1

Macroeconomic Derivatives: An Initial Analysis of Market-Based Macro Forecasts, Uncertainty, and Risk

Refet S. Gürkaynak, *Bilkent University and CEPR*
Justin Wolfers, *University of Pennsylvania, CEPR, IZA, and NBER*

1. Introduction

In 1993 Robert Shiller forcefully argued for the creation of a new set of securities tied to the future path of the macroeconomy. He argued that existing equity markets represent future claims on only a small fraction of future income, and that active "macro markets" would allow for more effective risk allocation, allowing individuals to insure themselves against many macroeconomic risks.

In October 2002, Goldman Sachs and Deutsche Bank set up the first markets tied directly to macroeconomic outcomes; they call these products "Economic Derivatives." These new markets allow investors to purchase options whose payoff depends on growth in non-farm payrolls, retail sales, levels of the Institute for Supply Management's manufacturing diffusion index, initial unemployment claims, and the Euro-area harmonized CPI. New U.S.-based markets have recently been created for GDP and the international trade balance, and plans are underway for securities on U.S. CPI.[1]

In this market "digital" or "binary" options are traded, allowing traders to take a position on whether economic data will fall in specified ranges, thereby providing market-based measures of investors' beliefs about the likelihoods of different outcomes. That is, the option prices can be used to construct a risk-neutral probability density function for each data release. Until the introduction of these Economic Derivatives such information was unavailable and probabilistic or density forecasts still remain quite rare.

We now have data for the first 2½ years of this market, and use these to provide an initial analysis. Given that we have only 153 data releases, many of our results will be suggestive. To preview our findings, in section 3 we find that central tendencies of market-based forecasts are very similar to, but more accurate than surveys. Further, financial market responses to data releases are also better captured by surprises measured with respect to market-based expectations than survey-based expectations, again suggesting that they better capture investor expectations. Some behavioral anomalies evident in survey-based expectations—such as forecastable forecast errors—are notably absent from market-based forecasts.

The Economic Derivatives market prices options on many different outcomes, allowing us to assess forecasts of a full probability distribution. In section 4 we compare the dispersion of the option- and survey-based distributions, and exploit the unique feature of our data that allows us to address the distinction between disagreement and uncertainty. Distributions of survey responses are measures of disagreement, or heterogeneity of beliefs, across respondents. Measuring uncertainty requires knowing how much probability agents attribute to outcomes away from the mean expectation and economic derivatives prices at different strikes provide exactly that information. Although there appears to be some correlation between disagreement and uncertainty, we find that on a release-by-release basis disagreement is not a good proxy for uncertainty. The time series of market-based measures of uncertainty also provides some evidence in favor of the view that (at least market participants believe that) non-farm payrolls and retail sales follow GARCH-like processes. In section 5 we move beyond the first and second moments of the distribution, analyzing the efficacy of these option prices as density forecasts.

While most of our analysis proceeds as if market-prices correspond one-for-one with probabilities, in section 6 we ask whether it is reasonable to expect risk aversion to drive a wedge between prices and probabilities. We find that the risk premium is in most cases sufficiently small that it can be ignored for many applications. Finally, we investigate the extent to which pricing of Economic Derivatives can provide an informative estimate of the degree of risk aversion of investors.

We view part of our contribution as simply introducing these fascinating data to the research community and thus in the next section we provide some institutional background on the details of the contracts traded, and on the market clearing mechanism.

2. The Market for Economic Derivatives

The institutional features of these new macro markets are worthy of some comment. Economic Derivatives are securities with payoffs based on macroeconomic data releases. Non-farm payrolls options, for example, settle when the employment report is released and the payrolls number is known.

The standard instruments traded are a series of digital (binary) options. The digital call (put) options pay $1 if the release is above (below) the strike. Typically around 10–20 different options are traded, each at different strike prices. Both puts and calls are traded for each data release. For transparency we will focus on the price of a "digital range"—a contract paying $1 if the announced economic number lies between two adjacent strike prices. Other types of options, such as digital puts and calls, capped vanilla options and forwards, are also traded in these markets. Each of these can be expressed as portfolios of digital ranges and are priced as such.

Figure 1 shows the prices of digital ranges from the May 12, 2005 auction (more on auctions below) which traded on what the monthly percentage change in retail sales (excluding autos) in April 2005 would be. The data was released later in the same day. Assuming risk-neutrality (which we will assume and defend in section 6), this histogram corresponds to the forecast probability distribution of the possible outcomes of this release. The mean of the distribution, the market's expectation,

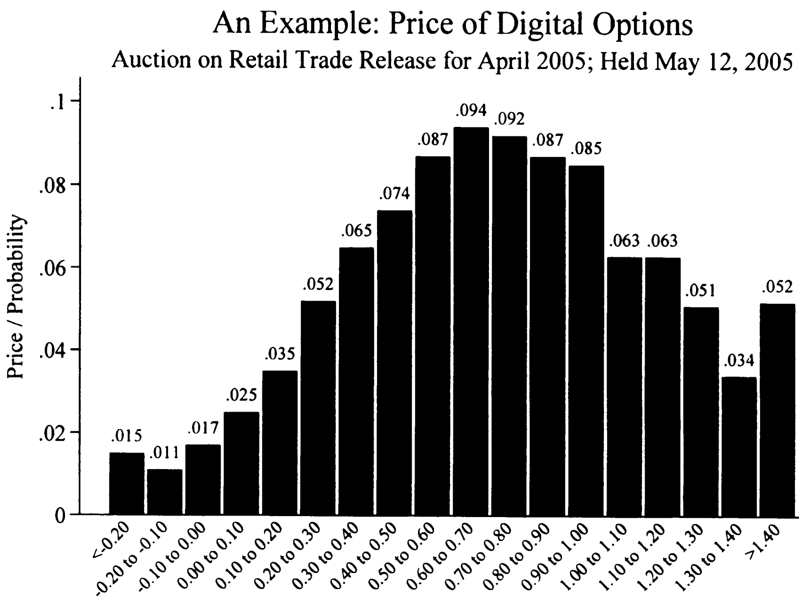

Figure 1
State-price distribution for the April 2005 retail sales release

was 0.72 percent, compared to the mean survey forecast of 0.5 percent. In the event, the released value came in at 1.07 percent, closer to the market-implied expectation. Assuming that probability is distributed uniformly within each bin, these market prices suggest that investors attributed about a 22 percent probability to the release coming in as high or higher. The major novelty of the economic derivatives market is that it allows the calculation of this implied probability.

While most financial markets operate as a continuous double auction, the market for economic derivatives is run as a series of occasional auctions, reflecting an attempt to maximize liquidity.[2] The auction mechanism is also noteworthy as it is a pari-mutuel system. That is, for a given strike price all "bets" (puts and calls) that the specified outcome either will or will not occur are pooled; this pool is then distributed to the winners in proportion to the size of their bet (the number of options purchased).[3] As such, the equilibrium price of these binary options is not known at the time the orders are made; indeed, it is only known when the last trade has occurred. Throughout the auction period (usually an hour) indicative price estimates are posted, reflecting what the price would be were no more orders to be made.

The use of pari-mutuel systems is unusual in financial markets, but common in horse race betting. Eisenberg and Gale (1959) provide useful results on the existence and uniqueness of equilibrium in such settings. The one important difference of this auction mechanism from horse race betting is that in the Economic Derivatives market it is possible to enter limit orders. This yields the possibility of multiple equilibria, which is resolved by an auction-clearing algorithm that chooses the equilibrium price vector that maximizes total trades.[4] As in traditional Dutch auctions, all trades (at a given strike) that take place are executed at the same price, regardless of the limit price.

This pari-mutuel mechanism is useful because it expands the number of ways to match buyers with sellers. While traders can be matched if one buyer's demand for calls matches another trader's demand for puts, the system does not require this. The horse track betting analogy is useful: even if nobody "sells" a given horse, as long as people bet on different horses the betting market clears. Similarly, buying a given digital range can be thought of as shorting all other outcomes and therefore having investors bidding at different strikes allows the pari-mutuel algorithm to clear the market and generate much greater volume.

In the economic derivatives market, option payoffs are determined with reference to a particular data release. Thus the payoff is based on, for example, the initial BLS estimate of growth in non-farm payrolls, rather than the best estimate of the statistical agencies (which will be subject to revision for years to come). In this sense these options provide hedges against event risk, where the events are data releases.

The events/auctions that are covered in the empirical analysis of this paper are growth of non-farm payrolls, the Institute for Supply Management manufacturing diffusion index (a measure of business confidence), change in retail sales ex-autos, and initial jobless claims. Options on GDP and trade balance releases commenced subsequent to our data collection efforts. Options on the Eurozone Harmonized Index of Consumer Prices also exist, but unfortunately we lack the high frequency financial market data for European securities required to analyze these data. Of the four markets that we do analyze, the non-farm payrolls market is the most liquid; business confidence and retail sales markets have liquidity comparable to each other but are less liquid. Initial claims options are the least liquid, however because this is a weekly release we have the largest number of observations in this market.[5]

Typically these auctions have taken place in the morning of the data release and they were sometimes preceded by another auction on the same release one or two days prior (non-farm payrolls auctions are held on both the morning the data are released and one day before).[6,7] Thus economic derivatives provide hedging opportunities against only very high frequency movements—event risk—and really cannot be said to provide the sorts of business cycle frequency risk-sharing opportunities envisioned by Shiller (1993). We return to a more careful assessment of the role of risk in these markets in section 6. But first we focus on the uses of market prices as forecasts.

3. The Accuracy of Market-based Forecasts

We begin by comparing forecasts generated by the Economic Derivatives market with an alternative information aggregator, the "survey forecast" released by Money Market Services (MMS) on the Friday before a data release.[8] Specifically, we compare the mean forecast from each mechanism, although our results are insensitive to the choice of mean versus median forecasts. For the MMS forecast, the "consensus" forecast typically averages across around 30 forecasters. For the market-based forecast, we aggregate across the distribution of outcomes and calculate the distribution's mean assuming that the probability distribution is uniform within each bin (boundaries of bins are defined by adjacent strikes).[9] As such, we implicitly assume that the price of a digital option is equal to the average belief that the specified outcome occurs. Wolfers and Zitzewitz (2005) discuss the relationship between prediction market prices and beliefs. We return to this issue in later sections, showing that ignoring risk aversion does very little violence to the data.

Figure 2 shows the relative forecasting performance of the survey- and market-based forecasts. Visual inspection suggests that the market-based forecast mildly dominates the survey forecast, a fact verified formally in Table 1.

Table 1 examines two specific measures of forecast accuracy: the mean absolute error and the root mean squared error, contrasting the performance of the Economic Derivatives market and the survey respondents. Each column reports these summary statistics for a different data series. In order to provide some comparability of magnitudes across columns we normalize the scale of each by dividing our measures of forecast errors by the historical standard deviation of survey forecast

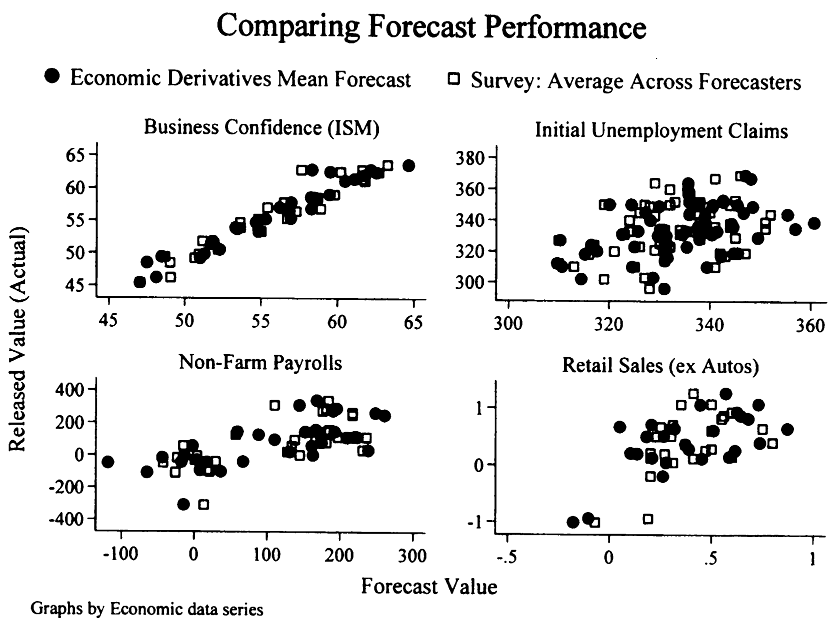

Figure 2
Comparing forecast performance

Table 1
Comparing the accuracy of mean forecasts

	Non-farm payrolls	Business confidence (ISM)	Retail sales (ex autos)	Initial unemployment claims	Pooled data
	Panel A: Mean Absolute Error				
Economic derivatives	0.723	0.498	0.919	0.645	0.680
	(.097)	(.090)	(.123)	(.061)	(.044)
Survey	0.743	0.585	0.972	0.665	0.719
	(.098)	(.093)	(.151)	(.063)	(.046)
	Panel B: Root Mean Squared Error				
Economic derivatives	0.907	0.694	1.106	0.808	0.868
	(.240)	(.257)	(.262)	(.126)	(.102)
Survey	0.929	0.770	1.229	0.831	0.921
	(.268)	(.296)	(.364)	(.130)	(.124)
	Panel C: Correlation of Forecast with Actual Outcomes				
Economic derivatives	0.700	0.968	0.653	0.433	0.631
	(.126)	(.047)	(.151)	(.114)	(.063)
Survey	0.677	0.961	0.544	0.361	0.576
	(.130)	(.052)	(.168)	(.117)	(.066)
	Panel D: Horse Race Regression (Fair-Shiller)				
	$Actual_t = \alpha + \beta * Economic\ Derivatives_t + \gamma * Survey\ Forecast_t$ (+*survey fixed effects*)				
Economic derivatives	1.06	0.91**	1.99**	1.64***	1.25***
	(0.78)	(.37)	(.79)	(.60)	(.29)
Survey	–0.14	0.17	–1.03	–1.21*	–0.24
	(0.89)	(.38)	(1.10)	(.68)	(.30)
Adjusted R^2	0.46	0.93	0.40	.20	.99
Sample size (Oct. 2002–Jul. 2005)	33	30	26	64	153

Notes: Forecast errors normalized by historical standard error of survey-based forecasts. (Standard errors in parentheses.) ***, **, and * denote statistically significant regression coefficients at 1 percent, 5 percent, and 10 percent, respectively.

errors over an earlier period.[10] Thus, the units in the table can be read as measures of forecast errors relative to an historical norm. This scaling makes the magnitudes sufficiently comparable that we can pool our observations across data series in the final column.

