THE RELATIONSHIPS AMONG EXPECTED INFLATION, DISAGREEMENT, AND UNCERTAINTY: EVIDENCE FROM MATCHED POINT AND DENSITY FORECASTS

Robert Rich and Joseph Tracy*

Abstract—This paper examines matched point and density forecasts of inflation from the Survey of Professional Forecasters to analyze the relationships among expected inflation, disagreement, and uncertainty. We undertake the empirical analysis within a seemingly unrelated regression framework and derive measures of uncertainty using a decomposition proposed by Wallis (2004, 2005) and by drawing on the concept of entropy. The results offer little evidence that disagreement is a useful proxy for uncertainty and mixed evidence that increases in expected inflation are accompanied by heightened uncertainty. Conversely, we document a quantitatively and statistically significant positive association between disagreement and expected inflation.

I. Introduction

AGREEMENT is widespread that inflation expectations are important for understanding the behavior of individuals and observed macroeconomic outcomes. While a great deal of research continues to focus on how people form expectations, there is also interest in examining other aspects of predictive behavior and characterizing their relationships. For example, Zarnowitz and Lambros (1987) and Giordani and Söderlind (2003) investigate the link between the dispersion of individual mean forecasts of inflation (a measure of disagreement over inflation forecasts) and the average dispersion of corresponding density forecast distributions (a measure of uncertainty over inflation forecasts). This issue bears on the validity of using disagreement as a proxy for inflation uncertainty in empirical investigations. Other studies seek to determine if changes in anticipated inflation are associated with parallel changes in uncertainty about inflation. If this relationship holds, then an additional cost of rising inflation is the adverse real effects associated with increased uncertainty. In another study, Mankiw, Reis, and Wolfers (2003) test the prediction of the sticky information model of Mankiw and Reis (2002) of a positive association between the dispersion of individual mean forecasts of inflation and expected inflation at the aggregate level.

This paper examines matched point and density forecasts of inflation from the Survey of Professional Forecasters (SPF) to analyze the relationships among aggregate expected inflation, disagreement, and uncertainty. Our study improves on previous studies in terms of data construction and estimation methodology. With regard to data construction, we derive empirical measures of uncertainty using two alternative approaches. One approach draws on the work of Wallis (2004, 2005) and uses a decomposition of the variance of the aggregate density forecast distribution. The second approach applies the concept of entropy from information theory. While we argue that each approach has its own merits, the use of both approaches has the added benefit of allowing us to check the robustness of the results.

With regard to estimation methodology, the matched point and density inflation forecasts from the SPF involve four forecast horizons. Previous studies have either selected a single horizon for analysis or examined the horizons separately. We adopt a seemingly unrelated regression (SUR) approach in which we group the equations for each horizon. This choice of estimation strategy stems not only from theoretical considerations suggesting the regression residuals should be correlated across horizons, but also from formal statistical tests that confirm this feature of the data. The SUR framework provides efficiency gains relative to conventional estimation methods and also allows us to check the robustness of the results across different forecast horizons.

In the next section, we provide an overview of the SPF inflation data. Section III describes our econometric methodology. We present the empirical results in section IV. We conclude with a short summary of our findings.

II. Data

A. *Variable Definitions*

The SPF was jointly initiated in late 1968 by the National Bureau of Economic Research (NBER) and the American Statistical Association (ASA) and was first known as the NBER-ASA Economic Outlook Survey.[1] The SPF originally asked respondents to provide point forecasts for ten variables over a range of forecast horizons. Unlike other surveys, the questionnaire also solicits density forecasts for aggregate output and inflation in the form of histograms. That is, respondents are asked to attach a probability to each of a number of preassigned intervals, or bins, in which output growth and inflation might fall. Because these forecasts relate to the spread of a probability distribution of possible outcomes, they provide a unique basis to derive empirical measures of uncertainty.

We restrict our attention to data on the inflation forecasts due to the lack of a homogeneous sample for the output forecasts.[2] With regard to the density forecasts of inflation, in the fourth quarter, the survey asks respondents about the annual average percentage change in prices between the current year and the following year. In the first, second, and third quarters, however, the survey asks respondents about the annual average percentage change in prices between the current year and the previous year. Consequently, the target variable for the density forecasts remains fixed for four consecutive surveys (from the fourth quarter of year t through the third quarter of year $t+1$), with a corresponding forecast horizon (h) that declines from

Received for publication July 3, 2006. Revision accepted for publication March 10, 2008.

* Rich and Tracy: Federal Reserve Bank of New York.

