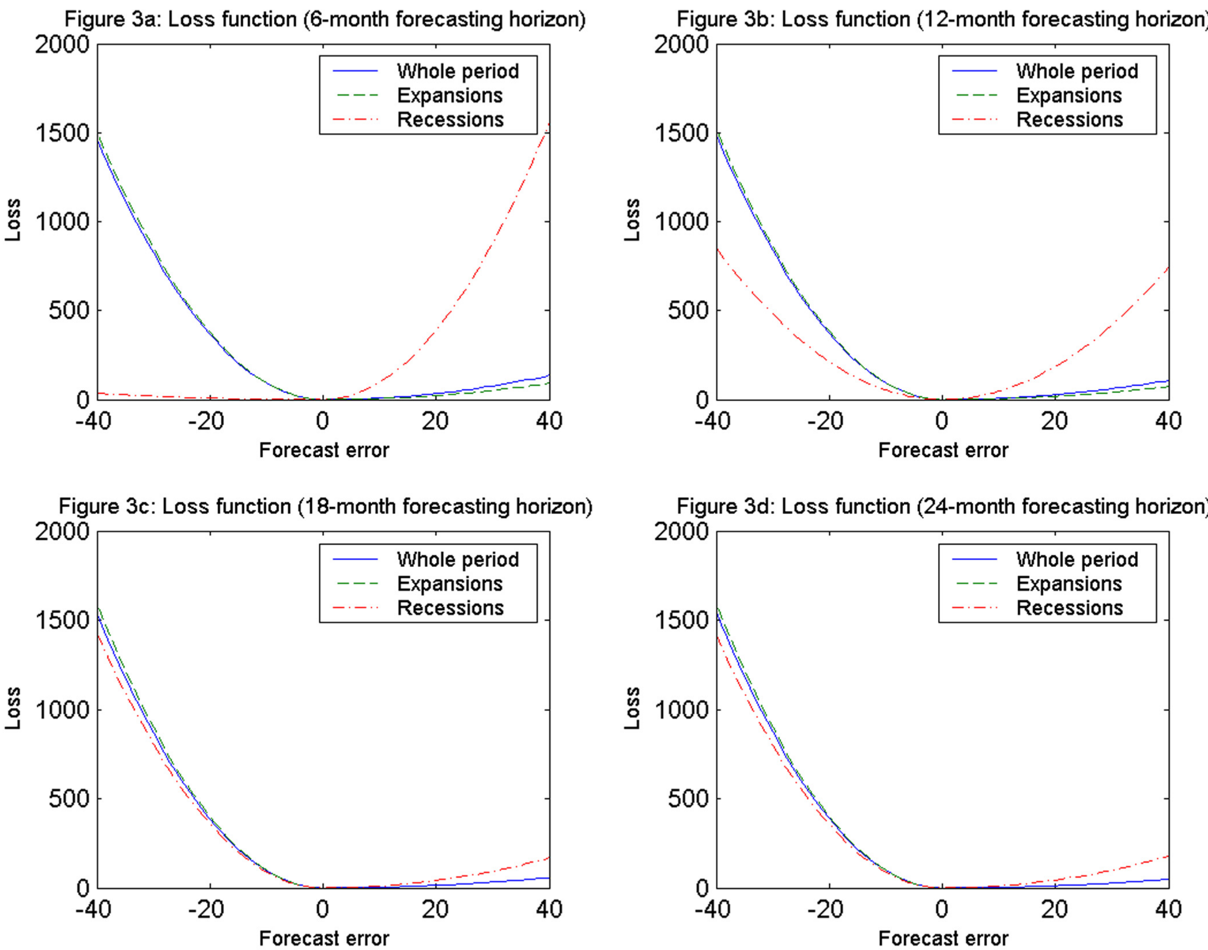

Figure 3: In this figure, we plot the loss functions implied by the α estimates obtained from the parametric approach of Elliott, Komunjer and Timmermann (2005). The loss functions are constructed from the α estimates which use all available instruments and assume that p – the power to which the forecast error is raised – equals 2 (see the estimates under "all instruments" in Table 3). On the horizontal axis, we plot the forecast error. The vertical axis indicates the value of the loss function. Panel A-D show the loss functions implied from the 6-month, 12-month, 18-month and 24-month ahead forecasts, respectively. Each subpanel contains both the loss function implied from the whole sample and the loss functions implied from two subsamples which are formed according to whether prior year's S&P 500 realized return was above (expansion) or below (recession) the sample period's median realized return of the S&P 500.