Comparing the two rows of Panel A shows that the market-based forecasts errors were on average smaller than the survey forecasts for all four data series. To interpret the magnitudes, start by noting that in all cases the estimates are less than one, implying that both sets of forecasts were more accurate than the survey forecast had been over the pre-2002 period. Beyond this, the improvements in forecast accuracy are meaningful, if not huge. For instance, pooling all of the data shows that relying on market-based forecasts rather than survey forecasts would have reduced the size of forecast errors by 0.04, which by virtue of the scaling is equivalent to 5½ percent of the average forecast error over the preceding decade. While meaningful, this reduction is not statistically significant. Panel B shows that analyzing the root mean squared error yields roughly similar results. In Panel C we compare the correlation of each forecast with actual outcomes. (Naturally these correlations can also be interpreted as the coefficient from a regression of standardized values of the outcome on standardized values of the forecast.) Each of these coefficients is statistically significant, suggesting that each forecast has substantial unconditional forecasting power. Even so, the market-based forecast is more highly correlated with outcomes than the consensus forecast for all four data series.

Panel D turns to a regression-based test of the information content of each forecast following Fair and Shiller (1990). Naturally there is substantial collinearity, as the market- and consensus-based forecasts are quite similar. Even so, we find rather compelling results. A coefficient of unity for the market-based forecast cannot be rejected for any of the indicators. By contrast, conditioning on the market-based forecast renders the survey forecast uninformative, and in three of four cases the survey-based forecast is not statistically different from zero and in the one case in which it is significant, it has a perverse negative coefficient. In the final column we pool the forecasts to obtain more precise estimates and again the market-based forecast dominates, and this difference is both statistically and economically significant.

These findings are probably partly due to the fact that the economic derivatives auction occurs on the morning of the data release, while the survey takes place up to a week before. Thus, option prices incorporate more information than was available to survey respondents. In an attempt to partly ameliorate this information advantage, we also reran our regressions in Panel D, controlling for two indicators of recent economic news: the change in equity prices and bond yields between the market close on the night prior to the release of the survey data to the night before the economic derivatives auction. These indicators for the release of relevant news were typically insignificant, and our main conclusions were not much altered by this control.

It seems likely that the improved performance is due to the market effectively weighting a greater number of opinions, or more effective information aggregation as market participants are likely more careful when putting their money where their mouth is.

We next ask which forecast aggregator better predicts the financial market reactions to the release of economic statistics. Or alternatively phrased, we ask: which forecast best embeds the forecasts of the equity and bond markets? In Figures 3A and 3B we show the short-term change in the S&P 500 and the 10-year Treasury note yield that result from the release of economic news. The solid dots measure the innovation as the deviation of the announced economic statistic from the economic derivatives forecast, while the hollow squares represent the innovation as the deviation from the consensus forecast.

Table 2 formalizes the comparisons in Figures 3A and 3B. Specifically, we run regressions of the form:

$$\Delta Financial\ variable_t = \alpha + \beta^* \left(Actual_t - Forecast_t^{Economic\ Derivs} \right) + \gamma^* \left(Actual_t - Forecast_t^{Survey} \right).$$

We measure changes in stock and Treasury markets around a tight window, comparing financial market quotes five minutes prior to the data release to 25 minutes after the event.[11] We analyze changes in implied Treasury yields, rather than changes in their prices, and report these changes in basis points; the stock market response is reported as percentage change. As before, we rescale our forecast error variables so that the estimates can be interpreted as the effect of a one-standard deviation forecast error.

Several patterns emerge in these data. First, comparing columns suggests that the non-farm payrolls release has the largest effect on financial markets; retail trade and business confidence are also important, but the weekly initial claims data rarely moves markets by much. Comparing panels shows that the yields on longer-dated securities more reliably and more forcefully respond to the release of these economic statistics than do yields on short-term Treasury bills. It is likely that short-term interest rate expectations have been strongly anchored by Federal Reserve statements recently, reducing the sensitivity of short-term yields to data release surprises. The stock market also responds quite vigorously to non-farm payrolls.[12] Lastly, comparing rows within each panel, financial markets appear to respond to economic data to the extent that they differ from the Economic Derivatives forecast; conditioning on this, the survey forecast has no statistically significant explanatory power in any individual regression.

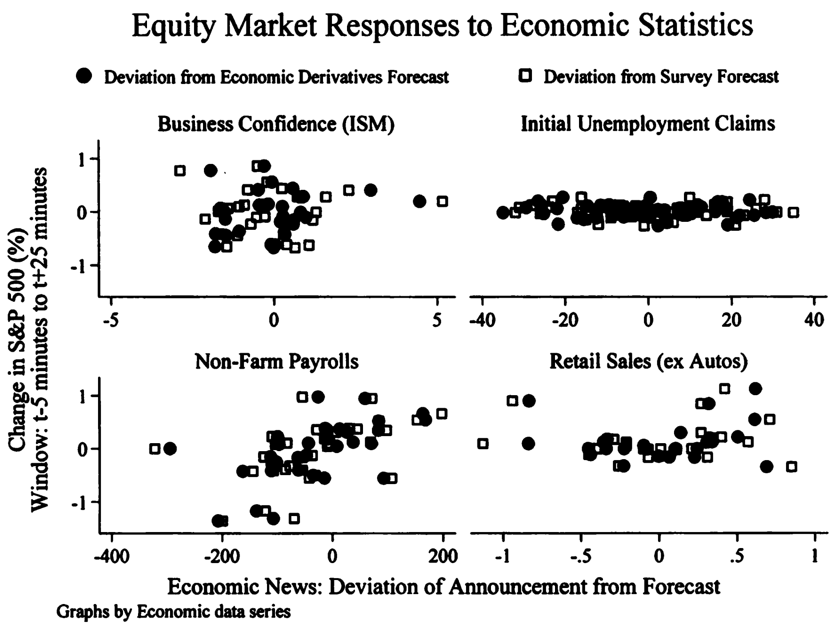

Figure 3A
Equity market responses to surprises

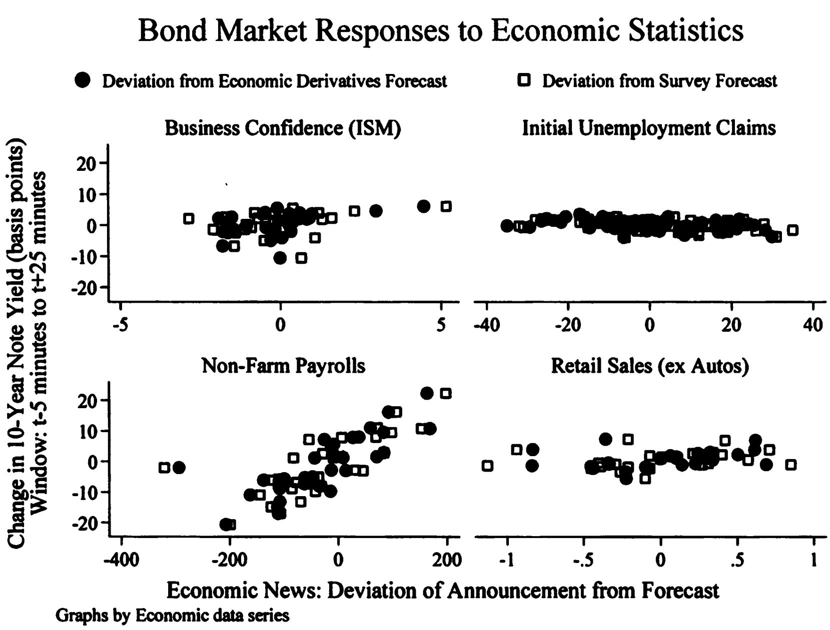

Figure 3B
Bond market responses to surprises

To maximize our ability to test the joint significance across columns, we pool our data across all four economic series and run:

$$\Delta Financial\ variable_t = \sum_{s \in Economic\ series} \alpha_s + \beta_s * \left(Actual_{s,t} - Forecast_{s,t}^{Consensus}\right)$$
$$+ \gamma_s * \left(Actual_{s,t} - Forecast_{s,t}^{Survey}\right).$$

The final column of Table 2 reports the joint statistical significance of the β's and the γ's, respectively. These joint tests clearly show that financial markets respond to the innovation as measured relative to the Economic Derivatives forecast and conditional on this, appear not to respond to the deviation of the data from the survey forecast.

In sum, Tables 1 and 2 establish that the Economic Derivatives forecast dominates the survey forecast (although survey forecasts perform quite well) both in predicting outcomes and in predicting market responses to economic news. Many previous papers have demonstrated that professional forecasters exhibit a range of predictable pathologies. For instance, Mankiw, Reis, and Wolfers (2003) analyze data on inflation expectations from the Survey of Professional Forecasters and the Livingstone Survey, finding that the median forecast yielded errors that were predictable based on recent economic developments, past forecast errors, or even the forecast itself. Were similar results to persist in the Economic Derivatives market, these predictable forecast errors would yield profitable trading opportunities.

In Table 3 we repeat many of the tests in that earlier literature, asking whether forecast errors are predictable based on a long-run bias (Panel A), on information in the forecast itself (Panel B), on previous forecast errors (Panel C), or on recent economic news (Panel D). We test the efficiency of the survey forecast and the Economic Derivatives forecasts separately, thus each cell in the table represents a separate regression. As before, we rescale the forecast errors by the historical standard deviation of the survey forecast errors for each indicator.

Each regression in Table 3 asks whether forecast errors are predictable; each panel tests different sets of predictors, and each column performs the test for a different economic indicator. The final column provides a joint F-test that the forecast errors are not predictable, aggregating across all four economic indicators in each row. In each succeeding panel we ask whether each forecast yields predictable on the basis of a simple constant term (Panel A), information in the forecast itself (Panel B), based on the forecast error from the previous month (Panel C), or based on recent economic information (Panel D).[13] Only

Table 2
Predicting market responses to economic statistics

	Non-farm payrolls	Business confidence (ISM)	Retail sales (ex autos)	Initial unemployment claims	Joint significance (F-test)
$\Delta Financial\ variable_t = \alpha + \beta * (Actual_t - Forecast_t^{Economic\ Derivs}) + \gamma * (Actual_t - Forecast_t^{Survey})$					
	Panel A: 3 Month Treasury Bill				
Economic derivatives	4.41**	0.428	–0.094	–0.087	p=.0006
	(1.71)	(.434)	(.491)	(.601)	
Survey	–2.50	–0.166	0.067	–0.123	p=.1374
	(1.66)	(.396)	(.442)	(.585)	
	Panel B: 6 Month Treasury Bill				
Economic derivatives	6.21**	1.034	0.221	–1.294	p=.0004
	(2.40)	(.786)	(.751)	(.785)	
Survey	–3.47	–0.483	–0.054	0.976	p=.1184
	(2.33)	(.769)	(.675)	(.764)	
	Panel C: 2 Year Treasury Note				
Economic derivatives	12.61**	3.96*	2.60	–1.40	p=.0016
	(6.04)	(1.98)	(2.16)	(1.15)	
Survey	–2.50	–1.71	–1.73	0.42	p=.7841
	(5.87)	(1.79)	(1.94)	(1.11)	
	Panel D: 5 Year Treasury Note				
Economic derivatives	14.94**	5.54**	3.66	–3.17**	p=.0001
	(6.39)	(2.07)	(2.44)	(1.22)	
Survey	–3.90	–2.56	–2.53	2.06*	p=.4254
	(6.21)	(1.86)	(2.19)	(1.19)	
	Panel E: 10 Year Treasury Note				
Economic derivatives	10.40*	5.09**	3.37	–2.12*	p=.0007
	(5.22)	(1.90)	(2.04)	(1.12)	
Survey	–1.64	–2.53	–2.36	1.22	p=.4955
	(5.07)	(1.71)	(1.83)	(1.09)	
	Panel F: S&P 500				
Economic derivatives	0.888**	0.575**	0.434*	–.106	p=.0001
	(.386)	(.226)	(.252)	(.084)	
Survey	–0.514	–0.466**	–0.367	0.092	p=.0058
	(.375)	(.204)	(.227)	(.082)	

Notes: Dependent variables normalized by historical standard error of survey-based forecasts. (Standard errors in parentheses) ***, **, and * denote statistically significant at 1 percent, 5 percent, and 10 percent.
For sample size, see Table 1.

Table 3
Tests of forecast efficiency

	Non-farm payrolls	Business confidence (ISM)	Retail sales (ex autos)	Initial unemployment claims	Joint significance (F-test)
		Panel A: Bias $Forecast\ error_t = \alpha$			
Economic derivatives	−0.29* (.15)	−0.03 (.13)	0.04 (.22)	−0.04 (.10)	p=.419
Survey	−0.29* (.16)	−0.06 (.14)	0.03 (.25)	0.05 (.10)	p=.371
		Panel B: Internal Efficiency $Forecast\ error_t = \alpha + \beta{*}Forecast_t$ [Square brackets shows test α=β=0]			
Economic derivatives	−0.049 (.174) [p=.161]	−0.078 (.053) [p=.345]	−0.309 (.310) [p=.604]	−0.371** (.167) [p=.031]	p=.182
Survey	0.043 (.204) [p=.196]	0.095 (.059) [p=.273]	0.512 (.476) [p=.564]	−0.398** (.197) [p=.127]	p=.173
		Panel C: Autocorrelation $Forecast\ error_t = \alpha + \rho{*}Forecast\ error_{t-1}$			
Economic derivatives	−0.091 (.183)	−0.008 (.191)	−0.383* (.188)	0.002 (.128)	p=.186
Survey	−0.078 (.183)	0.142 (.190)	−0.500** (.180)	−0.074 (.128)	p=.016
		Panel D: Information Efficiency $Forecast\ error_t = \alpha + \beta{*}Slope\ of\ yield\ curve_{t-1} + \gamma{*}\Delta S\&P\ 500_{t-1,t-10}$ [Square brackets shows *test* β=γ=0]			
Economic derivatives	β=−0.100 (.229) γ=0.051 (.060) [p=.640]	β=0.287 (.186) γ=−0.039 (.054) [p=.241]	β=0.078 (.322) γ=−0.073 (.094) [p=.735]	β=0.102 (.121) γ=−0.012 (.053) [p=.677]	p=.800
Survey	β=−0.031 (.237) γ=0.046 (.063) [p=.759]	β=0.390* (.201) γ=−0.043 (.059) [p=.127]	β=0.132 (.359) γ=−0.076 (.105) [p=.737]	β=0.137 (.123) γ=−0.018 (.054) [p=.502]	p=.672
		Panel E: Joint Test of All Predictors (p-value of joint significance) $Forecast\ error_t = \alpha + \beta{*}Survey\ Forecast_t + \beta_2{*}Market\ Forecast_t + \beta_3{*}Forecast\ error_{t-1} + \beta_4{*}Slope\ of\ yield\ curve_{t-1} + \beta_5{*}\Delta S\&P\ 500_{t-1,t-10}$			
Economic derivatives	p=.900	p=.129	p=.228	p=.015	p=.0664
Survey	p=.625	p=.036	p=.017	p=.004	p=.0003

Notes: Each cell represents a separate regression.
Dependent variables normalized by historical standard deviation of survey-based forecasts.
(Standard errors in parentheses) ***, ** and * denote statistically significant at 1 percent, 5 percent, and 10 percent.

