The Properties of Market-Based and Survey Forecasts for Different Data Releases

Lanne, Markku

HECER

2007

Helsinki Center of Economic Research

Discussion Papers

The Properties of Market-Based and Survey Forecasts for Different Data Releases

Markku Lanne
University of Helsinki, RUESG and HECER

Discussion Paper No. 167
May 2007

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 167

The Properties of Market-Based and Survey Forecasts for Different Data Releases*

Abstract

We compare the accuracy of the survey forecasts and forecasts implied by economic binary options on the U.S. non-farm payroll change. These options are available for a number of ranges of the announced figure, and each pays $1 if the released non-farm payroll change falls in the given range. For the first-release data both the market-based and survey forecasts are biased, while they are rational and approximately equally accurate for later releases. Both forecasts are more accurate for later releases. Because of predictability in the revision process, this indicates that the investors in the economic derivatives market are incapable of taking the measurement error in the preliminary estimates efficiently into account. This suggests that economic stability could be enhanced by more accurate first-release figures.

JEL Classification: C5, C82, D8, E44.

Keywords: Expectations, economic derivatives, data vintage, real-time data.

Markku Lanne

Department of Economics
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: markku.lanne@helsinki.fi

* I thank Heikki Kauppi and Panu Poutvaara for useful comments. Financial support from the Research Unit on Economic Structures and Growth (RUESG) in the University of Helsinki funded jointly by the Academy of Finland, University of Helsinki, Bank of Finland, Nokia Group, and the Yrjö Jahnsson Foundation, is gratefully acknowledged.

1 Introduction

Recently, a market for macroecnomic derivatives, i.e., securities involving bets on future macroeconomic development, has emerged. For instance, in October 2002, Goldman Sachs and Deutsche Bank set up a market that allows investors to buy options with payoffs dependent on growth in non-farm payrolls and Euro-area harmonized CPI, among others. Apart from providing means of hedging against macroeconomic risks, these derivatives yield economic forecasts. Data from the first 2.5 years of this market were analyzed by Gürkaynak and Wolfers (in press), who found the market-based forecasts more accurate than those based on a survey.

The payoffs of the current macroeconomic derivatives depend on the first-release figures that are inaccurate, and the "final" or "true" values of the economic variables only become available much later, after a large number of revisions. Typically the first-release figures are biased predictions of later releases that are likely to be close to the "true" values although revisions may still take place even after several years (see Swanson and van Dijk, 2006). Because the payoffs in the economic derivatives market are determined by the first-release figures, it should be optimal for investors in this market to attempt to predict these particular data. On the other hand, respondents in surveys, such as that studied by Gürkaynak and Wolfers (in press), have no such incentive, but for them it should be optimal to attempt to predict the "true" value that only becomes available after a long lag.[1]

Gürkaynak and Wolfers (in press) compared all the forecasts with the first-release figures. However, from the standpoint of forecasting, the ultimate goal is to forecast the "true" or "final" values, so it would be interesting to see whether the market-based forecasts are superior to survey forecasts for later releases as well. Moreover, by comparing the accuracy of the forcasts for different releases of data it is possible to assess the degree to which the investors are able to take the measurement error in

[1] Whether this is really the case, is controversial (see Sørensen and Ottaviani (2006) and the references therein), but our empirical results support this view.

the first-release data into account and determine what are the data they are actually attempting to forecast. From the practical point of view, it is important to know which of the two forecasts are more accurate for the variables we want to forecast, but the results may also have implications for economic theory. Recently, Bomfim (2001) showed that the measurement error in preliminary data releases can be a major source of economic fluctuations. This effect, however, depends on the economic agents' signal extraction capabilities such that if they cannot efficiently filter the noise out of the first-release data, reducing the measurement error would lead to less economic volatility. On the other hand, if they are able to efficiently take the measurement error into account, macroeconomic volatility increases with improvement in the accuracy of the data. Hence, if the investors appear to be forecasting the "final" figures instead of the first-release data, this suggests that they are unable to take the measurement error efficiently into account, and, according to Bomfim's (2001) model, decreasing the measurement error in first-release figures would then have a stabilizing effect on the economy.