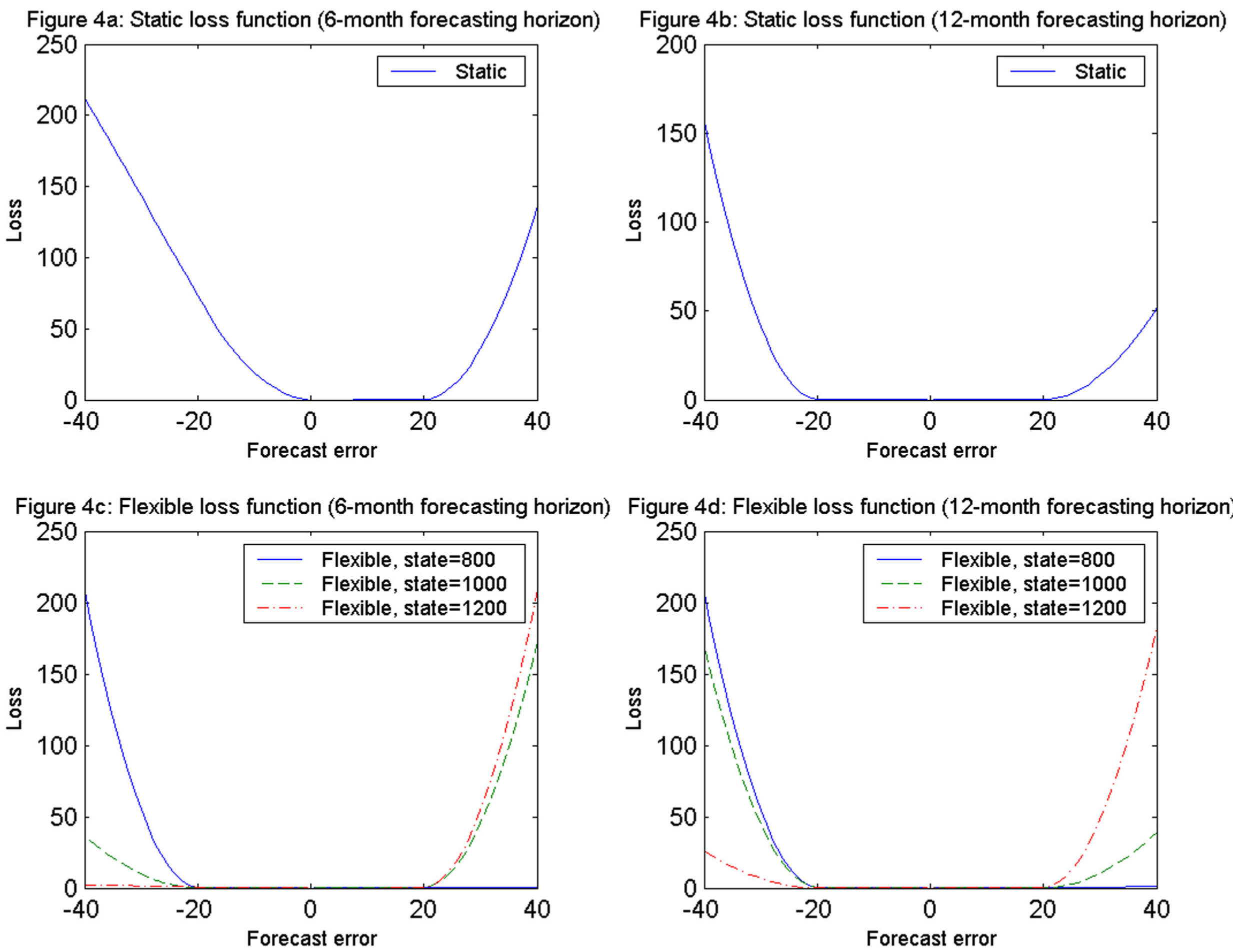

Figure 4: In this figure, we plot the loss functions implied by the estimates obtained from the non-parametric approach of Patton and Timmermann (2007). On the horizontal axis, we plot the forecast error. The vertical axis indicates the value of the loss function. Panel A and B show static (time-invariant) loss functions implied from the 6-month and 12-month ahead forecasts, respectively. Panel C and D show flexible loss functions, which explicitly depend on the price of the S&P 500 2 months prior to the surveys' publication date, implied from the 6-month and 12-month ahead forecasts. The flexible loss function are plotted using three potential prices of the S&P 500, which are 800, 1000 and 1200.

Table 1: Number of forecasts, realized return, and consensus expected return

	6-month horizon				12-month horizon				1 1/2-year horizon				2-year horizon				
Fct. Prod. Date	#	Realized return	Mean Fct.	StD Fct.	#	Realized return	Mean Fct.	StD Fct.	#	Realized return	Mean Fct.	StD Fct.	#	Realized return	Mean Fct.	StD Fct.	Riskfree rate (-2M)
1992.06	26	14.0	3.1	6.4	26	10.4	4.9	4.5	23	9.3	3.4	4.2					4.25
1992.12	21	6.9	6.6	7.0	21	7.1	4.7	4.2					20	2.7	4.5	3.6	3.25
1993.06	25	7.2	4.0	9.0	25	-1.4	3.8	6.4	24	1.3	2.9	4.7					3.25
1993.12	32	-9.3	4.1	7.6	32	-1.5	3.7	5.7					30	14.9	3.7	4.4	3.38
1994.06	35	6.9	2.6	5.5	35	22.6	2.6	4.0	34	24.3	2.1	4.5					4.00
1994.12	27	40.7	0.7	7.8	26	34.1	2.6	4.8					22	27.0	3.8	4.6	5.69
1995.06	35	27.8	-0.7	9.3	34	23.1	0.9	6.4	32	22.7	1.9	4.7					6.25
1995.12	31	18.6	4.5	10.1	32	20.3	3.6	6.5					27	25.5	2.5	4.6	5.94
1996.06	33	22.0	1.5	7.8	32	32.0	2.3	7.3	32	27.9	2.6	4.8					5.46
1996.12	30	42.8	-2.1	12.0	31	31.0	1.5	7.6					26	28.8	1.7	5.4	5.53
1997.06	26	20.2	-2.2	9.0	25	28.1	0.0	8.4	24	24.5	1.6	7.4					5.82
1997.12	27	36.5	4.8	9.9	29	26.7	3.4	8.6					24	23.1	3.1	6.2	5.72
1998.06	29	17.5	-1.1	10.8	28	21.1	0.3	7.1	25	18.9	1.2	5.0					5.66
1998.12	26	24.7	0.8	15.2	28	19.5	1.0	9.3					24	3.6	2.7	5.9	5.37
1999.06	24	14.6	2.8	13.3	24	6.0	2.1	10.8	22	-2.6	2.3	8.3					5.00
1999.12	21	-2.0	2.0	13.0	21	-10.1	4.5	8.3					17	-11.6	4.9	5.9	6.18
2000.06	23	-17.6	2.2	11.1	22	-15.8	3.9	7.9	21	-14.6	4.1	6.5					6.28
2000.12	20	-14.0	9.7	7.8	19	-13.0	9.0	5.5					14	-18.4	9.0	2.9	6.80
2001.06	14	-12.1	7.3	10.1	13	-19.2	9.4	4.0	13	-19.8	8.5	3.1					4.81
2001.12	17	-25.7	13.5	9.5	18	-23.4	12.9	8.7					14	-1.6	9.2	6.4	2.48
2002.06	19	-21.0	15.3	13.1	19	-1.6	11.8	5.1	18	8.1	10.7	4.2					2.01
2002.12	18	22.7	16.8	11.7	18	26.4	17.1	9.4					14	17.4	12.3	6.2	1.76
2003.06	21	30.2	9.2	11.4	20	17.1	11.0	5.6	17	15.7	11.7	6.5					1.29
2003.12	16	5.3	13.0	6.1	16	9.0	12.4	5.9					16	6.0	9.7	4.0	1.15
2004.06	18	12.9	10.4	6.0	17	4.4	10.5	5.1	17	6.2	9.5	3.4					1.14
2004.12	15	-3.4	6.9	4.4	16	3.0	8.0	3.4					15	8.2	6.7	2.9	2.06
2005.06	28	9.8	9.7	7.6	28	6.6	7.3	6.0	27	12.3	7.4	4.6					3.13
2005.12	26	3.5	6.3	7.6	27	13.6	7.1	4.5					24	8.5	6.3	3.4	4.12
2006.06	27	24.7	5.2	7.5	27	18.4	5.3	5.6	27	10.2	5.6	3.7					5.03
2006.12	24	12.4	5.1	6.3	25	3.5	5.8	4.0					22	-20.2	5.9	3.5	5.37
2007.06	23	-4.6	5.5	5.0	23	-14.9	6.1	3.1	21	-28.8	5.8	2.9					5.35
2007.12	23	-24.0	5.2	5.4	23	-38.5	6.2	3.1									5.24
2008.06	20	-50.2	7.8	8.0													2.71