Panel C seems to show evidence of behavioral biases, with the survey-based forecast yielding significantly negatively autocorrelated forecast errors, particularly for retail sales. Equally we should not overstate this result: while we cannot reject a null that market-based forecasts are efficient, we also cannot reject a null that they show the same pattern of predictable forecast errors as the survey-based forecasts.

Finally in Panel E we combine each of the above tests, testing whether forecast errors are predictable based on the full set of possible predictors (including both the market- and survey-based forecasts themselves). On this score the superior performance of the market-based forecasts is much more evident. The survey-based forecasts yield predictable forecast errors for three of the four statistical series; not surprisingly, the survey does best on non-farm payrolls, which is the most closely watched of these numbers. The market-based forecasts show no such anomalies except in the case of initial claims, which is easily the least liquid of these markets. Overall these results confirm the results in the earlier behavioral literature documenting anomalies in survey-based forecasts. Equally, they suggest that such inefficiencies are either absent, or harder to find in market-based forecasts.

This section compared the mean forecast from surveys and economic derivatives, with the basic finding that while surveys do well (despite some behavioral anomalies), markets do somewhat better in forecasting. If one is only interested in forecasting the mean, using surveys might suffice; however, Economic Derivatives provide a lot more information than just the mean forecast. Observing that the mean of the market-based probability distribution "works" the way it should is comforting and holds promise for the information content of the higher moments of the distribution, the subject of the next section.

4. Disagreement and Uncertainty

We now turn to analyzing the standard deviation of the state-price distribution. We will refer to this standard deviation as "uncertainty," reflecting the fact that this is the implied standard error of the mean forecast. Table 4 compares the market's average assessment of uncertainty with the realized root-mean-squared error of both the market- and survey-based forecasts over the same period. These results suggest that the market-based measure of uncertainty is reasonably well calibrated. We also include a third comparison: estimates by the official statistical agencies of the standard error of their measurements of these

Table 4
Expectations and realizations of forecast accuracy

RMSE of Forecasts (or standard deviation of forecast error)	Non-farm payrolls	Business confidence (ISM)	Retail sales (ex autos)	Initial unemployment claims
Expectations				
Market-implied standard deviation	96.1	2.01	0.44	12.5
Realizations				
SD of market forecast errors	100.7	1.40	0.42	15.1
SD of survey forecast errors	103.7	1.55	0.46	15.5
Sampling error				
Standard error of official estimate	81.5	n.a.	0.5	n.a.

Note: For estimates of the standard errors of the official estimates, see Wolfers and Zitzewitz (2004, p. 115).

economic statistics, where available. Market expectations of the RMSE of forecast errors are only slightly larger than sampling error in the case of non-farm payrolls, and slightly smaller in the case of retail sales.

Explicit measures of uncertainty are rare in macroeconomics, so we compare this market-based measure with the standard deviation of point forecasts across forecasters, and following Mankiw, Reis, and Wolfers (2003), we refer to the latter as "disagreement." The (previous) absence of useful data on uncertainty had led many researchers to analyze data on disagreement as a proxy for uncertainty. To date there has been very little research validating this approach, and indeed the only other measure of uncertainty we are aware of (from the Survey of Professional Forecasters) shows only weak comovement with measures of disagreement (Llambros and Zarnowitz 1987).

Figure 4 shows results consistent with Llambros and Zarnowitz: disagreement and uncertainty comove, but the correlation is not strong. The obvious difference in the levels is due to the fact that central expectations of respondents are close to each other even when each respondent is uncertain of their estimate.

In Table 5 we analyze these relationships a little more formally, regressing uncertainty against disagreement. Panel A shows that there is a statistically significant positive correlation between disagreement

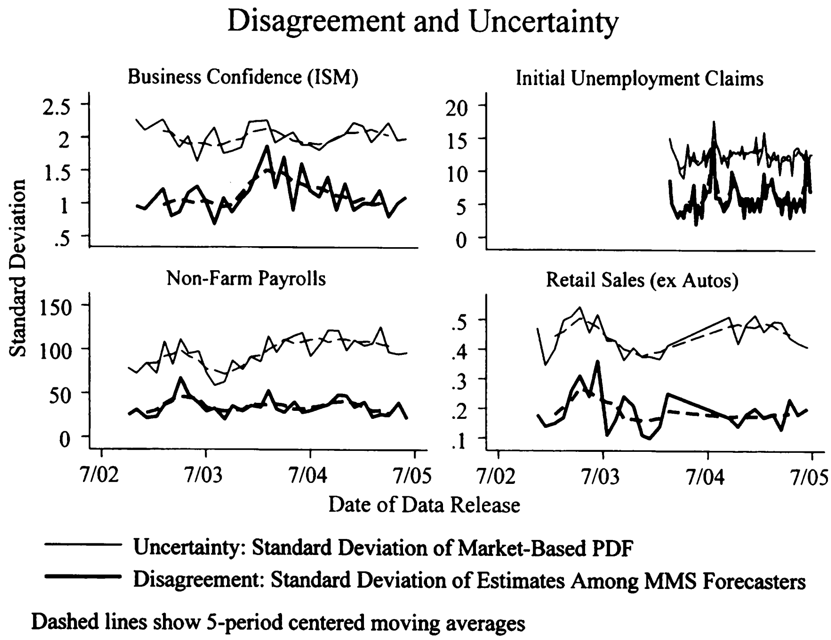

Figure 4
Disagreement and uncertainty

and uncertainty for all series except ISM. The final column shows the joint significance of the coefficients on disagreement, suggesting that the contemporaneous relationship is quite strong. Indeed, Chris Carroll has suggested that one can interpret these regressions as the first stage of a split-sample IV strategy, allowing researchers to employ disagreement as a proxy for uncertainty in another dataset. This, of course, depends on how high an R^2 one views as sufficient in the first stage regression.

Panel B of this table carries out a similar exercise focusing on lower-frequency variation. In this case, disagreement and uncertainty are still correlated but this correlation is substantially weaker. The 5-period moving average of disagreement is a significant explanator of the 5-period moving average of uncertainty only for retail sales and initial claims. (Even this overstates the strength of the relationship, as we do not correct the standard errors for the autocorrelation generated by smoothing.) Jointly testing the significance across all four indicators we find that the relationship between low frequency variation in disagreement and uncertainty is not statistically significant, and the R^2s

Table 5
Disagreement and uncertainty

	Non-farm payrolls	Business confidence (ISM)	Retail sales (ex autos)	Initial unemployment claims	Joint Significance (F-test)
	Panel A: Contemporaneous Relationship $Uncertainty_t = \alpha + \beta^* Disagreement_t$				
Disagreement	0.66**	–0.03	0.44**	0.27***	p=.0002
	(.29)	(.12)	(.16)	(.07)	
Constant	73.6	2.04	0.36	10.86	
	(10.39)	(.134)	(.03)	(.47)	
Adjusted R^2	0.11	–0.03	0.20	0.17	
	Panel B: Low Frequency – 5 Period Centered Moving Averages $Smoothed\ Uncertainty_t = \alpha + \beta^* Smoothed\ Disagreement_t$				
Disagreement	0.55	0.10	0.65**	0.32***	p=.1498
	(.47)	(.10)	(.24)	(.06)	
Constant	77.7	1.89	0.32	10.5	
	(16.8)	(.11)	(.05)	(.37)	
Adjusted R^2	0.01	–0.002	0.23	0.32	

Notes: (Standard errors in parentheses) ***, **, and * denote statistically significant at 1 percent, 5 percent, and 10 percent.

of these regressions are again sufficiently low and varied as to caution that disagreement might be a poor proxy for uncertainty in empirical applications.

Having demonstrated fairly substantial time series variation in uncertainty (albeit over a short period) naturally raises the question: What drives movements in uncertainty?

In Panel A of Table 6 we look to see whether any of the variation is explained by movements in expected volatility of equity markets. That is, our regressors include the closing price of CBOE's VIX index on the day prior to the economic derivatives auction, as well as the closing price one and two months prior (for the initial claims, these lags refer to one and two weeks earlier). As in Tables 1–3, we rescale the uncertainty measure by the standard deviation of historical forecast errors to allow some comparability across columns. Panel A shows that for all four indicators the contemporaneous values of the implied volatility index is uncorrelated with uncertainty about forthcoming economic data. While a couple of specific lags are statistically significant, they suggest a somewhat perverse negative correlation between

Table 6
Modeling uncertainty

	Non-farm payrolls	Business confidence (ISM)	Retail sales (ex autos)	Initial claims
	Panel A: Uncertainty and Expected Volatility			
	$Uncertainty_t = \alpha + \beta_1{*}VIX_t + \beta_2{*}VIX_{t-1} + \beta_3{*}VIX_{t-2}$			
VIX_t	0.76	0.41	0.04	0.10
	(.95)	(.72)	(1.07)	(.86)
VIX_{t-1}	−1.93**	0.79	1.15	−0.44
	(.86)	(.69)	(1.27)	(1.04)
VIX_{t-2}	0.23	−1.01*	−0.93	−0.22
	(.80)	(.57)	(.98)	(.85)
Joint sig?	p=0.02	p=0.31	p=0.73	p=0.80
Adjusted R^2	0.21	0.02	−0.07	−0.03
	Panel B: Persistence			
	$Uncertainty_t = \alpha + \beta_1{*}Uncertainty_{t-1} + \beta_2{*}Uncertainty_{t-2} + \beta_3{*}Uncertainty_{t-3}$			
$Uncertainty_{t-1}$	0.34*	0.24	0.43*	0.20
	(.19)	(.19)	(.23)	(.13)
$Uncertainty_{t-2}$	0.37*	−0.26	0.14	0.01
	(.19)	(.20)	(.23)	(.13)
$Uncertainty_{t-3}$	−0.12	0.11	−0.13	−0.24*
	(.19)	(.19)	(.21)	(.13)
Joint sig?	p=0.02	p=0.45	p=0.14	p=0.10
Adjusted R^2	0.24	−0.01	0.12	0.06
	Panel C: Pseudo-GARCH Model			
	$Uncertainty_t = \alpha + \beta_1{*}Uncertainty_{t-1} + \beta_2{*}Uncertainty_{t-2} + \beta_3{*}Uncertainty_{t-3} + \gamma_1{*}Forecast\ Error_{t-1}^2 + \gamma_2{*}Forecast\ Error_{t-2}^2 + \gamma_3{*}Forecast\ Error_{t-3}^2$			
$Uncertainty_{t-1}$	0.37*	0.21	0.47*	0.16
	(.21)	(.22)	(.25)	(.13)
$Uncertainty_{t-2}$	0.38	−0.12	−0.10	0.02
	(.22)	(.23)	(.25)	(.13)
$Uncertainty_{t-3}$	−0.13	0.05	0.12	−0.20
	(.19)	(.20)	(.24)	(.12)
Joint sig?	p=0.01	p=0.82	p=0.28	p=0.26
F'cast $error_{t-1}^2$	0.05**	0.02	0.05**	0.03**
	(.02)	(.02)	(.03)	(.01)
F'cast $error_{t-2}^2$	0.02	−0.02	−0.03	0.01
	(.02)	(.02)	(.03)	(.01)
F'cast $error_{t-3}^2$	−0.01	−0.00	−0.00	−0.00
	(.02)	(0.02)	(.02)	(.01)
Joint sig?	p=0.05	p=0.41	p=0.21	p=0.11
Adjusted R^2	0.38	−0.009	0.21	0.11
n [Panel A, B/C]	[33,30]	[30,27]	[26,23]	[64,61]

Notes: (Standard errors in parentheses) ***, **, and * denote statistically significant at 1 percent, 5 percent, and 10 percent. VIX_t refers to the close of CBOE's VIX index on the day prior to the auction. VIX_{t-1} refers to the day prior to the previous data release. $Uncertainty_{t-1}$ refers to the standard deviation of the state price distribution for the previous data release in that series. All of the uncertainty measures are rescaled by the historical standard deviation of forecast errors for that series.

uncertainty and expected volatility in the stock market. This lack of correlation likely suggests that uncertainty is usually not about the fundamental state of the economy but about the particular data release—perhaps because the seasonal factors are sometimes more difficult to forecast.

Panel B also examines the persistence of uncertainty, and uncertainty about non-farm payrolls and retail sales appears to show some degree of persistence. Finally Panel C jointly tests whether uncertainty is a product of both past uncertainty and past realizations, as posited in GARCH models. Market assessments of the uncertainty in non-farm payrolls, retail sales, and initial claims appears to be well-described by these variables, although we find no such evidence for ISM.[14] Finally we ask whether these market-based measures of uncertainty actually predict the extent of forecast errors.

Figure 5 seems to suggest that uncertainty is not strongly related to larger (absolute) forecast errors (note that these forecast errors are standardized by their historical standard errors). We perform a more formal test in Table 7. If the uncertainty measure is appropriately calibrated, we should expect to see a coefficient of one in the regression of absolute forecast errors on uncertainty.