In this paper, we compare the market-based and survey forecasts of the change in the U.S. non-farm payroll, for which the entire revision history is readily available. It turns out, that forecasts are not rational for the first-release figures, unlike later releases, indicating that the investors as well as professional forecasters are actually attempting to forecast the "true" data. This suggests that the investors are not capable of taking the measurement error in first-release data efficiently into account, lending support to the idea that enhancing the quality of first-release data would decrease economic fluctuations. Moreover, the differences between market-based and survey forecasts are relatively small, with the survey forecasts actually being somewhat more accurate for later releases.

The plan of the paper is as follows. The data set is described in Section 2. Section 3 contains the comparisons of the market-based and survey forecasts. In section 4, we present the results on the density forecasts implied by the option data. Finally, Section 5 concludes.

2 Data

In this paper, we consider the U.S. non-farm payroll options, for which the entire revision history is readily available. Specifically, we focus on the "digital range", a contract paying $1 if the announced economic number lies between two adjacent strike prices. Assuming risk neutrality, each price can be interpreted as probability of the change in the non-farm payroll falling in the given range. The options are traded in auctions for each data release, and our data set comprises the 33 monthly auctions from October 2002 until June 2005. The auction takes place on the morning of the data release. For details of the institutional arrangements, see Gürkaynak and Wolfers (in press).

The survey forecasts are released by Money Market Services (MMS) on the Friday before each data release. The survey sample consists mainly of professional economists in the financial markets, and the respondents are surveyed up to a week before the derivatives auction.[2]

The payoffs are determined with reference to the initial estimate of the growth in the non-farm payrorolls as reported by the Bureau of Labor Statistics (BLS). These first-release figures are inaccurate, and they will be subsequently revised multiple times. The entire revision history is extracted from the Real-Time Data Set of the Federal Reserve Bank of Philadelphia. In the empirical analyses we concentrate on the data released six and twelve months after the first release for ease of exposition. Because revisions do not take place every month, little additional insight would be provided by discussing all possible revision dates separately. Ideally, we would like to consider the "final" values that have undergone a greater number of revisions, but because the market for macroeconomic derivatives has been started only recently, only approximately one year's worth of revised estimates are available for the latest observations in the option data set. Hence, later revisions cannot be considered

[2]We are grateful to Justin Wolfers for providing the data on derivatives and survey forecasts on his web page.

without leaving out observations from our already very limited data set. Still, as the figures in Table 1 show, all the initial releases have been revised once after six months and most of them twice after 12 months, and the revisions can be quite sizeable.

3 Accuracy of Point Forecasts

Following Gürkaynak and Wolfers (in press), we begin the empirical analysis by comparing the market-based and survey forecasts to the released figures. While Gürkaynak and Wolfers (in press) only considered the first-release data, we present results also for revised data released six and twelve months after the expiration date of the option. By studying the forecast errors we get an idea of how accurate the forecasts implied by the option prices are for each release of data. Panels A and B of Table 2 contain the mean absolute forecast errors (MAE) and root mean squared forecast errors (RMSE) for each of the three releases considered. For comparison, we also present results for survey forecasts released by Money Market Services on the Friday before the first data release. The table shows quite clearly that, on average, both market-based and survey forecasts are most accurate for the twelve-month release and the accuracy improves monotonously after the first release. Moreover, the differences seem to be quite large. Assuming that the releases converge to the most accurate possible, "final", estimate, this suggests that the agents actually forecast the "final" value, indicating their inability to take the estimation errors made by the BLS into account. The differences between the market-based and survey forecasts are, in general, minor. However, for the first-release data, according to the MAE and RMSE criteria the market-based figures are more accurate. This is in line with Gürkaynak and Wolfers (in press) and not surprising since the payoffs of the derivatives depend on the first-release data. For the twelve-month figures that are supposedly closest to the "final" value, the survey forecasts are actually more accurate, but the differences are somewhat smaller than for the first-release data. These differences are not statistically significant at conventional significance levels, suggesting that the market-based forcasts are not superior to survey forecasts. Median forecasts yield similar results, and therefore they are not reported.[3]

To further compare the option-based and survey forecasts, we report in Panel C of Table 2 the results of the "horse race" regression suggested by Fair and Shller (1990), where the actual released figure is regressed on a constant, and the option-based and survey forecasts. The results for the 6 and 12-month releases differ markedly from those for the first release. The regressors are, of course, strongly correlated and, therefore, the standard errors are large. Still, in line with Gürkaynak and Wolfers (in press), the estimates suggest that the survey forecast is uninformative in the case of the first-release data, for which the the coefficients of the option-based and survey forecasts are very close to unity and zero, respectively. For the later releases, however, the estimated coefficients suggest that the two forecasts are more or less equally important, lending support to the results concerning the forecast errors in Panels A and B.