[a] In this table, we show the number of survey participants (#), the realized return over the forecast horizon, the consensus expected return over the forecast horizon (mean fct.), and the standard deviation of the expected returns over the forecast horizon (StD fct.) for each publication date of the Livingston Survey (Fct. prod. date). The forecast horizons shown in the table are equal to 6 months, 12 months, 1 and 1/2 years and 2 years. Our sample period includes all Livingston Survey publications from June 1992 to June 2008. We compute the consensus expected return as the average of all individual forecasters' expected returns at each point in time. We also provide the 3-month U.S. Interbank Offer Rate two months prior to the survey publication date (riskfree rate -2M).

Table 2: Summary statistics on forecast errors

	Obs	Mean (%)	Std (%)	Skew	Kurt	Min (%)	5pct (%)	Median (%)	95pct (%)	Max (%)
Panel A: S&P 500 stock return prediction errors (6-month forecasting horizon)										
Whole sample	800	4.7	24.2	-0.4	3.3	-82.5	-36.7	5.2	42.1	78.8
Expansion periods	445	12.0	21.4	-0.2	2.7	-36.9	-27.5	15.3	44.1	78.8
Recession periods	355	-4.4	24.5	-0.4	3.5	-82.5	-53.7	-1.6	36.4	57.7
Panel B: S&P 500 stock return prediction errors (1-year forecasting horizon)										
Whole sample	780	4.7	20.3	-0.5	2.8	-58.3	-31.9	6.7	32.3	54.8
Expansion periods	446	8.8	20.7	-0.8	3.3	-52.1	-39.7	13.5	34.4	54.8
Recession periods	334	-0.7	18.3	-0.2	2.8	-58.3	-31.8	0.4	29.4	43.1
Panel C: S&P 500 stock return prediction errors (1-1/2 year forecasting horizon)										
Whole sample	377	5.1	17.6	-0.6	2.8	-42.5	-32.0	5.6	27.3	43.7
Expansion periods	200	10.2	15.7	-0.6	2.3	-29.1	-19.9	15.3	28.6	43.7
Recession periods	177	-0.5	18.1	-0.6	2.7	-42.5	-34.8	3.1	23.6	41.2
Panel D: S&P 500 stock return prediction errors (2-year forecasting horizon)										
Whole sample	309	4.3	17.3	-0.3	2.3	-35.6	-27.0	4.3	28.1	39.9
Expansion periods	186	6.7	18.4	-0.5	2.2	-35.6	-26.8	11.0	30.3	39.9
Recession periods	123	0.7	14.8	-0.2	3.0	-33.7	-28.7	1.3	25.1	35.7

[a] In this table, we offer summary statistics on the forecast error. We define the forecast error as the realized return of the S&P 500 minus its return expectation obtained from the Livingston Survey of Professional Forecasters. The forecast horizon is equal to 6 months (Panel A), 12 months (Panel B), 18 months (Panel C) and 24 months (Panel D). In addition to the whole sample period, we also consider two subperiods, which are based on whether prior year's realized S&P 500 return was above ("expansion") or below ("recession") the sample median. The summary statistics we report are the number of observations per (sub)sample (obs), the mean, the standard deviation (Std), skewness (Skew), kurtosis (Kurt), the minimum (Min), the 5th percentile (5pct), the median, the 95th percentile (95pct) and the maximum (Max).

Table 3: Estimation of Parametric Loss Function[a]