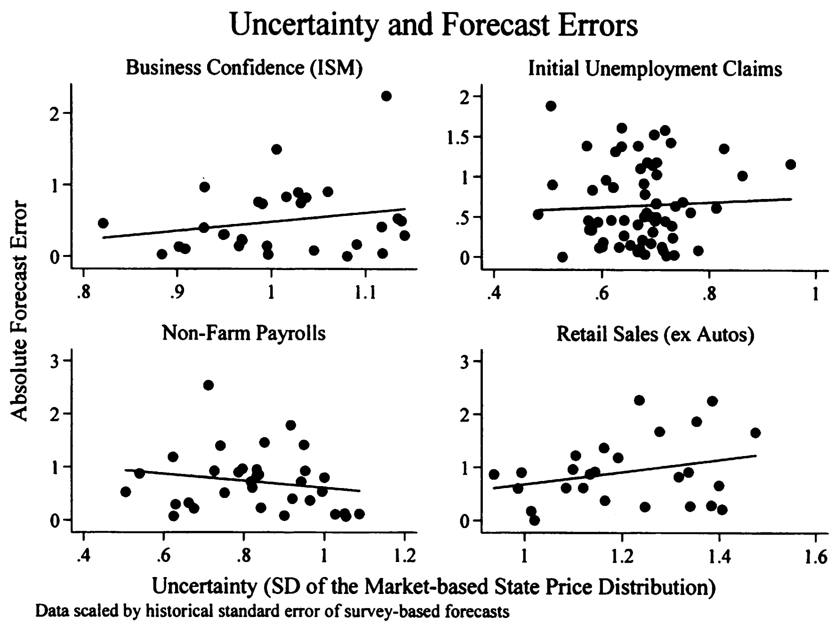

Figure 5
Uncertainty and forecast errors

Table 7
Uncertainty and forecast errors

	Non-farm payrolls	Business confidence	Retail trade (ex autos)	Initial claims	Joint Significance (F-test)
	*Absolute Forecast Error$_t$ = $\alpha + \beta$*Uncertainty$_t$*				
Uncertainty (β)	–0.65 (0.64)	1.27 (1.08)	1.16 (0.80)	0.31 (.77)	p=0.26
Test: β=0 (No information)	p=0.32	p=0.25	p=0.16	p=0.69	
Test: β=1 (Efficient forecast)	p=0.02	p=0.81	p=0.84	p=0.37	p=0.09

Notes: (Standard errors in parentheses)
***, **, and * denote statistically significant at 1 percent, 5 percent, and 10 percent, respectively.
Forecast errors normalized by historical standard error of survey-based forecasts.

Overall Table 7 suggests that these tests have very little power. In no individual case is the absolute forecast error significantly correlated with the market-based measure of uncertainty. The final column pools the data, again finding no evidence of a significant correlation. That is, the data cannot reject the null that there is no information in the time series variation in market-based uncertainty that helps predict time series variation in forecast errors. On the other hand, the estimates are imprecise enough that, as the second row shows, we cannot reject a coefficient of unity for three out of the four series either.

Of course the object of interest in these regressions—the standard deviation of the state price distribution—is a summary statistic from a much richer set of digital options or density forecasts, and so we will obtain greater power in the next section as we turn to analyzing these density forecasts more directly.

5. Full Distribution Implications

A particularly interesting feature of the Economic Derivatives market is that it yields not only a point estimate, but also a full probability distribution across the range of plausible outcomes. Exploiting this, we can expand our tests beyond section 3, which asked whether the mean forecast is efficient, to also ask whether the prices of these options yield efficient forecasts of the likelihood of an economic statistic falling in a given range.

Figure 6 provides an initial analysis, pooling data from all 2,235 digital call options (contracts that pay $1 if the announced economic statistic is above the strike price) across our 153 auctions. We grouped these options according to their prices, and for each group we show the proportion of the time that the economic statistic actually is above the strike price. These data yield a fairly close connection, and in no case do we see an economically or statistically significant divergence between prices and probabilities.

While the evidence in Figure 6 suggests that the Economic Derivatives prices are unbiased, it does not speak to the efficiency of these estimates, an issue we now turn to. Because density estimates are hard to come by (see Diebold, Tay, and Wallis 1999 for an example), the forecast evaluation literature has focused on evaluating point forecasts rather than densities. An intermediate step between point and density estimate evaluation is interval forecast evaluation. An interval forecast is a confidence interval such as "non-farm payrolls will be between 100,000 and 180,000 with 95 percent probability." Christoffersen (1998) shows that a correctly conditionally calibrated interval forecast will provide a hit sequence (a sequence of correct and incorrect predictions) that is independently and identically Bernoulli distributed with the desired coverage probability. A

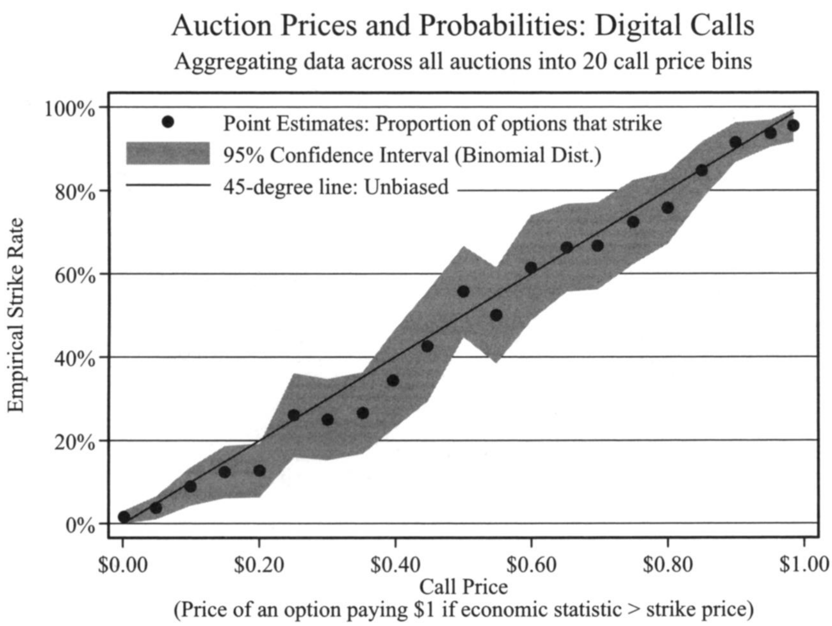

Figure 6
Prices and probabilities – digital call options

density forecast can be thought of as a collection of interval forecasts, and Diebold, Gunther, and Tay (1998) show that the i.i.d. Bernoulli property of individual interval forecasts translates into the i.i.d. uniform (0,1) distribution of the probability integral transform, z_t, defined as

$$z_t = \int_{-\infty}^{y_t} \pi(x)dx \overset{iid}{\sim} Uniform(0,1)$$

where $\pi(x)$ denotes the price of an option paying \$1 if the realized economic statistic takes on the value x, and y_t is the actual realized value of economic statistic. Thus z_t can be thought of as the "realized quantile," and the implication that this should be uniformly distributed essentially formalizes the argument that if the prediction density is correct, the "x" percent probability event should be happening "x" percent of the time. In the data we do not observe exact state-prices $\pi(x)$, but rather digital ranges, $\int_a^b \pi(x)dx$; to estimate the realized quantile we simply assume that $\pi(x)$ is uniformly distributed within each strike-price range.

In Figure 7 we calculate the realized quantile for each auction, pool the estimates across different economic statistics, and plot the relevant

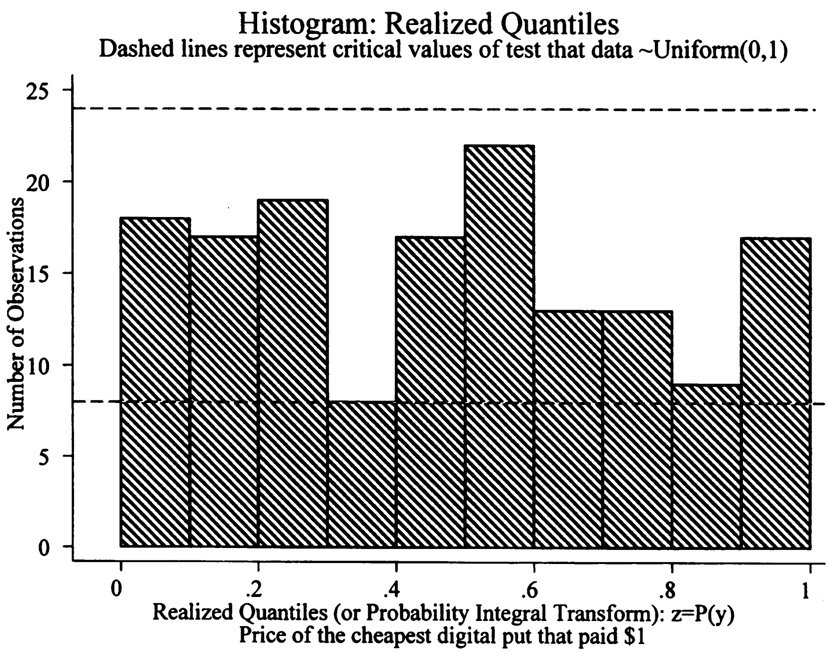

Figure 7
Histogram of realized quantiles

histogram. A simple way to test for deviations from uniformity derives from inverting the earlier logic: if the distribution is uniform, then the probability that any given realization is in any given bin should follow a Bernoulli distribution with the hit probability equal to the width of the bin, and hence the number of realizations in each bin should follow a binomial distribution. Thus in Figure 7 we show the relevant 95 percent critical values under the assumption of i.i.d. uniformity.

Figure 7 shows that the distribution is generally close to uniform, albeit with a peak around 0.5, which is suggestive of excess realizations close to the median forecast. That said, this distribution is statistically indistinguishable from a uniform distribution.[15]

The inference in this figure is partly shaped by the specific bin widths chosen for the histogram. Figure 8 shows an alternative representation, mapping both the entire cumulative distribution function of the probability integral transform and the uniform distribution. The figure also shows the deviations from uniformity that would be required for a Kolmogorov-Smirnov test to reject a null that the realized quantiles are drawn from a uniform distribution. As seen, this suggests that the data are fairly close to an idealized uniform (0,1) distribution, and that these data yield no statistically significant evidence falsifying this null.

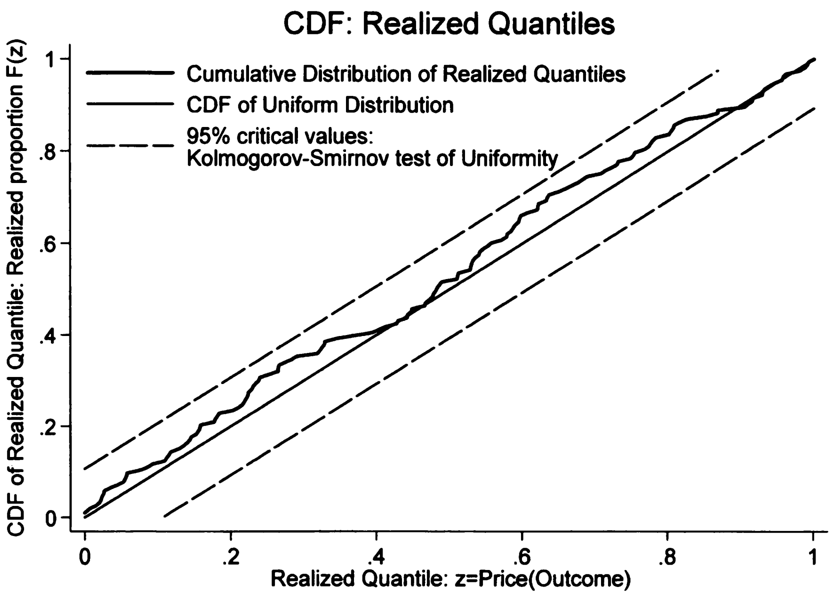

Figure 8
Cumulative distribution function of realized quantiles

Delving deeper, Figure 9 plots the same transformed variable for each data series separately.

Disaggregating the realized quantile by data series confirms that there is little evidence of non-uniformity of these distributions although there are some interesting hints of small miscalibrations in density forecasts. In particular, the ISM CDF is too steep in the central section, suggesting that too few realizations fall in the tails of the forecast distribution. The non-farm payrolls probability integral transform series is also very close to the upper critical value, suggesting too many realizations in the left tail. Neither of these leads to a rejection of the uniform distribution null hypothesis, however.

Figures 8 and 9 show that the economic derivatives based density forecasts have correct coverage. Efficient density forecasts also require independence of the probability integral transform variables over time. We therefore now turn to examining the time series of the probability integral transforms in Figure 10.

The time series plots do not suggest any clear time series correlation. To be sure, we have run simple AR(3) models, and found no statistically significant evidence of autocorrelation.