We have benefited from the suggestions of attendees of the System Committee on Business and Financial Conditions Conference in Chicago. We thank J. S. Butler, Tim Cogley, John Ham, Bart Hobijn, Spence Krane, John Leahy, José Lopez, Simon Potter, Til Schuermann, Tom Stark, and Giorgio Topa for comments. We are especially grateful to Julio Rotemberg (the editor), Kenneth Wallis, and two anonymous referees for many helpful suggestions on previous drafts of this paper. Bess Rabin, Ariel Zetlin-Jones, and Kieran Walsh provided excellent research assistance. The views expressed in this paper are our own and do not reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

[1] Details about the SPF and its history can be found in Rich and Tracy (2006).

[2] Specifically, respondents switched from forecasting nominal output to real output in the early 1980s.

The Review of Economics and Statistics, February 2010, 92(1): 200–207

approximately 4½ quarters to 1½ quarters.[3] For convenience, we refer to these horizons as $h = 4, \ldots, 1$.

Defining notation, let $\phi_{h,t}(\pi)$ denote a respondent's h-quarter-ahead density forecast of inflation (π) in year t. Therefore, $\phi_{4,t}(\pi)$ will denote the respondent's density forecast in the fourth quarter ($h = 4$) of year t, while $\phi_{3,t+1}(\pi)$ will denote the subsequent density forecast in the first quarter ($h = 3$) of year $t + 1$. We then let $\phi^e_{h,t}(\pi)$ and $\sigma^2_{h,t}(\pi)$ denote, respectively, the mean and variance of the corresponding density forecast.

With regard to the point forecasts, the SPF asks respondents for predictions of the price level for the current quarter and the next four quarters. Because data are available on the price index in preceding quarters, a point forecast, $f^e_{h,t}$, can be constructed that matches each density forecast. Therefore, we will let $f^e_{4,t}$ denote a respondent's point forecast of the annual average percentage change in prices in the fourth quarter ($h = 4$) of year t. The subsequent point forecast of the annual average percentage change in prices in the first quarter ($h = 3$) of year $t + 1$ will be denoted by $f^e_{3,t+1}$.

Our study considers two alternative approaches to derive measures of uncertainty. The first is based on the statistical framework of Wallis (2004, 2005), which yields a formal relationship among measures of uncertainty and disagreement. Specifically, let $\bar{\phi}_{h,t}(\pi)$ denote the h-quarter-ahead aggregate density forecast of inflation in year t that averages the density forecasts across all respondents. As Wallis notes, the combined density forecast is an example of a finite mixture distribution. If we assume that each individual's point forecast ($f^e_{h,t}$) is the mean of the individual's forecast density ($\phi^e_{h,t}(\pi)$), then Wallis (2004, 2005) shows that the variance of $\bar{\phi}_{h,t}(\pi)$ can be expressed as[4]

$$\mathrm{Var}[\bar{\phi}(\pi)] = \bar{\sigma}^2_{\phi(\pi)} + s^2_{f^e}, \tag{1}$$

where, for convenience, we suppress the subscripts denoting the specific forecast horizon and year. The first term on the right-hand side of (1) is the average individual variance ($\bar{\sigma}^2_{\phi(\pi)}$), which provides the basis for a measure of aggregate uncertainty. The second term is the cross-sectional variance of the point forecasts ($s^2_{f^e}$), which provides the basis for the measure of disagreement. Given an estimate of $\mathrm{Var}[\bar{\phi}(\pi)]$ and a calculated value for $s^2_{f^e}$, the decomposition in (1) can be used to back out the values for the series $\bar{\sigma}^2_{\phi(\pi)}$. Because equation (1) is derived from moment conditions relating to the individual forecast densities and the aggregate density, we will refer to the uncertainty measure $\bar{\sigma}^2_{\phi(\pi)}$ as an (implied) moment-based measure of uncertainty.

Our second approach to derive a measure of uncertainty draws on information theory and the concept of entropy.[5] We can calculate the entropy of an individual SPF histogram as

$$H(\pi) = -\left(\sum_{k=1}^{n} p(k)[\ln(p(k))]\right), \tag{2}$$

where $p(k)$ denotes the probability that an individual attaches to interval k. The entropy is nonnegative and can attain a value of 0 when $p(k) = 1$ for one of the n bins. If we hold the number of bins fixed at n, then the entropy is maximized when $p(k) = (1/n)$. However, this maximum increases when the number of possible outcomes (n) increases and the bin width is held constant.[6]