Panel A: Homogeneous Loss Functions

	Obs	c + all instruments				c + first 2 instruments				c + last 3 instruments			
		$\hat{\alpha}$	Asy. p-value $\alpha = 0.5$	Asy. p-value α =0 or 1	Boot. p-value $\alpha = 0.5$	$\hat{\alpha}$	Asy. p-value $\alpha = 0.5$	Asy. p-value α =0 or 1	Boot. p-value $\alpha = 0.5$	$\hat{\alpha}$	Asy. p-value $\alpha = 0.5$	Asy. p-value α =0 or 1	Boot. p-value $\alpha = 0.5$
Panel A.I: 6-month forecasting horizon													
Whole sample	800	**0.085**	(0.00)	(0.00)	(0.05)	**0.286**	(0.00)	(0.00)	(0.15)	**0.253**	(0.00)	(0.00)	(0.19)
Expansion periods	445	**0.057**	(0.00)	(0.00)	(0.19)	**0.193**	(0.00)	(0.00)	(0.23)	**0.094**	(0.00)	(0.00)	(0.22)
Recession periods	355	**0.978**	(0.00)	(0.00)	(0.00)	**0.914**	(0.00)	(0.00)	(0.06)	**0.911**	(0.00)	(0.00)	(0.00)
Panel A.II: 1-year forecasting horizon													
Whole sample	780	**0.068**	(0.00)	(0.00)	(0.04)	**0.180**	(0.00)	(0.00)	(0.12)	**0.142**	(0.00)	(0.00)	(0.14)
Expansion periods	446	**0.045**	(0.00)	(0.03)	(0.08)	**0.136**	(0.00)	(0.00)	(0.10)	**0.065**	(0.00)	(0.00)	(0.13)
Recession periods	334	**0.467**	(0.31)	(0.00)	(0.80)	**0.439**	(0.10)	(0.00)	(0.44)	**0.686**	(0.00)	(0.00)	(0.25)
Panel A.III: 1 1/2 year forecasting horizon													
Whole sample	377	**0.036**	(0.00)	(0.02)	(0.00)	**0.042**	(0.00)	(0.03)	(0.00)	**0.147**	(0.00)	(0.00)	(0.19)
Expansion periods	200	**0.000**	(0.00)	(0.07)	(0.00)	**0.004**	(0.00)	(0.05)	(0.00)	**-0.002**	(0.00)	(1.00)	(0.00)
Recession periods	177	**0.105**	(0.00)	(0.00)	(0.34)	**0.094**	(0.00)	(0.00)	(0.19)	**0.192**	(0.00)	(0.00)	(0.64)
Panel A.IV: 2-years forecasting horizon													
Whole sample	309	**0.030**	(0.00)	(0.04)	(0.09)	**0.146**	(0.00)	(0.00)	(0.13)	**0.088**	(0.00)	(0.00)	(0.17)
Expansion periods	186	**0.000**	(0.00)	(0.07)	(0.00)	**0.013**	(0.00)	(0.18)	(0.00)	**0.031**	(0.00)	(0.05)	(0.14)
Recession periods	123	**0.112**	(0.00)	(0.00)	(0.25)	**0.134**	(0.00)	(0.00)	(0.00)	**0.404**	(0.05)	(0.00)	(0.38)

Panel B: Heterogeneous Loss Functions

	Obs		$\hat{\alpha}$ (c + all instruments)				$\hat{\alpha}$ (c + first 2 instruments)				$\hat{\alpha}$ (c + last 3 instruments)			
			< 0.1	0.1-0.5	0.5-0.9	> 0.90	< 0.1	0.1-0.5	0.5-0.9	> 0.90	< 0.1	0.1-0.5	0.5-0.9	> 0.90
Panel B.I: 6-month forecasting horizon														
Whole sample	433	% in group	**0.50**	**0.20**	**0.15**	**0.15**	**0.00**	**0.60**	**0.35**	**0.05**	**0.10**	**0.50**	**0.40**	**0.00**
		% $\neq$ 0.5	1.00	0.25	0.33	1.00	-	0.33	0.29	1.00	1.00	0.40	0.38	-
		% $\neq$ boundary	0.00	0.75	1.00	0.00	-	0.67	1.00	0.00	0.00	0.70	1.00	-

(continued on next page)

Table 3: Estimation of Parametric Loss Function (continued)[a]

			$\hat{\alpha}$ (c + all instruments)				$\hat{\alpha}$ (c + first 2 instruments)				$\hat{\alpha}$ (c + last 3 instruments)			
Panel B: Heterogeneous Loss Functions (continued)	Obs		< 0.1	0.1-0.5	0.5-0.9	> 0.90	< 0.1	0.1-0.5	0.5-0.9	> 0.90	< 0.1	0.1-0.5	0.5-0.9	> 0.90
Panel B.II: 1-year forecasting horizon														
Whole sample	421	% in group	**0.30**	**0.25**	**0.30**	**0.15**	**0.10**	**0.50**	**0.30**	**0.10**	**0.10**	**0.40**	**0.35**	**0.15**
		% $\neq$ 0.5	1.00	0.60	0.83	1.00	1.00	0.40	0.17	1.00	1.00	0.63	0.14	1.00
		% $\neq$ boundary	0.00	0.60	0.67	0.00	0.00	0.80	1.00	0.00	0.00	0.75	1.00	0.00

[a] In this table, we show the estimation outcomes of agents' loss functions from the Elliott, Komunjer and Timmermann (2005) methodology with squared forecast errors (p=2). In Panel A, we impose the condition that the loss functions of survey participants have to be equal to one another. In contrast, in Panel B we allow for variation in loss functions. The forecast horizon is equal to 6 months (subpanel I), 12 months (subpanel II), 18 months (subpanel III) and 24 months (subpanel IV). Due to data constraints, the two longer forecast horizons are not considered in Panel B. In addition to the whole sample period, we also consider two subperiods in Panel A, which are based on whether prior year's realized S&P 500 return was above ("expansion") or below ("recession") the sample median. We use three sets of instrumental variables. The first set contains a constant (c) plus the return spread between 3-month and 1-month U.S. government bills, the S&P 500 dividend yield, the yield spread between Aaa- and Baa-rated corporate bonds, the yield spread between 10-year and 1-year U.S. government bonds, and the yield on a 3-month U.S. government bill. The second set contains a constant plus the first two of the former instruments, whereas the third set contains a constant plus the final three of the former instruments. Obs is the total number of observations. In Panel A, we show the alpha estimate next to both an asymptotic (asy. p-value α=0.5) and a bootstrap test (boot. p-value α=0.5) checking whether the estimate is significantly different from 0.5. We also report the outcomes from a test proposed by Andrews (2002) (asy. p-value $\alpha = 0$ or 1) which checks whether the alpha estimate is significantly different from the closer boundary (zero or one). In Panel B, we report the percentage of alpha estimates which are smaller than 0.1, between 0.1 and 0.5, between 0.5 and 0.9 and above 0.9. Within each group, we show the proportion of alpha estimates which are significantly different from 0.5 and from the closer boundary at the 95% confidence level.