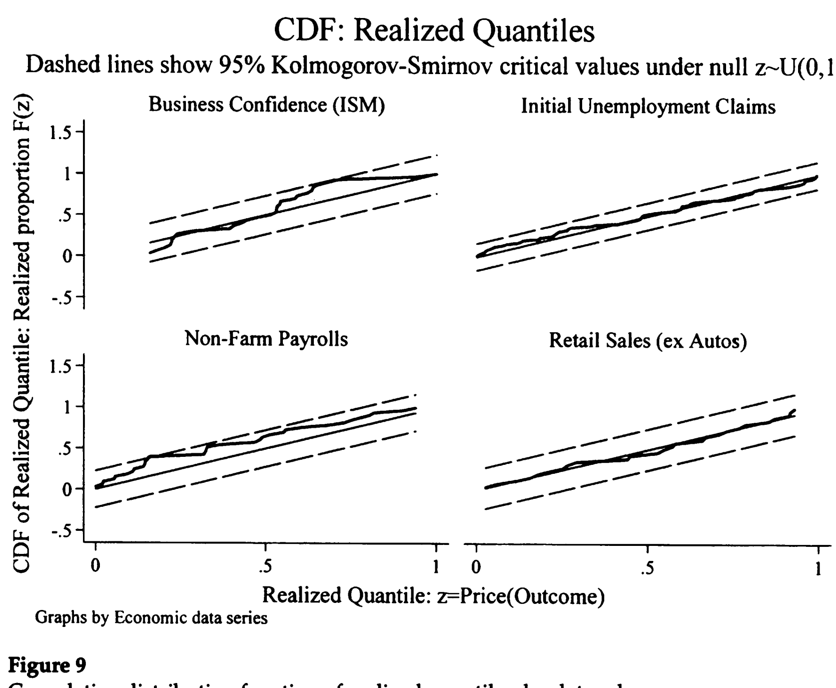

Figure 9
Cumulative distribution function of realized quantiles, by data release

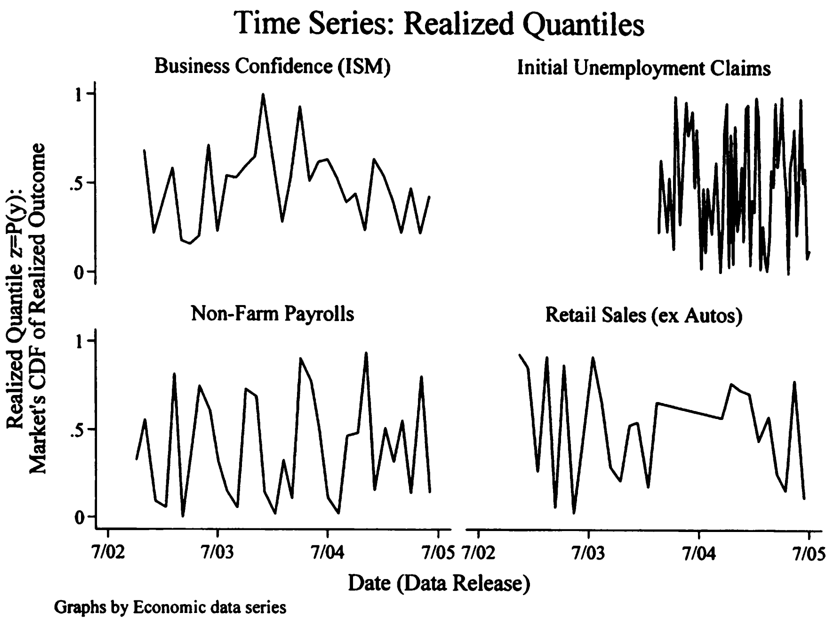

Figure 10
Time series of probability integral transforms

Finally we turn to a test that allows us to test jointly for both serial independence and uniformity of the realized quantile, maximizing our statistical power. Berkowitz (2001) notes that there exist more powerful tests for deviations from normality than from uniformity, particularly in small samples. He suggests analyzing a normally-distributed transformation of the probability integral transform. Specifically, he advocates analyzing:

$$n_t = \Phi^{-1}(z_t) = \Phi^{-1}\left(\int_{-\infty}^{y_t} \pi(x)dx\right)$$

where $\Phi^{-1}(z_t)$ is the inverse of the standard normal distribution function. Thus, if z_t is i.i.d.~U(0,1), then this implies that n_t is i.i.d.~N(0,1). We can thus test this null against a first-order autoregressive alternative allowing the mean and variance to differ from (0,1) by estimating:

$$n_t - \mu = \rho(n_{t-1} - \mu) + \varepsilon_t .$$

We estimate this regression by maximum likelihood. Berkowitz shows the exact log-likelihood function for the univariate case; it is simple to adapt this to the case of an unbalanced panel as in the present case:

$$L=\sum_{t=1}^{Unobs.\ Lag}\left[-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right)-\frac{\left(n_{s,t}-\mu/(1-\rho)\right)^2}{2\sigma^2/(1-\rho^2)}\right]$$

$$+\sum_{t=1}^{Observe\ Lag}\left[-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log(\sigma^2)-\frac{\left(n_{s,t}-\mu-\rho n_{s,t-1}\right)^2}{2\sigma^2}\right]$$

where the first term aggregates over observations where the lagged dependent variable is not observed, and the second term aggregates over all others.

Table 8 reports our estimation results. Estimating 3 parameters across each of 4 data series we find only two coefficients that are individually statistically distinguishable from the efficiency null. For each series we perform a likelihood ratio test that jointly tests whether the estimated models significantly deviate from the efficiency null. For none of our series is there significant evidence that the realized quantiles violate the i.i.d. uniform requirement. Finally, in order to maximize our statistical

Table 8
Testing for autocorrelation in the probability integral transform

	Non-farm payrolls	Business confidence	Retail trade (ex autos)	Initial claims	Pooled data
	$n_t-\mu=\rho(n_{t-1}-\mu)+\varepsilon$, where $n_t=\Phi^{-1}\left(\int_{-\infty}^{Outcome_t}\pi(x)dx\right)$				
Mean (μ)	−0.46**	0.03	0.04	−0.04	−0.10
	(.19)	(.15)	(.17)	(.15)	(.09)
Variance (σ^2)	1.05	0.70	0.76	1.46*	1.16
	(.26)	(.18)	(.21)	(.26)	(.13)
Autocorrelation (ρ)	−0.11	0.23	−0.31	0.05	0.001
	(.17)	(.26)	(.19)	(.13)	(.09)
$LL(\hat{\mu},\hat{\sigma}^2,\hat{\rho})$	−18.20	−12.59	−11.45	−51.45	−100.42
$LL(0,1,0)$	−21.34	−13.65	−12.82	−54.19	−101.99
LR test	6.27	2.12	2.73	5.48	3.16
	(p=0.10)	(p=0.55)	(p=0.44)	(p=0.14)	(p=0.37)
Sample size	33	30	26	64	153

Notes: (Standard errors in parentheses)
***, **, and * denote statistically significant deviations from the null at 1 percent, 5 percent, and 10 percent, respectively.
Forecast errors normalized by historical standard error of survey-based forecasts.

power we pool the estimates across all four indicators, and once again the test suggests that these density forecasts are efficient.

The evidence presented in this section shows that economic derivatives option prices are accurate and efficient predictors of the densities of underlying events. This finding is surprising in the sense that asset prices usually embed a risk premium due to risk aversion and for this reason tend to be systematically biased—a bias that does not seem to be present in this market. The implications of risk and risk aversion in the pricing of economic derivatives are the subjects of the next section.

6. The Role of Risk

Thus far we have interpreted the prices of digital options as density forecasts—an approach that would be warranted if investors were risk-neutral. Yet options and option markets exist precisely because there is risk, and it seems plausible that agents willingly pay a risk premium for the hedge offered by macroeconomic derivatives. We now turn to assessing the magnitude of this risk premium. To preview, we find that for an investor who holds the S&P 500 portfolio the aggregate risks that are hedged in these markets are sufficiently small that for standard assumptions about risk aversion the premium should be close to zero. Further, we show that option prices are typically quite close to the empirical distribution of outcomes. We then explore the corollary of these results, investigating what the pricing of these options implies about risk aversion.

Using option prices to make inference about risk and risk aversion is not a new idea, but is seldom attempted in the literature due to the complications arising from properties of standard options—complications that are not present in the economic derivatives market. In important papers, Jackwerth (2000) and Aït-Sahalia and Lo (2000) analyzed options on the S&P 500 to derive measures of risk aversion. Using economic derivatives to measure perceived risk and risk attitudes is far easier for several reasons. First of all, the options in these markets provide direct readings of state-prices; these do not have to be constructed from portfolios of vanilla options. More importantly, since the options expire within the same day of the auction, time discounting is not an issue and the discount factor can be set to zero. Similarly none of the concerns arising from the presence of dividends are present here.

To illustrate the relationship between risk aversion and the pricing of economic derivatives, we start by considering a representative investor who is subject to some risk that with probability p will change her wealth to β percent of its current value, w. The investor can buy or sell economic derivatives to protect herself against this shock. We consider the purchase of a derivative that pays \$1 per option purchased if the event occurs. Thus, the investor chooses how many derivatives to purchase (x) at a price π to maximize her expected utility:

$$\underset{\{x\}}{Max}\ EU(w) = pU\left(\beta w + (1-\pi)x\right) + (1-p)U(w - \pi x).$$

The first-order condition yields an optimal quantity of options, x^*:

$$\frac{U'(\beta w + (1-\pi)x^*)}{U'(w - \pi x^*)} = \frac{\pi(1-p)}{p(1-\pi)}.$$

That is, the investor purchases options until the marginal rate of substituting an additional dollar between each state is equated with the ratio of the marginal cost of transferring a dollar between states.

Because these economic derivatives are in zero net supply, in a representative agent model equilibrium requires that $x^* = 0$, yielding the equilibrium price:

$$\pi = \frac{p}{p + (1-p)\dfrac{U'(w)}{U'(\beta w)}}.$$

This expression yields some very simple intuitions. If β is unity then the probability and the state price are the same regardless of the degree of risk aversion. Indeed, such an option would be redundant because there is no risk to be hedged. Alternatively if agents are risk-neutral ($U'(w) = U'(\beta w)$), then again the option price represents the probability that the event will occur. If investors are risk averse and the option pays off following a negative shock to wealth ($\beta < 1$) then the state price is higher than the true probability. If the option pays off following a positive wealth shock ($\beta > 1$) then the risk-averse investors will price it at a value lower than its probability. Alternatively phrased, risk aversion leads the state-price distribution to shift to the left of the probability distribution, and this shift is larger the smaller the ratio $U'(w)/U'(\beta w)$; that is, distribution shifts further left for more risk-averse investors, and for larger adverse shocks.

Extending this logic to the case where the investor is subject to many possible shocks, and where there are markets available for her to hedge each risk is somewhat cumbersome, but yields only a minor modification. Specifically, the investor may face a variety of shocks where each specific shock, indexed by i, changes wealth to β_i percent of baseline and occurs with probability p_i. Investors hedge these risks so as to maximize expected utility by purchasing x_i options at price π_i, and each such option pays \$1 if the specified shock occurs. We refer to π_i as a state-price, and the distribution as the state-price distribution. The representative consumer's problem is:

$$\underset{\{x_i\}}{Max}\ E[U(w)]=\sum_i p_i U\left(\beta_i w + x_i - \sum_j \pi_j x_j\right).$$

We combine the first-order condition with the pari-mutuel mechanism constraint that total premiums paid should cover total payoffs in all states of the world ($\forall i: x_i = \Sigma_j \pi_j x_j$), to derive the following fairly intuitive expression for the risk premium:

$$\frac{\pi_i}{p_i}=\frac{U'(\beta_i w)}{\sum_j p_j U'(\beta_j w)}.$$

In Figure 11 we use this equilibrium relationship to assess the relationship between state prices and probabilities at different levels of risk aversion. Specifically, to make this exercise relevant to assessing the pricing of economic derivatives, we solve for the entire state-price distribution when the investor risks being hit by wealth shocks that are drawn from a normal distribution. In this example a one-standard deviation negative shock causes wealth to decline by 1 percent (That is, $\beta = 1 + 0.01z$ where $z \sim N(0,1)$). We calculate option prices for the log-utility case ($\gamma = 1$), a substantially more risk averse case ($\gamma = 5$) at the upper end of values usually assumed to be plausible by macroeconomists, and for a level of risk aversion typically thought implausible, but required to generate the observed equity premium ($\gamma = 20$).

As can be seen fairly clearly, for standard levels of risk aversion, the price distribution closely resembles the risk-neutral distribution. Increasing risk-aversion shifts the distribution to the left and the higher the risk aversion the more the state-price and data generating distributions are different.

More generally, our option pricing formula allows us to utilize data on two objects of the utility function, the distribution of shocks and the state prices, to make inferences about the third, the risk premium. In order to assess the likely magnitude of the risk premium, we begin by

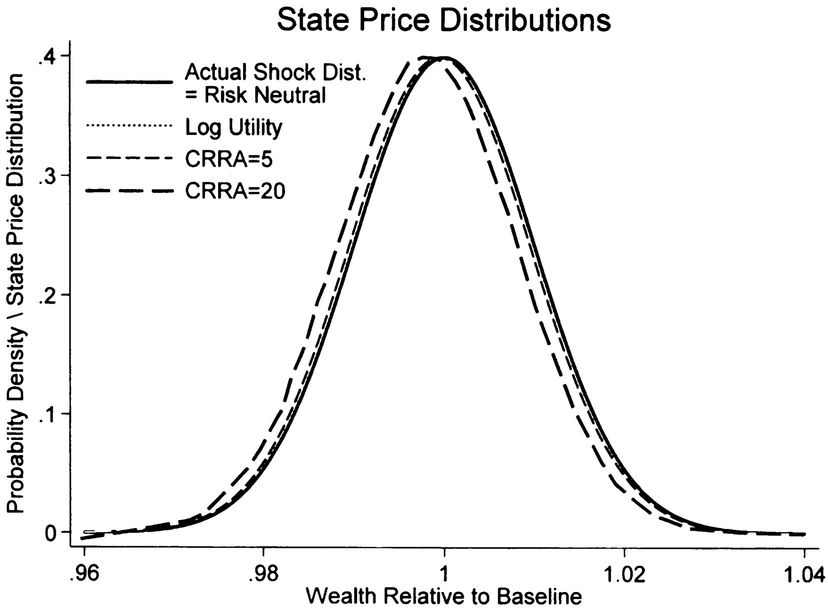

Figure 11
Risk aversion and state-price distributions

analyzing the divergence between the state-price distribution and the shock distribution that would be implied by specific utility functions and the economic shocks we see in our data. This requires us first to map the relationship between economic shocks and changes in wealth, then to map the empirical distribution of such economic shocks, before plugging these data into the above equation to back out the risk premium suggested by the theory.

Our analysis in section 3 (and specifically Figure 2) shows that the economic statistics have important effects on equity and bond markets. Backing out the implications of these shocks for wealth requires us to be more precise about a specific model of the economy. We assume complete markets, which imply the existence of a representative investor (Constantinides 1982). Following Jackwerth (2000) and Ait-Sahalia and Lo (2000) we assume that movements in the S&P 500 are representative of shocks to the entire stock of wealth. While one might be concerned that news about the economy affects different sectors differently, these are diversifiable risks, and so with complete markets should not affect wealth. Thus to recover the shock to wealth that macroeconomic derivatives allow one to hedge, we analyze the stock-market response to economic shocks in Table 9. That is, we run:

$$\Delta S\&P\ 500_t = \alpha + \beta^* \left(Actual_t - Forecast_t^{\text{Economic Derivs}} \right) .$$

As before, we examine changes in the 30-minute window around the announcement, and we scale the forecast error by the historical standard deviation of forecast errors for that series.

As expected, we find that positive shocks to non-farm payrolls, business confidence and retail trade are positive shocks to wealth, while higher initial claims is a negative shock. Comparing columns, it is clear that the non-farm payrolls surprise is easily the most important shock. The coefficient is also directly interpretable: a one standard deviation shock to non-farm payrolls raises wealth by 0.37 percent and the 95 percent confidence interval extends from +0.17 percent to +0.54 percent. These magnitudes are all much smaller than those used to construct Figure 11, suggesting that the relationship between prices and probabilities is even closer than that figure suggested. More to the point, these coefficient estimates correspond to $\beta - 1$ in the simple model presented above, allowing us to calculate the risk premium directly.