Because the entropy relates to the extent to which the probability is concentrated on a few points or dispersed over many points, it lends itself to being interpreted as a measure of uncertainty. After calculating the value of (2) for each respondent, we then average the individual values to produce an entropy-based measure of aggregate uncertainty ($\bar{H}(\pi)$). An attractive feature of our entropy-based measure of uncertainty is that it can be derived directly from the individual histograms. This contrasts with the measure of average individual uncertainty in (1), which is derived from the aggregate density and is not free of the influence of disagreement about point forecasts.[7]

The previous discussion provides a general description of the approaches used to derive some of the key variables of interest. For the empirical analysis, however, we slightly modify the forecast density data due to periodic changes in the SPF's survey instrument. Specifically, the survey initially provided 15 intervals. From 1981:Q3–1991:Q4, the number of intervals was reduced to six. Since 1992:Q1, there have been ten intervals. More important, the interval widths also varied from 1 percentage point before 1981:Q3 and after 1991:Q4 to 2 percentage points in the intervening period.[8] The presence of varying interval widths poses a concern because it will have an impact on some of our summary measures and their movements across subperiods. This is particularly true for the measures of uncertainty. Therefore, we redefine the intervals to impose a common 2 percentage point width throughout the sample period.[9] This modification results in some loss of information, but it has the benefits of maintaining consistency across time in the data and ensuring that the results are not being driven by differential interval widths.[10]

The empirical analysis will also consider slightly modified versions of the measures of uncertainty and disagreement. From (1), our preferred measures of aggregate uncertainty and disagreement are $\bar{\sigma}_{\phi(\pi)}(= \sqrt{(\bar{\sigma}^2_{\phi(\pi)})})$ and $s_{f^e}(= \sqrt{s^2_{f^e}})$, respectively. Taking the square root of the uncertainty and disagreement variables makes their units of measurement coincide with that of inflation. With regard to the entropy-based measure of uncertainty in (2), we will make an additive log adjustment. Because the histogram-based entropy in (2) holds only when the bin width is equal to unity, we add an adjustment factor of ln(2).[11]

[3] Zarnowitz and Lambros (1987) select these values for the distances between the dates of the surveys and the end of the target year, where the horizons also reflect publication lags in the price index.

[4] Engelberg, Manski, and Williams (2009) provide evidence that most SPF forecasters give point predictions that are consistent with the means, medians, and modes of their density forecast distributions.

[5] Interested readers can consult Harris (2006) for a helpful discussion of entropy and its use in information theory.

[6] As discussed in Wallis (2006), equation (2) can be interpreted as a histogram-based approximation to the entropy of a continuous random variable.

[7] We recognize that it would be useful if the entropy-based measures of disagreement and uncertainty could be constructed along the same lines as in Wallis (2004, 2005). However, while the entropy for the aggregate density forecast of inflation can be calculated and decomposed into two terms, their interpretation would not be identical to those in (1). One of the terms would correspond to average uncertainty, but the other term would correspond to the dispersion in respondents' forecast uncertainty, not in their inflation forecasts.

[8] See table 1 of Rich and Tracy (2006) for details.

[9] Due to the odd number of intervals used over the subperiod 1968:Q4–1981:Q2, we use a unit interval length for the middle interval.

[10] The importance of this adjustment is readily apparent from a comparison of the entropy using the original and adjusted (2 percentage point) forecast density data. The profiles of the entropy differ markedly before 1981:Q3 and after 1991:Q4, with the use of the original data resulting in an artificial increase in the measure of uncertainty during the subperiods associated with the narrower intervals.

[11] See Wallis (2006) for further discussion of this point.

Further details on the construction of the variables for the empirical analysis as well as particular features of the SPF inflation data that bear on estimation of the relationships of interest can be found in Rich and Tracy (2006).[12] With regard to the latter point, some of the important features include dropping surveys from the analysis that were inadvertently subject to a mismatch between intended and requested forecast horizon, changes in the price index (and base year) used to define inflation, and the exclusion of respondents due to either their failure to provide matching point and density forecasts or discrepancies between their point and density forecasts that are judged to be excessive.[13] Rich and Tracy (2006) also provide a more detailed comparison to the approach in this study to that in Zarnowitz and Lambros (1987) and Giordani and Söderlind (2003).

III. Empirical Framework

The previous discussion focused on the construction of measures of expected inflation, disagreement, and uncertainty. We now turn to evaluating the economic and statistical significance of the various relationships of interest. To do so, we adopt the following linear regression model

$$Y = a + bX + \varepsilon, \tag{3}$$

where (Y, X) denote the relevant variables of interest, (a, b) denote the parameters of the model, and ε is a mean-zero, random disturbance term.