The results of the test of forecast efficiency are presented in Panel D of Table 2. Three conclusions emerge. First, both the option-based and survey forecasts are downward biased for all three releases of data, but the bias is much larger for the first than the later releases. Second, the differences between the option-based and survey forecasts are very small. Third, the bias is statistically significant at the 10% level only for the first-release data, with p-values of approximately 0.06 for both forecasts. For the later releases, the p-values exceed 20%. These results can be interpreted in favor of the agents' inability to efficiently take into account the estimation error inherent in the first-release data. In other words, the later releases are closer to what the agents actually are forecasting. This is good news from the viewpoint of using the market-based and survey forecasts, as they do not seem to be contaminated by the estimation error inherent in the first-release data. On the other hand, the result reinforces the impression that the agents are not acting rationally in forming expectations, because the profits accruing from option trading depend on the ability

[3]The results are available upon request.

to specifically forecast the first-release data.

4 Accuracy of Density Forecasts

In addition to the point forecast, the derivatives market yields a full probability distribution for each date, which facilitates the analysis of the accuracy of density forecasts for any release of data. Based on the results for the mean predictions above, our expectation is that the agents' predictive distribution is consistent rather with later releases than the first release.

To examine the correctness of the conditional predictive distribution of the market for the different releases, we consider the probability integral transforms.

$$z_{kt} = \int_0^{y_{kt}} f_{k,t-1}(x)\,dx, \qquad t = 1, 2, ..., T,$$

where $f_{k,t-1}(\cdot)$ is the conditional predictive density extracted from the derivatives data and y_{kt} is the observed value of the change in the non-farm payroll given by the kth release at date t. If the conditional density is correct, the distribution of the sequence $\{z_{kt}\}_{t=1}^{T}$ is independently and unformly on [0, 1] distributed for each k. To check for uniformity, Pearson's goodness-of-fit test can be used. The test statistic is based on a histogram of $\{z_{kt}\}_{t=1}^{T}$ consisting of m bins,

$$\sum_{t=1}^{T} \frac{(T_i - T/m)}{T/m},$$

where T_i is the number of observations in the ith bin. If z_{kt} really is uniformly distributed, T_i should equal T/m for all i, i.e., there should be equal number of observations in each bin. Under the null hypothesis this test statistic follows asymptotically the chi-square distribution with $m - 1$ degrees of freedom. In addition to a formal test, Diebold et al. (1998) recommended examining the uniformity graphically by plotting the histogram on which the test is based, as this may give further information on the shortcomings of the predictive distribution. The hypothesis of no serial correlation in the level and square of the probability integral transform can be tested by using the standard Ljung-Box test.

The histograms of the probability integral transform series along with the pointwise 95% confidence bands are presented in Figure 1. Only the first-release data exhibits a violation of uniformity; for the later releases, all the bins are included in the 95% confidence bands. It seems that the market forecasts put too much emphasis on the lower tail of the distribution to conform to the first-release values. This is in accordance with the result above that the market forecasts are downward biased for the first-release data but not for later releases. The results of Pearson's goodness-of-fit test based on this histogram in Table 3 confirm this finding. Uniformity of the probability integral transform is clearly rejected for the first-release data (p-value equals 0.02), while it cannot be rejected for the other two releases at reasonable significance levels. Autocorrelation in the levels and squares of the probability integral transform cannot be rejected for any of the releases.

All in all, the results thus indicate that the distribution of the market forecast is more accurate for the distribution of later releases than that of the first release, in accordance with the results obtained for the mean forecasts above. Because the later releases are presumably more accurate in the sense of being closer to the "final" or "true" values, these findings thus suggests that it is the "true" distribution that the agents are attempting to forecast. In other words, they are unable to take the measurement error of the BLS into account, although in the derivatives market it clearly would be optimal to forecast the erroneous first-release figure, upon which the payoffs are determined.