Rather than make specific parametric assumptions, we simply observe the distribution of different sized economic shocks in our data, and use a kernel density smoother to recover the shock distribution, using the estimates in Table 9 to rescale forecast errors into the corresponding wealth shocks. In this framework the frequency of specific shocks, their effects on wealth, and assumptions about risk aversion are sufficient to yield an estimate of the expected risk premium embedded

Table 9
Effects of economic news on the S&P 500

Dependent variable: %ΔS&P 500	Non-farm payrolls	ISM	Retail sales (ex autos)	Initial claims
$Actual_t - Forecast_t^{\text{Economic Derivs}}$ (Normalized by historical SD)	+0.37%*** (.10)	+0.11% (.11)	+0.04% (.06)	–0.01% (.02)
Adjusted R^2	0.31	0.005	–0.03	–0.006
n	33	30	26	64

Notes: Forecast errors normalized by historical standard error of survey-based forecasts. (Standard errors in parentheses)
***, **, and * denote statistically significant at 1 percent, 5 percent, and 10 percent, respectively.

in any particular strike price. Consequently in Figure 12 we show the state price distribution that the theory implies, based on the empirical shock distribution and assumptions about risk aversion. The risk-premium is simply the difference between the state price distribution, and the risk-neutral or empirical shock distribution.

Clearly for most plausible utility functions the risk premium is extremely small. Indeed, for log utility the risk premium is less than 1 percent of the price even for very extreme outcomes. Even with rates of constant relative risk aversion as high as five, the risk premium is still essentially ignorable; the only real exception to this is the non-farm payrolls release, which constitutes a much larger shock to wealth. In that instance, the price of an option with a strike price two standard deviations from the mean may be inflated by around 4 percent (and hence a call option would be priced at \$0.026 instead of \$0.025). If the relevant relative risk aversion parameter is as high as 20, then the data suggest that option prices might be somewhat more biased.

Of course, for many applications, the mean forecast implicit in the state price distribution is the object of interest. Thus in Table 10 we compute

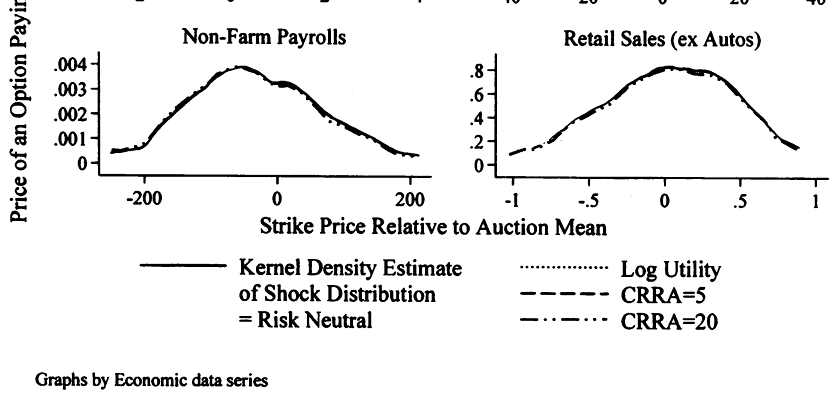

Figure 12
Effects of risk on the state price distribution

Table 10
Measures of central tendency of the probability and state-price distribution

	Non-farm payrolls	ISM	Retail sales (ex autos)	Initial claims
Panel A: Risk Premium: Mean of probability distribution less mean of state-price distribution				
Risk-neutral (γ=0)	0	0	0	0
Log utility (γ=1)	–0.32	–0.001	–0.0002	0.002
Risk-averse (γ=5)	–1.60	–0.005	–0.0009	0.008
Extremely risk averse (γ=20)	–6.40	–0.021	–0.0034	0.033
Panel B: Risk Premium Measured relative to historical standard deviation of forecast error				
Risk-neutral (γ=0)	0	0	0	0
Log utility (γ=1)	–0.0028	–0.0005	–0.0005	0.0001
Risk-averse (γ=5)	–0.0137	–0.0028	–0.0023	0.0004
Extremely risk averse (γ=20)	–0.0553	–0.0107	–0.0094	0.0018

Notes: In panel A, the units are thousands of non-farm payroll jobs, points on the ISM index, percentage growth in retail sales, and thousands of initial claims. Panel B measurements are relative to a one standard deviation shock.

the difference between the mean of the state price distribution and the mean of the underlying probability distribution for different values of assumed risk aversion. Again these numbers are based on the empirical distribution of shocks, although assuming normally distributed shocks yields similar magnitudes. Our aim is simply to provide a rule-of-thumb adjustment for calculating the mean of the probability distribution from the widely reported mean of the auction price distribution.

Panel A shows that, under risk aversion, the mean of the state price distribution will under-estimate the mean of the risk-neutral ("true") distribution for the three pro-cyclical series, but will lead to a minor overstatement of initial claims, which is countercyclical. The adjustments in Panel A are in the same underlying units as the statistics are reported in, and hence suggests, for instance, that if the relative risk aversion of investors is five, then the mean of the state price distribution understates the mean forecast by about 1600 jobs. Panel B presents these same results in a metric that better shows that these magnitudes are small, scaling the risk-premium adjustment by the standard error of the forecast. In each case the bias from simply assuming risk-neutrality is less than one-tenth of a standard error, and in most cases, it is orders of magnitude smaller.

While Table 10 suggests that risk *should* lead the market-based forecast to be only slightly lower than the risk-neutral forecast, we can take advantage of the time series movement in uncertainty to test this.[16] In Figure 13 we show forecast errors and uncertainty for each data series. In no case is the regression line statistically significant, suggesting that the data do not falsify the implications in Table 10 that the slope should be approximately zero. Notice that this exercise is slightly different from the one in Table 10 as here we look at the consequences of time-variance in the amount of risk, while in Table 10 the amount of risk is implicitly taken as invariant but the price of risk changes.

In sum, Figure 12 and Table 10 imply that under standard assumptions about risk, the state price distribution is a reasonable approximation to the true underlying probability distribution, and this conclusion holds even when we make fairly extreme assumptions about risk aversion. Indeed, Figure 13 and our analysis of the probability integral transform in the previous section confirmed precisely this point and in most cases market prices provided quite successful estimates of empirical realizations.

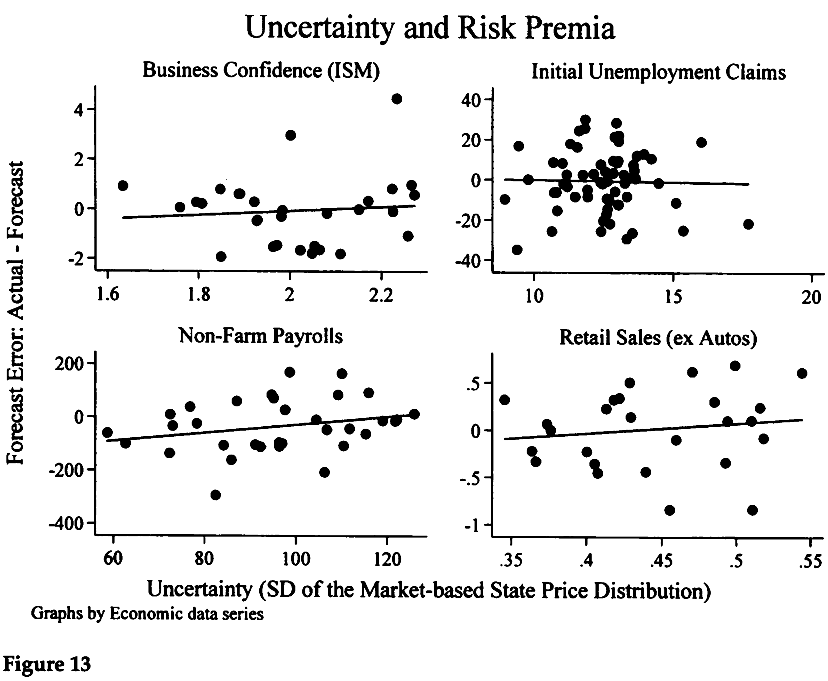

Figure 13
Uncertainty and risk premia

Figure 14 makes this point in an alternative manner, pooling the data across all auctions within each data series to map both the empirical shock distribution and the average state price distribution. The two appear remarkably close given the limited number of observations identifying the distribution of outcomes.

Our option-pricing formula also suggests that we can compare option prices and observed outcomes to back out an estimate of risk aversion. Indeed, under the assumption of constant relative risk aversion of γ, our option pricing formula directly yields a log-likelihood function:

$$L=\sum_{a}^{Auctions}\left[log(\pi_a^*)+\gamma\, log(\beta_a^*)-log\left(\sum_{s}^{Strikes_a}\pi_{s,a}\beta_{s,a}^{-\gamma}\right)\right]$$

where auctions are indexed by a and digital options within each auction are further indexed by s, the asterisk indexes the winning digital option, and thus π^* and $\pi_{s,a}$ come from the data, while estimates of the wealth impacts of shocks, β_i are taken from Table 8, and β^* is the relative wealth position given the observed shock.

We pooled all of our data to estimate the coefficient of relative risk aversion (γ), but these data do not yield much power: the 95 percent confidence interval around our estimate of γ extends from –182 to +27, with a central estimate that suggests risk-loving behavior. This is readily apparent in Figure 13, which shows that the state price distribution is to the right of the outcome distribution for non-farm payrolls, and to the left of the outcome distribution for the counter-cyclical initial claims data. (As Figure 12 shows, risk aversion would suggest the opposite pattern.)

However, rather than highlight our point estimate, we regard its enormous imprecision as arguably more interesting.[17] This imprecision derives from the fact that under our complete market assumptions the economic risks that can be hedged in this market are sufficiently small that alternative views about risk aversion do not affect all that much how one would price options tied to these risks. From an estimation standpoint this implies small amounts of noise in the option prices potentially yield very different implications for point estimates of implied risk aversion. Again, Figure 12 is instructive: essentially our estimates suggest that the data cannot distinguish between any of the state price distributions drawn on that figure, and given how close they are, this is not particularly surprising.

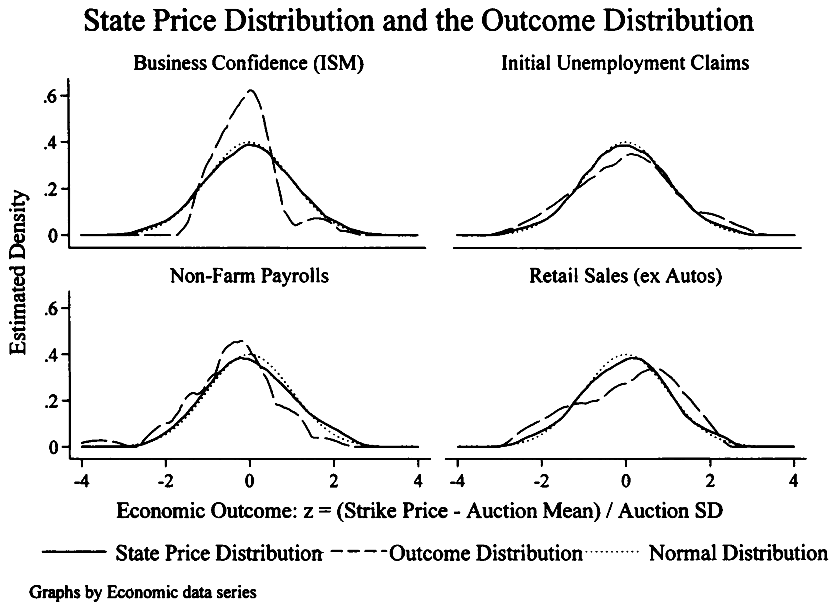

Figure 14
State price distribution and the distribution of outcomes

Thus while this market does not yield particularly useful estimates of risk aversion, the flipside is that this is driven by the fact that option prices are relatively insensitive to assumptions about risk aversion. From a practical perspective this is good news: the option prices that we observe in this market are a reasonable approximation to the risk-neutral distribution.

7. Conclusions

In this paper we provided a first analysis of the option prices from the new economic derivatives market. Economic derivatives (which have an interesting, pari-mutuel, market clearing mechanism) are novel because these binary options are written on economic data releases and state-prices of different strikes provide information not only about markets' central belief but also about implied probabilities of outcomes away from the mean. This information is not available from surveys.

We dwelled on several aspects of the economic derivatives, starting with their predictive performance. These options appear to yield efficient *density* forecasts, a rarity. Knowing that event probabilities are correctly priced in this market makes inference using the dispersion statistics convincing. In particular, this justifies using the option-based standard deviation to measure uncertainty about a data release. Comparing uncertainty with disagreement, the standard deviation of survey responses, showed that these two measures of dispersion do not have a high degree of correlation. It may not be advisable to use disagreement as a proxy for uncertainty.

The density forecast efficiency tests, when applied to market-based measures, are joint tests of efficient pricing and absence of risk premia. Our finding that economic derivatives based densities are efficient therefore indicate that risk premia in this market are unlikely to be sizable. We exploited the institutional structure of economic derivatives to study risk and risk aversion. This is quite straightforward when options from this market are used, compared to using S&P 500 options, which require taking into account time discounting and dividends. We believe economic derivatives are promising instruments for economists who would like to use asset prices to learn about agents' beliefs and preferences.

We should emphasize that we view this paper as an initial exploration. We showed that economic derivatives correctly capture subjective beliefs and provided some applications of this information. Having these subjective probabilities will facilitate future research to study how expectations are formed and how they relate to actions, as well as to analyze agents' responses to occurrence of events of different prior subjective probabilities.

Acknowledgements

The authors would like to thank Jeffrey Crilley and Banu Demir for excellent research assistance. Thanks also to Chris Carroll, Frank Diebold, John Fernald, Jeffrey Frankel, Jose Lopez, Glenn Rudebusch, Betsey Stevenson, Adam Szeidl, Jonathan Wright, and Eric Zitzewitz for useful discussions and to Bill Cassano and Kevin Keating of Goldman Sachs for helping us with institutional detail. This paper was partly written while Gürkaynak was at the Federal Reserve Board. He thanks that institution for the outstanding research environment it provided. Wolfers gratefully acknowledges the support of a Hirtle, Callaghan and Co. – Arthur D. Miltenberger Research Fellowship, Microsoft Research, the Zell/Lurie Real Estate Center, and the Mack Center for Technological Innovation.

Notes

1. Beyond these markets, the Chicago Board of Trade is offering federal funds rate futures and options and the Chicago Mercantile Exchange has a thinly traded CPI futures contract. Online markets such as Hedgestreet and Tradesports also offer an array of economic derivatives to retail investors.