Table 4: Rationality of Expected Returns Under Parametric Loss Function[a]

Panel A: Homogeneous Loss Functions

	Obs	c + all instruments J	c + all instruments Asy. p-value	c + all instruments Boot. p-value	c + first 2 instruments J	c + first 2 instruments Asy. p-value	c + first 2 instruments Boot. p-value	c + last 3 instruments J	c + last 3 instruments Asy. p-value	c + last 3 instruments Boot. p-value
Panel A.I: 6-month forecasting horizon										
Whole sample	800	**79.8**	(0.00)	(0.00)	**39.7**	(0.00)	(0.00)	**59.8**	(0.00)	(0.00)
Expansion periods	445	**44.1**	(0.00)	(0.00)	**7.4**	(0.06)	(0.00)	**38.7**	(0.00)	(0.00)
Recession periods	355	**47.3**	(0.00)	(0.00)	**43.7**	(0.00)	(0.00)	**44.5**	(0.00)	(0.00)
Panel A.II: 1-year forecasting horizon										
Whole sample	780	**75.8**	(0.00)	(0.00)	**52.4**	(0.00)	(0.00)	**67.8**	(0.00)	(0.00)
Expansion periods	446	**45.4**	(0.00)	(0.00)	**37.6**	(0.00)	(0.00)	**41.1**	(0.00)	(0.00)
Recession periods	334	**43.5**	(0.00)	(0.00)	**41.3**	(0.00)	(0.00)	**40.1**	(0.00)	(0.00)
Panel A.III: 1 1/2 year forecasting horizon										
Whole sample	377	**43.7**	(0.00)	(0.00)	**41.4**	(0.00)	(0.00)	**38.2**	(0.00)	(0.00)
Expansion periods	200	**37.4**	(0.00)	(0.00)	**24.8**	(0.00)	(0.00)	**25.6**	(0.00)	(0.00)
Recession periods	177	**24.8**	(0.00)	(0.00)	**20.6**	(0.00)	(0.00)	**19.4**	(0.00)	(0.00)
Panel A.IV: 2-years forecasting horizon										
Whole sample	309	**36.6**	(0.00)	(0.00)	**31.8**	(0.00)	(0.00)	**34.9**	(0.00)	(0.00)
Expansion periods	186	**35.7**	(0.00)	(0.00)	**21.2**	(0.00)	(0.00)	**18.4**	(0.00)	(0.00)
Recession periods	123	**17.6**	(0.01)	(0.12)	**14.8**	(0.00)	(0.09)	**12.5**	(0.01)	(0.00)

Panel B: Heterogeneous Loss Functions

	Obs		J (c + all instruments) < 5	J (c + all instruments) 5-10	J (c + all instruments) > 10	J (c + first 2 instruments) < 5	J (c + first 2 instruments) 5-10	J (c + first 2 instruments) > 10	J (c + last 3 instruments) < 5	J (c + last 3 instruments) 5-10	J (c + last 3 instruments) > 10
Panel B.I: 6-month forecasting horizon											
Whole sample	433	% in group	**0.15**	**0.75**	**0.10**	**0.90**	**0.10**	**0.00**	**0.75**	**0.25**	**0.00**
		% $\neq 0$	0.00	0.00	1.00	0.00	0.50	-	0.00	0.00	-
Panel B.II: 1-year forecasting horizon											
Whole sample	421	% in group	**0.20**	**0.70**	**0.10**	**0.65**	**0.35**	**0.00**	**0.60**	**0.40**	**0.00**
		% $\neq 0$	0.00	0.00	0.50	0.00	0.43	-	0.00	0.13	-

[a] In this table, we show the estimation outcomes of rationality tests (J-tests) from the Elliott, Komunjer and Timmermann (2005) methodology with squared forecast errors (p=2). In Panel A, we impose the condition that the loss functions of survey participants have to be equal to one another. In contrast, in Panel B we allow for variation in loss functions. The forecast horizon is equal to 6 months (subpanel I), 12 months (subpanel II), 18 months (subpanel III) and 24 months (subpanel IV). Due to data constraints, the two longer forecast horizons are not considered in Panel B. In addition to the whole sample period, we also consider two subperiods in Panel A, which are based on whether prior year's realized S&P 500 return was above ("expansion") or below ("recession") the sample median. We use three sets of instrumental variables. The first set contains a constant (c) plus the return spread between 3-month and 1-month U.S. government bills, the S&P 500 dividend yield, the yield spread between Aaa- and Baa-rated corporate bonds, the yield spread between 10-year and 1-year U.S. government bonds, and the yield on a 3-month U.S. government bill. The second set contains a constant plus the first two of the former instruments, whereas the third set contains a constant plus the final three of the former instruments. Obs is the total number of observations. In Panel A, we show the J-test statistic next to both an asymptotic (asy. p-value) and a bootstrap test (boot. p-value) checking whether the estimate is significantly different from 0. In Panel B, we report the percentage of J-test statistics which are smaller than 5, between 5 and 10, and above 10. Within each group, we show the proportion of J-test statistics which are significantly different from 0 at the 90% confidence level.