5 Conclusion

We consider the forecasts of the U.S. non-farm payroll change implied by the market for economic derivatives and the MMS survey. Unlike the recent study by Gürkaynak and Wolfers (in press), we study the properties of these forecasts with respect to revised values in addition to the first-release data.

The implications of our findings are twofold. First, contrary to what the results of Gürkaynak and Wolfers (in press) might lead one to believe, the differences in forecast accuracy between the market-based and survey forecasts are minor, and, as a matter of fact, the survey forecasts tend to be more accurate for the "final" values. This result emerges from analyzing both the point and density forecasts. Hence, in light of these findings, there is little reason to prefer the market-based to survey forecasts. Second, investors in the economic derivatives market are unable to take the measurement error in the initial estimates of the BLS efficiently into account, but they seem to be forecasting the "true" non-farm payroll change, although their profits depend on accurately forecasting the first-release data. According to Bomfim's (2001) model, this lends support to the idea that improvements in the quality of first-release data would decrease economic fluctuations.

Because the market for economic derivatives has existed for only a short time, the sample considered is relatively small, and although the results are qualitatively consistent, strong statistical significance is difficult to establish. The derivatives data are also very recent, so that the even the most recent revised estimates of the non-farm payroll changes available may still deviate from the "final" values. Therefore, it would be interesting to see whether the obtained conclusions hold in an extended data set. It might also be worth studying whether the predictability in the revision process could be exploited to make excess profits in the economic derivatives market. We only considered the market for derivatives on the U.S. non-farm payroll change, but it would be interesting to see whether similar results are obtained for other economic derivatives. We leave these issues for future research.

References

Bomfim, A.N. (2001), Measurement error in general equilibrium: the aggregate effects of noisy economic indicators. Journal of Monetary Economics 48, 585–603.

Diebold, F.X., T. Gunther, and A.S. Tay (1998), Evaluating density forecasts, with applications to financial risk management. International Economic Review 39, 863–883.

Fair, R.C., and R.J. Shiller (1990), Comparing information in forecasts from econometric models. American Economic Review 80, 375–389.

Gürkaynak, R., and J. Wolfers (in press), Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk. NBER International Seminar on Macroeconomics, 2005.

Ottaviani, M., and P.N. Sørensen (2006), The strategy of professional forecasting. Journal of Financial Economics 81, 441–466.

Swanson, N.R., and D. van Dijk (2006), Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry. Journal of Business and Economic Statistics 24, 24–42.