2. Currently every order must go through Goldman Sachs, Deutsche Bank, or ICAP (an interdealer broker). As of the writing of this paper (September 2005) an agreement was in place to involve the CME in the auction process.

3. The transaction cost—the fee paid to Goldman Sachs and Deutsche Bank—is 1 percent of the notional amount (one cent per digital option) capped at 10 percent of the price of the option.

4. The auction clearing pari-mutuel algorithm, called "Parimutuel Derivative Call Auction technology" is patented by Longitude Inc., who also license their product to create markets in mortgage prepayment speeds and natural gas and crude oil inventories (see Baron and Lange, 2003, for more on this algorithm).

5. Auctions of initial claims options are not held for the releases that immediately precede the employment report. Our data set consists of 33 non-farm payrolls auctions, 30 business confidence auctions, 26 retail trade auctions, and 64 initial claims auctions.

6. Some auctions on European inflation take place two months prior to the data release.

7. When more than one auction was held for a single data release, we analyze data from the latest auction.

8. MMS was acquired by Informa in 2003 and no longer exists; Action Economics is now providing the same survey service. We use the MMS numbers for most of our sample and the Action Economics survey for the more recent period. Bloomberg survey numbers were used to fill some gaps. Despite using more than one source, we call our survey numbers "the MMS survey" as most of our data is from this source. The MMS survey sample consists mainly of professional economists working in financial markets, and many of the firms surveyed are probably also participants in the economic derivatives market.

9. More specifically, throughout the paper we treat the distribution as discrete, assuming that all probability mass occurs at the midpoint of the relevant bin. For the tails we impute an upper- and lower-bound so that the midpoint would be equal to the mean of that bin if the pdf were normal. Our results are invariant to different treatments of tail probabilities.

10. In order to maintain a non-overlapping sample, we calculated the standard deviation of the survey-based forecast errors for samples ending in October 2002. The "historical" sample begins in January 1990 for non-farm payrolls and retail sales, in July 1991 for ISM, and in July 1997 for initial claims. The historical standard errors of these forecasts are 115,600 non-farm payroll jobs, 18,500 initial unemployment claims, 0.37 percent growth in retail sales and 1.99 points of the ISM index.

11. The intraday data we use help us isolate the market reaction to the data release in question with minimum noise. The yields we use are yields of on-the-run Treasury securities. The stock price changes are from S&P futures contracts as the stock market is not open at 8.30 a.m. (EST), when the three of the four macroeconomic data series we are interested in are released (ISM is a 10.00 a.m. release). In taking the market snapshots, if

there is no trade in a given security five minutes before the event, we search back in time until we find a trade or a settlement price. If there is no trade exactly 25 minutes after the event we again search back in time, until the data release moment. If there are no trades in this 25 minute interval we mark a zero change, assuming that if there was a surprise in the data release that changed the shadow price of a security there would have been a trade over this time period. We do not search for a trade forward in time so as to ensure that the price change we observe is not due to another event that took place later in the same day. The data set is described in detail in Gürkaynak, Sack, and Swanson (2005).

12. Note that while a strong data release for an important statistic should unambiguously push yields up, the effect on stock prices is not as clear. The news that the state of the business cycle is better than expected will lift the S&P index, but the associated increase in interest rates has a dampening effect on equities.

13. Panel D controls for the slope of the yield curve (measured as the difference between the ten year and 3-month yields), and the change in the S&P 500 over the preceding ten trading days as regressors.

14. While Table 6 provides useful descriptive detail, it is silent on the issue of driving forces. There are potentially three important influences that may be driving variation in uncertainty about a particular economic statistic: fundamental uncertainty about the true underlying state of the economy, data-driven uncertainty whereby other data series have not spoken clearly about the state of the economy, and uncertainty about the extent of possible measurement error in the underlying economic statistic. Financial market responses to economic news can potentially help sort out which driving forces are important as economic news has its largest impact on beliefs (and hence on financial markets) when there is greater uncertainty about the true state of the economy. By contrast, traders will be more likely to discount the same sized shock if their uncertainty reflects concerns about measurement. Our statistical tests for these produced very imprecise estimates that we do not report, but we note this potential use of economic derivates based information.

15. Note that the critical values are appropriate for each bin separately, but they are inappropriate for jointly testing that the heights of all bins are drawn from a binomial distribution.

16. We thank Jeffrey Frankel for suggesting this test to us.

17. Note that when estimating implied risk aversion in this fashion, we treated the β_i as known. The confidence interval would have been even wider had we accounted for the variance imparted by having the β's estimated.

References

Aït-Sahalia, Yacine, and Andrew W. Lo. 2000. "Nonparametric Risk Management and Implied Risk Aversion." *Journal of Econometrics* 94: 9–51.

Baron, Ken, and Jeffrey Lange. 2003. "From Horses to Hedging." *Risk* 16(2): 73–77.

Berkowitz, Jeremy. 2001. "Testing Density Forecasts, with Applications to Risk Management." *Journal of Business and Economic Statistics* 19(4): 465–474.

Christoffersen, Peter. 1998. "Evaluating Interval Forecasts." *International Economic Review* 39: 841–862.

Constantinides, George. 1982. "Intertemporal Asset Pricing with Heterogeneous Consumers and without Demand Aggregation." *Journal of Business* 55(2): 253–276.

Diebold, Frank X., Todd Gunther, and Anthony S. Tay. 1998. "Evaluating Density Forecasts, with Applications to Financial Risk Management." *International Economic Review* 39: 863–883.

Diebold, Frank X., Anthony S. Tay, and Kenneth F. Wallis. 1999. "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters." In R. Engle and H. White (eds.), *Festschrift in Honor of C.W.J. Cranger.* Oxford: Oxford University Press, 76–90.

Eisenberg, Edmund, and David Gale. 1959. "Consensus of Subjective Probabilities: The Pari-Mutuel Method." *Annals of Mathematical Statistics* 30: 165–168.

Fair, Ray C., and Robert J. Shiller. 1990. "Comparing Information in Forecasts from Econometric Models." *American Economic Review* 80(3): 375–389.

Gürkaynak, Refet S., Brian Sack, and Eric Swanson. 2005. "Did Financial Markets Think Disinflation Was Opportunistic?" Manuscript, Federal Reserve Board.

Jackwerth, Jens Carten. 2000. "Recovering Risk Aversion from Option Prices and Realized Returns." *The Review of Financial Studies* 13(2): 433–451.

Llambros, Louis, and Victor Zarnowitz. 1987. "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy* 95(3): 591–562.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2003. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual* 18: 209–248.

Shiller, Robert. 1993. *Macro Markets*. Oxford: Oxford University Press.

Wolfers, Justin, and Eric Zitzewitz. 2004. "Prediction Markets." *Journal of Economic Perspectives* 18(2): 107–126.

Wolfers, Justin, and Eric Zitzewitz. 2005. "Interpreting Prediction Market Prices as Probabilities." *Mimeo*, University of Pennsylvania.

Comment

Christopher D. Carroll, Johns Hopkins University and NBER

This paper opens up what promises to be a whole new approach to macroeconomic research. Market-based forecasts of macroeconomic variables provide a promising way to neatly sidestep the intractable, insoluble, and semi-theological debates about how expectations are formed that have plagued macroeconomics since Keynes first speculated that "animal spirits" were a driving force in business cycles.

So you might say I'm a fan.

In fact, the first part of my discussion will argue that the results of the paper are even more important than one might conclude from the authors' own analysis, because they focus on the (microscopic) differences between survey-based forecasts and market-based forecasts, rather than on the impressive similarities between them. The brief latter part of the discussion raises some reasons for caution about the institutional design and operation of these markets.

1. Comparing Survey and Auction Based Expectations

A substantial part of the paper (Tables 1–3) compares expectations as revealed by the auction market to the mean forecasts of a survey of professional forecasters. An incautious reader might get the impression that these results suggest the market-based expectations are notably better than those of the survey. In fact, I think the opposite interpretation is the right one: When used to measure the same thing, survey-based expectations are, for analytical purposes, indistinguishable from market-based expectations.

Consider, for example, the non-farm payrolls data, which are for most purposes the most important single U.S. data release.[1] The authors present the following comparative statistics about the two. (These are taken from their Table 1).

Table 1
Prediction errors from auction and survey (non-farm payrolls)

	Mean absolute error (AbsErr)	Root mean squared error (RMSE)	Correlation with actual outcome
Auction	0.723	0.907	0.700
Survey	0.743	0.929	0.677

The table speaks for itself.

The authors emphasize the results for their other data series, which could be described as providing a smidgen of evidence that the market forecasts are more accurate than the survey forecasts. I will shortly express some quibbles with this interpretation. But before doing so, I would like to point out that even under the authors' interpretation, the superiority of the auction forecast is generally small.

This is important because the macroeconomic derivatives markets have been operating only for a short time. Since, according to the NBER Business Cycle Dating Committee, the average postwar business cycle in the U.S. has had a duration of about eight years, the usefulness of these data for macroeconomic analysis will arguably be modest for at least a decade. If instead we draw the conclusion that the macroeconomic derivatives markets have definitively revealed the impressive qualities of survey-based expectations, the scope of the paper's usefulness is vastly expanded, since various kinds of survey-based expectations have been collected for a very long time (for example, the Survey of Professional Forecasters has been conducted since 1968).

1.1 *Quibbles*

As the authors note, the auctions they analyze do not provide any real opportunity for hedging macroeconomic risks in the sense Shiller (1993) originally proposed because they are generally conducted only a few hours (or at most a few days) before the data are released.

This timing, however, means that participants in the auctions have more recent information than survey participants, whose views are collected every Friday. In the case of a data series released on a Thursday, the auction participants' information set could incorporate nearly a week's worth of extra knowledge about the state of the economy.

This problem is particularly serious for initial claims for unemployment insurance, since this is a weekly series released on Thursday mornings. Indeed, it is remarkable that the almost week-old surveys do almost as well as the previous-day auctions in forecasting this weekly series.

An alternative way of analyzing the authors' data (and one that is fairer to the forecasters) would be to hypothesize that both forecasters' and auction participants' views are rational; in that case, Hall (1978) taught us that the auction results should equal the survey results plus a random expectational error that reflects the forecasters' extra information:

$$A_t = S_{t-1} + \varepsilon_t, \tag{1}$$

which can be tested by estimating a regression

$$A_t = z_0 + z_1 S_{t-1} \tag{2}$$

and testing $z_0 = 0$ and $z_1 = 1$.

To test this proposition as an overall characterization of the authors' data, it is necessary to put the various statistics on a common footing in the sense of having comparable means and measures of variability. I did so by subtracting, for each series, the mean realized value over the sample period, and dividing by the gap between the maximum and minimum realized sample values.[2]

Results are plotted in Figure 1. As the figure illustrates, there is a very strong association between the survey and the auction predictions.

The point is illustrated statistically by Table 2, which reports the results of a regression like the one contemplated in equation (2). The hypotheses that $z_0 = 0$ and $z_1 = 1$ cannot be rejected at standard significance levels, and the $\bar{R}^2$ for the regression is over 90 percent. When the sample is restricted to the crucial non-farm payrolls data, similar results obtain.

One way of testing whether the more up-to-date information held by auction market participants could plausibly explain a modest superiority in their forecasts is to see whether auctions that are held closer to the date of the data release produce forecasts that are more accurate. Unfortunately, the authors' dataset contains only a few auctions that were held earlier than the day on which a data series was released. Most of these were for the ISM data. Table 3 calculates the size of the absolute error for the 21 auctions that were held on the morning of the data release, the four auctions that were held one day before, and the three auctions that were held three days before. (There seem to be no examples of auctions conducted two days before the release). The mean absolute error is notably larger for the auctions conducted rela-

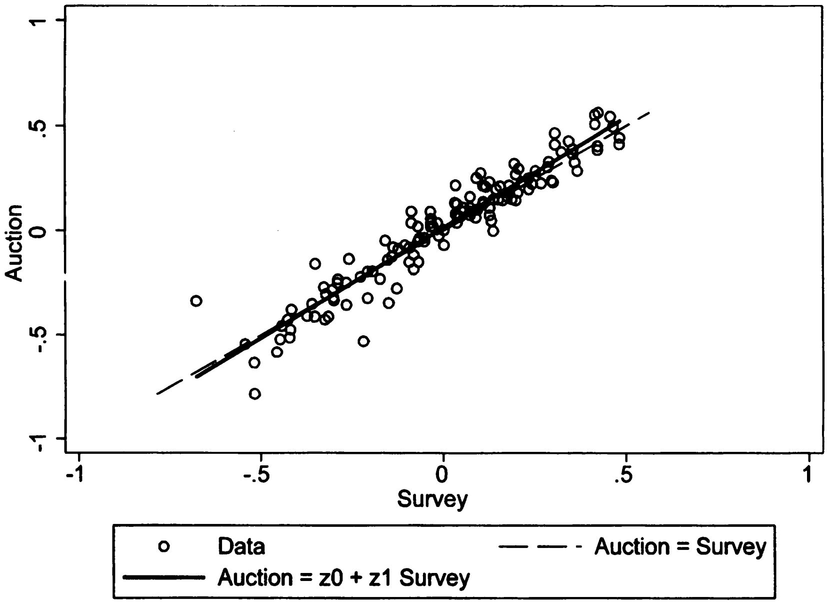

Figure 1
Survey expectations versus auction expectations

Table 2
Regression of auction on survey expectations

	Auction = z0 + z1 Survey		
Data series	z0	z1	$\bar{R}^2$
All	0.013 (0.007)	1.055 (0.039)	0.91
Payrolls	0.001 (0.014)	1.096 (0.052)	0.95

Robust standard errors in parentheses.

Table 3
Absolute error for different ISM auction horizons

Days between auction and data release	Number of auctions	Mean absolute error
0	21	0.48
1	4	0.57
3	3	0.56

tively earlier, as would be true if significant news generally arrives in the period leading up to the release (though separate tests (not shown) indicate that these differences are not statistically significant).

The authors emphasize the results of a final horse race (in Table 2) between the two series. They show (convincingly) that financial market reactions to the actual data release are stronger when the "surprise" is measured as the deviation from the auction forecast than when it is measured as the deviation from the survey forecast, at least for the payrolls data.

Again a possible explanation is the later date of the auction than the survey. Another possibility that the authors suggest is that the participants in the auctions are precisely the same people whose financial transactions, post-release, will determine the market reaction. If this is true, it would be puzzling if their opinions did *not* have more influence on financial market outcomes than the opinions of bystanders like the economists participating in the surveys.