Table 5: Rationality of Expected Returns Under Non-Parametric Loss Function[a]

Panel A: Homogeneous Loss Functions

	Obs	static loss J	static loss Asy. p-value	static loss Boot. p-value	flexible loss J	flexible loss Asy. p-value	flexible loss Boot. p-value
Panel A.I: 6-month forecasting horizon							
Whole sample	800	**97.2**	(0.00)	(0.00)	**39.9**	(0.00)	(0.00)
Panel A.II: 1-year forecasting horizon							
Whole sample	780	**80.8**	(0.00)	(0.00)	**29.7**	(0.00)	(0.00)
Panel A.III: 1 1/2 year forecasting horizon							
Whole sample	377	**56.1**	(0.00)	(0.00)	**15.7**	(0.00)	(0.76)
Panel A.IV: 2-years forecasting horizon							
Whole sample	309	**47.8**	(0.00)	(0.11)	**123.2**	(0.00)	(0.80)

Panel B: Heterogeneous Loss Functions

	Obs		J (static loss) < 5	J (static loss) $5-10$	J (static loss) > 10	J (flexible loss) < 5	J (flexible loss) $5-10$	J (flexible loss) > 10
Panel B.I: 6-month forecasting horizon								
Whole sample	389	% in group	**0.00**	**0.53**	**0.47**	**0.06**	**0.75**	**0.19**
		$\% \neq 0$	-	0.11	1.00	0.00	1.00	1.00
Panel B.II: 1-year forecasting horizon								
Whole sample	377	% in group	**0.00**	**0.65**	**0.35**	**0.12**	**0.65**	**0.24**
		$\% \neq 0$	-	0.09	0.83	0.00	1.00	1.00

[a] In this table, we show the estimation outcomes of rationality tests (J-tests) from the Patton and Timmermann (2007) methodology. In Panel A, we impose the condition that the loss functions of survey participants have to be equal to one another. In contrast, in Panel B we allow for variation in loss functions. The forecast horizon is equal to 6 months (subpanel I), 12 months (subpanel II), 18 months (subpanel III) and 24 months (subpanel IV). Due to data constraints, the two longer forecast horizons are not considered in Panel B. As instrumental variables, we use a constant, the returns on the 1-month U.S. government bill and on the 3-month U.S. government bill, the S&P 500 dividend yield, the yields on Aaa-rated corporate bonds, on Baa-rated corporate bonds, on 1-year government bonds, on 10-year U.S. government bonds and on 3-month U.S. government bills. Obs is the total number of observations. In Panel A, we show the J-test statistic next to both an asymptotic (asy. p-value) and a bootstrap test (boot. p-value) checking whether the estimate is significantly different from 0. In Panel B, we report the percentage of J-test statistics which are smaller than 5, between 5 and 10, and above 10. Within each group, we show the proportion of J-test statistics which are significantly different from 0 at the 90% confidence level.

A Appendix: Inferences When the α Parameter is on the Boundary of the Parameter Space

In the methodology of Elliott, Komunjer, and Timmermann (2005), the α parameter, which controls the skewness of the parametric loss function L, is bounded between zero and one. As a result, their asymptotic results on the distribution of the parameter estimate fail to hold if the true value of the α parameter is on the boundary of the parameter space. In this situation, Andrews (2002) suggests to simulate the distribution of the parameter estimate. To this end, he first approximates the GMM criterion function through a quadratic function which depends on the following two matrices:

$$\mathfrak{J} = \Gamma' M \Gamma \qquad \text{and} \qquad Z_T = \mathfrak{J}^{-1}\Gamma' M T^{\frac{1}{2}} G_T(\theta_0), \tag{14}$$

where Γ is the probability limit of the derivative matrix of the moment conditions with respect to the free parameters, M is the limit of the GMM weighting matrix, $G_T(\theta_0)$ is the vector of moment conditions and T is the number of observations. Moreover, θ and θ_0 are the vectors of parameter estimates and true parameter values, respectively. If we denote the asymptotic variance of $T^{\frac{1}{2}}G_T(\theta_0)$ by $\mathfrak{V}$, then it is possible to show that Z_T converges in distribution to Z, which is normally distributed with expectation equal to zero and variance equal to $\mathfrak{J}^{-1}\Gamma' M \mathfrak{V} M \Gamma \mathfrak{J}^{-1}$. Andrews (2002) then proves that the quadratic form approximating the criterion function converges in distribution to:

$$(\lambda - Z)'\mathfrak{J}(\lambda - Z), \tag{15}$$

where λ are the parameters of the quadratic form.

Using a concept due to Chernoff (1954), the shifted parameter space $\Theta - \theta_0$ can be approximated through a convex cone. In our case, we can use assumption *GMM4** and *Lemma 4* in Andrews (2002) to determine that, if $\alpha_0 = 0$, then the convex cone is specified by $\lambda \geq 0$. In contrast, if $\alpha_0 = 1$, then the convex cone is specified by $\lambda \leq 0$. The main outcome of Andrews (2002) then is:

$$T^{\frac{1}{2}}(\theta - \theta_0) \rightarrow_d \lambda, \text{ where } \lambda \text{ minimizes } (\lambda - Z)'\mathfrak{J}(\lambda - Z) \text{ over the convex cone.} \tag{16}$$

In practice, we (1) compute the $\mathfrak{J}$ matrix and the variance-covariance matrix of Z, (2) construct a realization of Z, and (3) find the value of λ which minimizes equation [15] subject to the constraint on λ. Repeating this procedure 1,000 times allows us to construct the empirical distribution of the α estimate if the true α parameter is on one of the two boundaries of the parameter space.

An interesting question is how well the approach of Andrews (2002) approximates variation in the α estimate compared to a simple t-test. To find out, we conduct a Monte-Carlo analysis of the finite-sample properties of these two approaches. For the Monte-Carlo analysis, we set the true value of α to 0.01, 0.05, 0.50, 0.95 and 0.99. In addition to the variable to forecast, we create one constant and two instruments. Consistent with Elliott and Timmermann (2004), the variable to forecast and the instruments follow a mixture of normals distribution with two states and a switching probability equal to 0.40. The vector of expectations and the variance-covariance matrix across the two states can be seen from the description of Table A1. We construct optimal forecasts through minimizing expected loss with respect to the parameters of the linear forecasting model over all observations from the initial period and all already-known observations from the holdout period. Finally, we use data from the holdout period on the forecast error and the instruments to estimate α and to test its significance with both an asymptotic t-test and the test proposed by Andrews (2002).