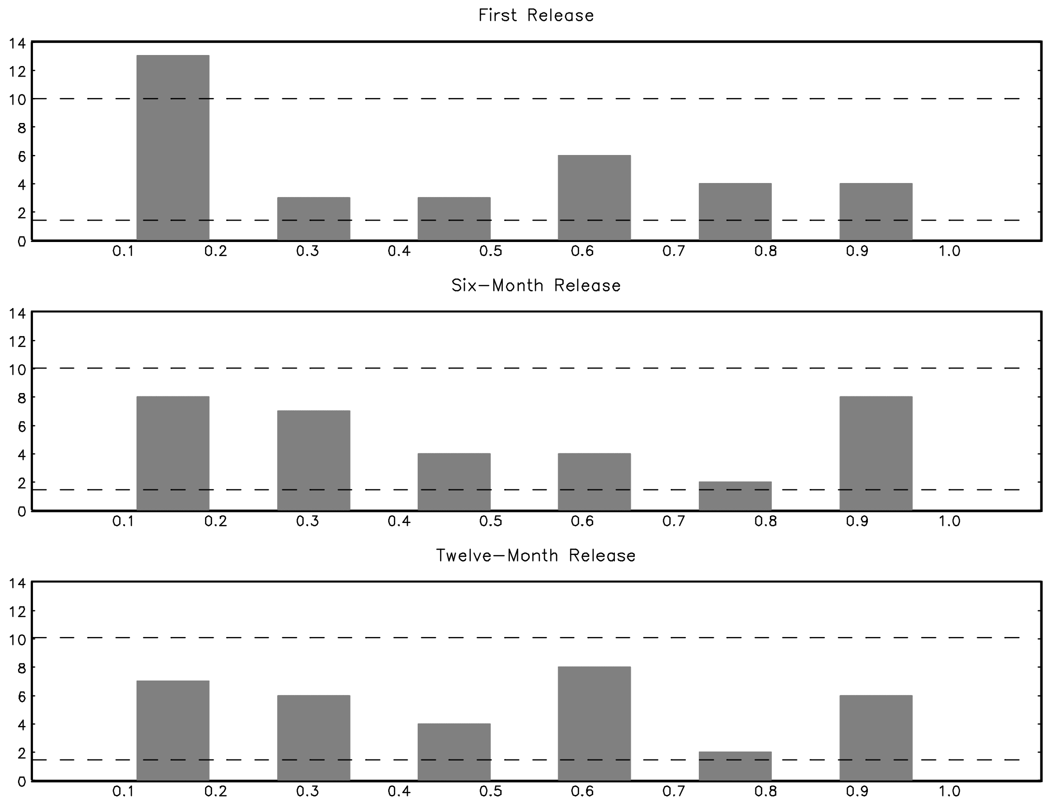

Figure 1: Probability Integral Transforms of the First, Six-Month and Twelve-Month Releases of the Non-Farm Payroll Change based on the Derivatives Data.

Table 1: First, Six-Month and Twelve-Month Releases of the Change in the Non-Farm Payroll (Thousands of New Jobs).

Date	Release			Date	Release		
	First	6-Month	12-Month		First	6-Month	12-Month
2002:10	–43	–84	65	2004:3	21	83	94
2002:11	–5	69	119	2004:4	308	353	320
2002:12	–40	1	1	2004:5	288	324	337
2003:1	–101	–211	–211	2004:6	248	208	250
2003:2	143	158	94	2004:7	112	96	106
2003:3	–308	–121	–159	2004:8	32	83	83
2003:4	–108	–151	–110	2004:9	144	188	188
2003:5	–48	–22	–20	2004:10	96	130	130
2003:6	–17	–76	–28	2004:11	337	282	282
2003:7	–30	–83	–14	2004:12	112	132	132
2003:8	–44	–45	–45	2005:1	157	155	155
2003:9	–93	–25	–25	2005:2	146	124	76
2003:10	57	67	67	2005:3	262	300	265
2003:11	126	88	88	2005:4	110	122	140
2003:12	57	83	83	2005:5	274	292	228
2004:1	1	8	8	2005:6	78	126	106
2004:2	112	159	117				

Table 2: Comparing the Accuracy and Efficiency of the Mean Forecasts for the First, Six-Month and Twelve-Month Releases of the Change in the Non-Farm Payroll.

	Release		
	First	6-Month	12-Month
	Panel A: Mean Absolute Error (MAE)		
Market	84.2 (11.3)	82.9 (9.6)	77.5 (10.2)
Survey	86.1 (11.4)	82.1 (10.1)	76.8 (10.1)
	Panel B: Root Mean Squared Error (RMSE)		
Market	105.6 (48.9)	99.1 (47.8)	96.9 (47.2)
Survey	107.6 (57.8)	100.0 (48.9)	95.2 (47.4)
	Panel C: Horse Race Regression		
	$Actual_t = \alpha + \beta Market_t + \gamma Survey_t$		
Market	0.97 (0.87)	0.57 (0.83)	0.43 (0.80)
Survey	–0.02 (0.97)	0.44 (0.92)	0.50 (0.89)
	Panel D: Efficiency Regression		
	$Forecast\ error_t = \alpha$		
Market	–33.79 (18.05)	–20.79 (17.29)	–17.48 (16.54)
Survey	–33.92 (17.68)	–20.92 (17.12)	–17.61 (16.84)

Market and Survey denote the market-based and survey forecasts, respectively. The figures in parentheses are standard errors.

Table 3: Testing the Accuracy of the Predictive Distributions implied by the Derivatives for the Non-Farm Payroll Change.

	Release		
	First	6-Month	12-Month
Pearson's Goodness-of-Fit Test	0.020	0.337	0.511
Level Autocorrelation Test	0.458	0.358	0.333
Square Autocorrelation Test	0.110	0.296	0.627

The figures are marginal significance levels.