None of this is meant to dispute the proposition that the auction based forecasts are a superior source of information, when both auction and survey data exist. As the authors show, the auction data paint a much richer picture of expectations than is available from the surveys, particularly with respect to the probability distribution over possible outcomes, which can be condensed (as the authors show) in any of several ways to measure uncertainty. In 30 years there may be no reason to use survey data at all because a sufficient amount of auction data will be available. But for the time being, the authors' results provide compelling evidence that surveys capture an enormous amount of useful information.

This richness is used in section 4 of the paper to examine a question that heretofore has been a matter of speculation: whether disagreement among survey participants can be interpreted as a measure of uncertainty.

On the whole their conclusion is that such an interpretation is problematic. Table 4 reproduces the key results from their analysis of this question, in which they regress measures of uncertainty on measures of disagreement. The absolute magnitudes of the coefficients are not meaningful, because there is no obvious mapping between the cross-forecaster standard deviation of forecasts of the mean value of the release, and the standard deviation of the released data itself. The right questions are the degree of statistical significance of the relationship

Table 4
Uncertainty versus disagreement

Uncertainty = $\alpha + \beta$ Disagreement		
Series	β	$\bar{R}^2$
Payrolls	0.66** (0.29)	0.11
Retail sales	0.44** (0.16)	0.20
Initial claims	0.27*** (0.07)	0.17
ISM	−0.03 (0.12)	−0.03

between uncertainty and disagreement, and the total proportion of uncertainty that can be measured by disagreement. Except for the ISM series, the authors find a highly statistically significant relationship between disagreement and uncertainty.

They tend to emphasize, however, the finding that the $\bar{R}^2$ is well below one in all cases. But there is clearly sampling error in the survey of forecasters; how to think about this is not entirely obvious, since there are forecasters who exist but are not in the survey and the survey participants vary over time. By itself this would be enough to prevent an $\bar{R}^2$ equal to one even if the authors' measures of uncertainty were perfect.

My own sense is that the more important question is whether disagreement can be interpreted as a statistically reliable indicator of the degree of uncertainty, rather than a direct measure. One way to make the question concrete is to ask whether the regression the authors report can be thought of as the first stage of a two-stage least squares regression of uncertainty on disagreement. One could then use the prediction of the estimated equation as a contemporaneous measure of appropriately calibrated uncertainty. Judged in this way, the $\bar{R}^2$'s for the first stage regressions and the high statistical significance of the coefficients are plenty good enough to interpret the prediction of the model as an (instrumented) measure of uncertainty. (Of course, careful econometrics would have to make sure that this cross-section disagreement is not perfectly correlated with some other macro variable (like the inflation rate).)

2. Caveats about Macro Markets

Despite their many attractive properties, it is worth worrying a little bit (at this early stage) about the longer term consequences of the creation of macro markets, especially for the data collection process.

I have the fullest faith in the integrity and objectivity of the staff at the agencies that produce economic data. But there can be no doubt that the creation of macro markets will increase both the pressure on the staff and the ease with which an unscrupulous employee could exploit inside information. Data security procedures need not only to be objectively rigorous but also to be transparently seen to be rigorous. Possibly there should be a systematic ongoing program (by the Securities and Exchange Commission?) to monitor trading in macro markets for any signs of insider trading.

Another concern is that if macro markets become sufficiently popular (and lucrative), the economic agencies may have a problem of retaining senior staff. If senior officials were regularly lured away from their posts by the offer of salaries many times higher than the government can provide, it might be difficult to preserve the institutional memory and expertise necessary for guaranteeing the consistency and high quality of U.S. statistics. Probably the only appropriate measure that could be taken to prevent this (in addition to paying appropriately high salaries to the senior staff) would be to impose strict ethics rules that require a substantial waiting period (say, five years) between the time of departure from a statistical agency and any employment that exploits that expertise in the context of macro markets.

Finally, and perhaps most significantly, the existence of macro markets could influence the data collection procedures themselves. Although the currently existing auction markets probably do not pose much risk in this dimension, when markets are created for longer-term forecasts (as they inevitably will be), the holders of those auction contracts will have the incentive to become lobbying groups for or against changes in the methods of data collection. Imagine, for example, that macro markets had existed at the time of the Boskin Commission on reform of the CPI in the mid-1990s, or the redefinition of the unemployment rate in the early 1990s. If each decision a commission announces results in immediate capital gains or losses of billions of dollars for holders of contingent securities, there will be extraordinary incentives to subvert the objectivity of the decision makers. Good institutional design could certainly circumvent these pressures, but if data collection procedures are perceived to be able to be influenced by the appointment of ad hoc committees nominated by politicians there is reason to worry.

This risk could perhaps be alleviated if the agencies that produce the data were to create standing committees of scientific advisors associated with each of the major statistical releases for which macro markets exist or are in contemplation. For example, a panel of distinguished labor economists might be recruited to monitor proposed changes to the non-farm payrolls survey. These committees might borrow the model of the NBER Business Cycle Dating Committee: Meetings only when warranted by some event, but a committee that is always well defined. This would provide some transparent insulation against the political forces that might otherwise mobilize to have commissions appointed whose members would be picked to reach preordained conclusions.

It is important to resolve these issues early, because the whole superstructure of macro markets will be undermined if the integrity of the data collection process comes into question. But if addressed early, these problems should not be serious.

3. Conclusions

All quibbles aside, this paper, and the macro markets that it is the first to explore, represent a tremendous innovation in macroeconomic analysis. I look forward with great anticipation to the literature that will undoubtedly flow from them.

Notes

1. Like the authors, Fleming and Remolona 1997 find that this data release moves the bond market more than any other, and more recently Faust et. al. 2003 have found that this data release moves exchange rates even more than monetary policy surprises.

2. Results were similar when the data were scaled, following the authors, by the presample standard error; the resulting figure is slightly more legible using my scaling method.

References

Faust, Jon, John H. Rogers, Eric Swanson, and Jonathan H. Wright. 2003. "Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data." *Journal of the European Economic Association* 1: 1031–1057.

Fleming, Michael, and Eli Remolona. 1997. "What Moves the Bond Market?" *Federal Reserve Bank of New York Economic Policy Review* 3: 31–50.

Hall, Robert E. 1978. "Stochastic Implications of the Life-Cycle/Permanent Income Hypothesis: Theory and Evidence." *Journal of Political Economy* 96: 971–987.

Shiller, Robert J. 1993. *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks.* Oxford: Oxford University Press.

Comment

Adam Szeidl, *University of California, Berkeley*

1. Introduction

This is an interesting and informative paper that explores pricing behavior in a new market for macroeconomic derivatives. Asset markets where risk associated with future macroeconomic events can be traded are a recent financial innovation. These markets may allow more efficient sharing of macro risks and increase economic welfare. To assess their potential, it is important to understand how well existing economic derivatives markets function. Analyzing data from one such market where claims on macroeconomic indicators including non-farm payrolls are traded, this paper argues that (1) Expectations derived from market prices are more accurate than survey-based forecasts and less subject to behavioral biases; (2) The market predicts the probability distribution of outcomes remarkably well; (3) Risk aversion plays at most a small role in determining prices in this market.

I begin by discussing potential theoretical foundations for the empirical findings. Then I briefly discuss features of the market mechanism, and finally turn to the role of risk aversion. My comments suggest additional empirical tests that can sharpen our understanding of how markets for economic derivatives function.

2. Theory

Perhaps surprisingly, it is not easy to come up with plausible microfoundations for findings (1) and (2). Why are prices accurate predictors of outcomes? And why are prices more accurate than survey-based forecasts, when in many economic models, prices are functions of the beliefs that forecasts measure? To answer these questions, I begin by exploring the mechanism through which markets may aggregate information. A large theoretical literature (e.g., Grossman 1976 or more recently Reny and Perry 2003) argues that markets correctly aggregate heterogeneous information in the presence of common prior beliefs. In practice, however, the common prior assumption appears to be at odds with often-observed disagreement in survey forecasts among professional forecasters, because different individuals with common priors cannot agree to disagree (Aumann 1976). A plausible alternative in this context is to assume that disagreement is due to heterogeneous prior beliefs.

However, with heterogeneous beliefs, as argued for example by Manski (2004), it is not a-priori clear that predictive markets should correctly aggregate information. To see the logic, note that in principle, a wealthy individual with incorrect beliefs may be able to push prices away from fundamental values by the sheer size of her investment. More formally, Wolfers and Zitzewitz (2005) show that with risk-averse investors and a competitive market, the price will equal the wealth-weighted average belief in the population. This result confirms that market prices can depart from true expectations if the distribution of beliefs is correlated with wealth. On the other hand, in this model, accurate market prices obtain if the average belief in the population correctly predicts outcomes. This suggests that the reason why predictive markets function so well is that the average belief of investors is correct.

To test this proposition, one can look for alternative empirical measures of beliefs. A natural candidate, used for example by Mankiw, Reis, and Wolfers (2003), is survey-based forecasts. If one accepts that such surveys are a good measure of beliefs, then the Wolfers-Zitzewitz model predicts that surveys will forecast outcomes at least as well as market prices. However, this prediction contradicts finding (1) of this paper. How can prices be more accurate than surveys, when surveys are a direct measure of investors' beliefs?

To resolve this contradiction, one has to relax one of the assumptions of the previous argument. It must be that either (a) prices are not more accurate than survey-based forecasts; or (b) surveys do not reflect true beliefs; or (c) prices are accurate not because they reflect average beliefs, but for some different reason. Distinguishing between these alternatives would be useful to better understand the workings of predictive markets.

Let us address each possibility in turn. Case (a) suggests that finding (1) in the paper is due to other differences between the survey and market data. Timing is one such difference: while the predictive market meets on the morning of the data release, the survey is collected up to a week earlier. Given such differences in timing, information that becomes available after the survey is collected may be reflected in the market price. This explanation suggests that surveys are good measures of expectations. From a practical perspective, this would be useful, because survey data is more widely available than data from predictive markets. Using the data of the current paper, this explanation can be tested by comparing the differential accuracy between surveys and forecasts depending on the difference in timing. When this explanation is correct, surveys that take place later should be closer in accuracy to market prices.

Case (b) may hold for example if survey respondents have little to lose from making incorrect predictions, while market participants have money at stake. In this case, earlier work where beliefs are measured using survey based forecasts is potentially misleading. While there is little doubt that predictions do improve when the stakes are higher, the question is quantitative. How much does precision increase when the stakes go up? A preliminary empirical approach to explore this question is to compare the accuracy of predictions across markets with different stakes, as measured perhaps by total investment in short and long positions. In markets with higher total investment, we should find that prices are better predictors of outcomes.

In my view, case (c) is the least likely. If prices do not reflect average beliefs, then we are back to the original puzzle: Why do prices in predictive markets forecast outcomes so accurately?

To summarize, the most plausible theory raises the question of whether finding (1) is caused by the different nature of surveys versus markets or their differential timing, and suggests additional empirical tests to help sort out whether markets are just as accurate as surveys or more accurate because the stakes are small for survey participants.

3. The Pari-Mutuel Mechanism

Understanding the logic of information aggregation in predictive markets is further complicated by the fact that the market mechanism is not competitive. The market is a modified version of the pari-mutuel mechanism often used in horse race betting. Eisenberg and Gale (1959) explore Nash equilibrium in a simple version of the basic pari-mutuel model. They establish existence and uniqueness of equilibrium; however, the equilibrium they find need not involve prices that correctly predict outcomes. To quote the last sentence in their paper: "In the case of two bettors with equal budgets if the first bettor's subjective probability distribution on two horses is ((1/2),(1/2)) then the equilibrium probabilities will be ((1/2),(1/2)) regardless of the subjective probabilities of the second bettor, as the reader will easily verify." Therefore, in the special case discussed in the quote, the price will be independent of the beliefs of the second bettor. This example suggests that exploring the actual market mechanism in more detail can lead to useful insights about the logic of information aggregation.

4. Risk Aversion

My final topic is the role of risk aversion. Using a simple model with power utility investors, the paper shows that for reasonable coefficients of relative risk aversion the risk premium of holding economic derivatives should be very small. Based on this argument, the authors conclude that risk is unlikely to affect asset prices in predictive markets.

One problem with this logic is that the same calibration argument, if applied to the aggregate stock market, would imply that risk plays at most a minor role in determining expected stock returns, and that the equity risk premium should be very small. As it is well known, this implication of the model is robustly contradicted in the data (e.g., Mehra and Prescott 1985). This equity premium puzzle suggests that the standard power utility model should not be used to assess the effect of risk in influencing asset prices. An alternative approach to gauge the impact of risk on prices is to note that for most investors, investing in predictive markets is likely to be a relatively small risk. There are studies suggesting that decision making in the presence of small risk is well-described by loss-aversion preferences that have a kink at the status quo level of wealth (see for example, Thaler, Tversky, Kahneman, and Schwartz 1997). Calibrating a model with such loss-averse investors would be an empirically more plausible way to assess the role of risk in affecting predictive market prices.

To conclude, this is an interesting paper that documents useful facts about the functioning of economic derivatives' markets. I hope that my discussion helps in suggesting additional empirical tests to sharpen our understanding of the mechanism through which these markets aggregate information.

References

Aumann, Robert. 1976. "Agreeing to Disagree." *The Annals of Statistics* 4: 1236-1239.

Eisenberg, Edmund, and David Gale. 1959. "Consensus of Subjective Probabilities: The Pari-mutuel Method." *Annals of Mathematical Statistics* 30: 165–168.

Grossman, Sanford. 1976. "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information." *Journal of Finance* 31: 573–585.

Mankiew, Gregory N., Ricardo Reis, and Justin Wolfers. 2003. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual.*

Manksi, Charles F. 2004. "Interpreting the Predictions of Prediction Markets." Working Paper no. 10359. Cambridge, MA: National Bureau of Economic Research.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15: 145–161.

Reny, Philip J., and Perry, Motty. 2003. "Toward a Strategic Foundation for Rational Expectations Equilibrium." Working Paper. University of Chicago and Hebrew University.

Thaler, Richard H., Amos Tversky, Daniel Kahneman, and Alan Schwartz. 1997. "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test." *Quarterly Journal of Economics* 112: 647–661.

Wolfers, Justin, and Eric Zitzewitz. 2005. "Interpreting Prediction Market Prices as Probabilities." Working Paper. University of Pennsylvania and Stanford University.