We have chosen the mixture of normals distribution to induce skewness in the variable to forecast and the instruments. This choice is important for several reasons. First, if the random variables had been elliptically distributed, then only the constant of the forecasting model would have changed with the true value of α (Elliott and Timmermann, 2004). Second, our summary statistics in Table 1 and 2 suggest that the real-world data can show substantial deviations from normality. Finally, we expect that skewed data should make it harder for the finite-sample distributions of the test statistics to converge to their asymptotic counterparts.

Our findings are in Table A1. When α_0 is far away from 0.50, the t-test has poor size properties. For example, if p equals 2, the lengths of the initial period and the holdout period are 100 and 200 observations, respectively, and α_0 equals 0.01, the t-test rejects the true null hypothesis in 26.3% of all cases at the 95% confidence level. In contrast, if α_0 equals 0.99, it rejects the null hypothesis in 29.7% of all cases. As α_0 moves to 0.50, the size of the t-test improves considerably, although it still rejects the null hypothesis slightly too often (8.2%). The size of the t-test depends crucially on the length of the holdout period. For short holdout periods, e.g., 50 observations, the size of the test is still extremely poor, even if $\alpha_0 = 0.50$. Our findings are similar for $p = 1$.

When α_0 is far away from 0.50, the simulation test proposed by Andrews (2002) performs markedly better than the t-test, although its size properties are still not perfect. As an example, for α_0 close to zero, it rejects the true null hypothesis in too few cases (1%). In contrast, for α_0 close to one, it rejects the null hypothesis too often (10%). Consistent with intuition, if α_0 is 0.50, then the simulation test has worse size properties than the t-test. As before, our outcomes are similar across absolute and quadratic forecast errors. Overall, our Monte-Carlo evidence thus suggests that, if α_0 is close to the boundary of the parameter space, the simulation test is more appropriate than an asymptotic t-test to evaluate whether the α estimate is significantly different from its true value.

Table A1: Size of Asymptotic t-test and Andrews (2002) test[a]

Initial	Holdout	$\alpha_0 = 0.01$		$\alpha_0 = 0.05$		$\alpha_0 = 0.50$		$\alpha_0 = 0.95$		$\alpha_0 = 0.99$	
Period	Period	t-test	Andrews	t-test	Andrews	t-test	Andrews	t-test	Andrews	t-test	Andrews
Panel A: Quadratic loss functions (p=2)											
50	50	0.482	0.005	0.357	0.006	0.178	0.201	0.335	0.106	0.465	0.105
50	100	0.351	0.016	0.233	0.006	0.109	0.153	0.245	0.117	0.355	0.132
100	50	0.590	0.004	0.391	0.007	0.182	0.215	0.433	0.068	0.573	0.066
100	100	0.458	0.001	0.306	0.002	0.111	0.159	0.293	0.054	0.452	0.076
100	200	0.263	0.003	0.155	0.010	0.082	0.129	0.158	0.106	0.297	0.112
Panel B: Linear loss functions (p=1)											
50	50	0.601	0.005	0.399	0.013	0.117	0.173	0.442	0.098	0.624	0.082
50	100	0.598	0.006	0.205	0.015	0.073	0.133	0.241	0.109	0.637	0.055
100	50	0.665	0.004	0.428	0.008	0.139	0.195	0.453	0.064	0.685	0.039
100	100	0.654	0.003	0.272	0.010	0.105	0.143	0.249	0.081	0.702	0.033
100	200	0.476	0.007	0.133	0.008	0.072	0.129	0.157	0.123	0.515	0.065

[a] In this table, we report on the size of an asymptotic t-test and a simulation test proposed by Andrews (2002) which both check whether the α estimate obtained from the methodology of Elliott, Komunjer and Timmermann (2005) is significantly different from its true population value. We set the true value of α equal to 0.01, 0.05, 0.50, 0.95 and 0.99. In Panel A, the underlying (true) loss function raises the forecast error to the power of 2 (p=2), whereas in Panel B it simply takes the absolute value of the forecast error (p=1). The variable to forecast and the two true instrumental variables follow a mixture of normals distribution, whose switching parameter equals 0.40. The expectations of the three variables are $[0.0, 0.0, 0.0]'$ in the first stage and $[0.5, 0.5, 0.5]'$ in the second stage. The covariance matrices in the two states are respectively:

$$\Sigma_1 = \begin{bmatrix} 1.00 & & \\ 0.10 & 0.25 & \\ 0.10 & 0.10 & 0.15 \end{bmatrix}, \qquad \Sigma_2 = 0.1\Sigma_1.$$

We generate optimal forecasts through minimizing the expectation of the loss function on simulated data from the initial period and the fraction of the holdout which is known to the (hypothetical) forecaster. Once we have constructed the time-series of optimal forecasts, we can then estimate the value of the α parameter using data from the holdout period. Subsequently, we check whether the α estimate is significantly different from its true value at the 95% confidence level according to both the asymptotic t-test and the simulation test proposed by Andrews (2002). Repeating this procedure 1,000 times allows us to compute the proportion of times that the null hypothesis is rejected.

B Appendix: GMM Block Bootstrap

In this Appendix, we describe our implementation of the Hall and Horowitz (1996) GMM bootstrap with non-overlapping blocks. To start with, we outline the computation of the α parameter and the J-test statistic using all available observations, which, for the purpose of the bootstrap, are treated as the population. The final α parameter solves:

$$\min_{\alpha} J_N(\alpha, \tilde{\alpha}_N) \equiv \left[N^{-1} \sum_{i=1}^{N} g(X_i, \alpha) \right]' \Omega_N(\tilde{\alpha}_N) \left[N^{-1} \sum_{i=1}^{N} g(X_i, \alpha) \right], \tag{17}$$

where N is the number of observations, $N^{-1} \sum_{i=1}^{N} g(X_i, \alpha)$ is a vector of moment conditions (see equation [7]), α is the decision variable, $\tilde{\alpha}_N$ is the estimated α parameter from an earlier stage (one-step before) estimation, and $\Omega_N(\tilde{\alpha}_N)$, which is the weighting matrix, equals:

$$\Omega_N(\alpha_N) = \left\{ N^{-1} \sum_{i=1}^{N} \left[g(X_i, \alpha) g(X_i, \alpha)' + \sum_{j=1}^{\kappa} H(X_i, X_{i+j}, \alpha) \right] \right\}^{-1}, \tag{18}$$

where $H(X_i, X_{i+j}, \alpha) = g(X_i, \alpha) g(X_{i+j}, \alpha)' + g(X_{i+j}, \alpha) g(X_i, \alpha)'$. Since $\tilde{\alpha}_N$ depends on a weighting matrix, Ω_N, which in turn depends on $\tilde{\alpha}_N$, in practice the estimation process initiates with the weighting matrix set to the identity matrix. Iteration continues until convergence of the α parameter estimate.

Now define:

$$D_N = N^{-1} \sum_{i=1}^{N} \partial g(X_i, \hat{\alpha}_N) / \partial \alpha. \tag{19}$$

It is well known that the variance of the α parameter equals $\sigma_N = N^{-1}(D' \Omega_N D)^{-1}$, and thus that the t-statistics for testing $H_0 : \alpha_N = \alpha_0$ equals $T_N = (\hat{\alpha}_N - \alpha_0)/(\sigma_N)^{1/2}$. In contrast, the J-test can be written as $J_N = K_N(\hat{\alpha}_N)' K_N(\hat{\alpha}_N)$, where $K_N(\alpha) = \Omega_N(\alpha)^{1/2} N^{-1/2} \sum_{i=1}^{N} g(X_i, \alpha)$. Asymptotically, the t-test follows a $N(0, 1)$ distribution, and the J-test a χ^2 distribution with degrees of freedom equal to the number of instruments.

Under bootstrap sampling, we treat the whole sample data as the population. Drawing random samples from this population and computing the corresponding t-test and J-test statistics, we can develop bootstrap empirical distributions of these test statistics, which allow us to infer the α level bootstrap critical values. In order to take account of the dependency in the data, we resample in blocks. In particular, we divide the total sample into b non-overlapping blocks of length l, and then sample the b blocks randomly with replacement to form a bootstrap sample. There is, however, one complication. Since only under exceptional circumstances $N^{-1}\sum_{i=1}^{N} g(X_i,\alpha)=0$, the bootstrap would normally sample from a population which does not fulfill the moment conditions. As a result, the bootstrap and sample versions of the test statistics under consideration would have different asymptotic distributions. This problem can be solved by recentering, i.e., by defining the bootstrap sample moment conditions as $g^*(x,\alpha)=g(x,\alpha)-E^*g(X,\hat{\alpha}_N)$, where x is the bootstrap sample, and E^* denotes the expectation relative to the distribution of the bootstrap samples.

The bootstrap versions of the t-test and the J-test statistic are computed in an analogous way as those for the whole sample, but replacing J_N, Ω_N, D_N, H_N, α_N, $\tilde{\alpha}_N$, g and X_i in equations [10], [11] and [12] with their bootstrap counterparts. We call these bootstrap counterparts J_N^*, Ω_N^*, D_N^*, H_N^*, α_N^*, $\tilde{\alpha}_N^*$, g^* and X_i^*, respectively. The t-test statistic equals $T_N^*=(\hat{\alpha}_N^*-\hat{\alpha}_N)/(\sigma_N^*)^{1/2}$, while the J test equals $J_N^*=K_N^*(\hat{\alpha}_N^*)'K_N^*(\hat{\alpha}_N^*)$, where $K_N^*(\alpha_N^*)=\Omega_N^*(\alpha_N^*)^{1/2}N^{-1/2}\sum_{i=1}^{N} g^*(X_i^*,\alpha_N^*)$. Unfortunately, the asymptotic distributions of T_N^* and J_N^* conditional on the whole sample are not identical to the asymptotic distributions of T_N and J_N. Hall and Horowitz (1996) show that this problem can be solved through the following correction factors. Define:

$$\tilde{W}_N=N^{-1}\sum_{i=0}^{b-1}\sum_{j=1}^{l}\sum_{k=1}^{l} g^*(X_{il+j},\hat{\alpha}_N)g^*(X_{il+k},\hat{\alpha}_N)', \tag{20}$$

$$\bar{\sigma}_N=(D_N'\Omega_N D_N)^{-1}D_N'\Omega_N\Omega_N^{-1}(\hat{\alpha}_N)\Omega_N D_N(D_N'\Omega_N D_N)^{-1}, \tag{21}$$

and

$$\tilde{\sigma}_N=(D_N'\Omega_N D_N)^{-1}D_N'\Omega_N\tilde{W}_N\Omega_N D_N(D_N'\Omega_N D_N)^{-1}. \tag{22}$$

The corrected bootstrap t-test statistic equals $T_N^*=(\bar{\sigma}_N/\tilde{\sigma}_N)^{1/2}(\hat{\alpha}_N^*-\hat{\alpha}_N)/(\sigma_N^*)^{1/2}$. Let V_N^+ be the Moore-Penrose generalized inverse of V_N. Then define:

$$M_N=I-\Omega_N(\hat{\alpha}_N)^{1/2}D_N[D_N'\Omega_N(\hat{\alpha}_N)D_N]^{-1}D_N'\Omega_N(\hat{\alpha}_N)^{1/2} \tag{23}$$

and

$$V_N = M_N \Omega_N(\hat{\alpha}_N)^{1/2} \tilde{W}_N \Omega_N(\hat{\alpha}_N)^{1/2} M_N. \tag{24}$$

The corrected version of $K_N(\hat{\alpha}_N)$ equals:

$$K_N^*(\alpha_N^*) = (V_N^+)^{1/2} \Omega_N^*(\alpha_N^*)^{1/2} N^{-1/2} \sum_{i=1}^{N} g^*(X_i^*, \alpha_N^*). \tag{25}$$

The empirical distributions of the t-test and the J-test statistics then allow us to infer the α level critical values, which in turn can be used to derive bootstrap p-values.