To gauge whether disagreement is a useful proxy for uncertainty, we consider the cases where $Y = [\ln(\bar{\sigma}_{\phi(\pi)}), \bar{H}(\pi)]$ and $X = s_{f^e}$, where the additional log transformation of $\bar{\sigma}_{\phi(\pi)}$ allows a more direct basis of comparison between the measures of uncertainty.[14] The use of the density forecast data to construct the uncertainty measures and the use of the point forecast data to construct the disagreement measure is similar to the approach of Zarnowitz and Lambros (1987). More important, the latter choice is appropriate given that the disagreement measure used as a proxy for uncertainty has always been calculated from point forecast data.

With regard to analyzing the contribution of expected inflation to movements in uncertainty and disagreement, we differentiate between the use of the density forecast data and the point forecast data.[15] However, we allow differences in the construction of the measures within each of the relationships we examine. To investigate the link between expected inflation and uncertainty, we consider $Y = [\ln(\bar{\sigma}_{\phi(\pi)}), \bar{H}(\pi)]$ and now set $X = \bar{\phi}^e(\pi)$, where $\bar{\phi}^e(\pi)$ denotes the mean of the aggregate density forecast. In the case of the link between expected inflation and disagreement, we consider $Y = \ln(s_{f^e})$ and $X = \bar{f}^e$, where $\bar{f}^e$ denotes the mean of the point forecasts.

The forecasting horizon for the SPF inflation data is not constant and instead declines from the fourth quarter of year t through the third quarter of year $t + 1$. Because of the variation in forecast horizons, it is more reasonable to treat the data as annual observations on four different series than as quarterly observations on a homogeneous series. By itself, this consideration would suggest estimation of the following regression equations across the individual horizons:

$$Y_{h,t} = a_h + b_h X_{h,t} + \varepsilon_{h,t}, \quad h = 4, 3, 2, 1, \tag{4}$$

where we allow the intercept and slope coefficients in (3) to vary across forecast horizons. Equation (4) is consistent with the approach of previous studies that have almost exclusively based their analysis on data for a single horizon or for individual horizons. We will argue, however, that the nature of the data lends itself to applying the SUR method.

While the different forecast horizons argue for separate equations for the data, it does not seem reasonable to view the equations as completely unrelated due to their sharing a common inflation target over four contiguous quarters. This feature of the survey suggests that the corresponding error terms $[(\varepsilon_{4,t}, \varepsilon_{3,t+1}, \varepsilon_{2,t+1}, \varepsilon_{1,t+1})]$ are likely correlated with each other. If this is the case, then it is possible to exploit the correlation structure of the error terms and apply the generalized-least-squares estimators proposed by Zellner (1962) to generate more efficient parameter estimates than those obtained by the application of OLS to each equation individually.[16]

Our SUR estimation strategy is standard except for one minor modification in which we group the equations based on their affiliation with the forecast horizon and target rate of inflation. In particular, we stack the four time series regressions as follows:

$$\begin{bmatrix} Y_{1,2} \\ \vdots \\ Y_{1,T} \\ \vdots \\ Y_{4,1} \\ \vdots \\ Y_{4,T-1} \end{bmatrix} = \begin{bmatrix} a_1 + b_1 X_{1,2} + \varepsilon_{1,2} \\ \vdots \\ a_1 + b_1 X_{1,T} + \varepsilon_{1,T} \\ \vdots \\ a_4 + b_4 X_{4,1} + \varepsilon_{4,1} \\ \vdots \\ a_4 + b_4 X_{4,T-1} + \varepsilon_{4,T-1} \end{bmatrix}, \tag{5}$$

where we order the equations from horizon $h = 1$ to horizon $h = 4$.[17] We will follow convention with regard to the structure of the variance-covariance matrix Ω. Specifically, we assume the disturbance term in any single equation is conditionally homoskedastic and nonautocorrelated, although allowance is made for the data to be conditionally heteroskedastic across equations.[18]

Following Breusch and Pagan (1980), we can construct the following Lagrange multiplier test to formally test for nonzero correlations between the disturbance terms in the four equations:

$$\lambda = T \sum_{m=1}^{i} \sum_{n=1}^{i-1} \rho_{mn}^2, \tag{6}$$

[12] Whenever the construct of variables required an estimate of the mean or variance of the aggregate or individual histograms, we adopted the assumption that all the probability mass is located at the interval midpoints.

[13] This is similar to Engelberg et al., (2009), who also find that some SPF forecasters, point predictions appear to be inconsistent with the means, medians, or modes of their density forecasts.

[14] Given that the dependent variables are strictly positive, the log transformation helps support the assumption of a normally distributed disturbance term.

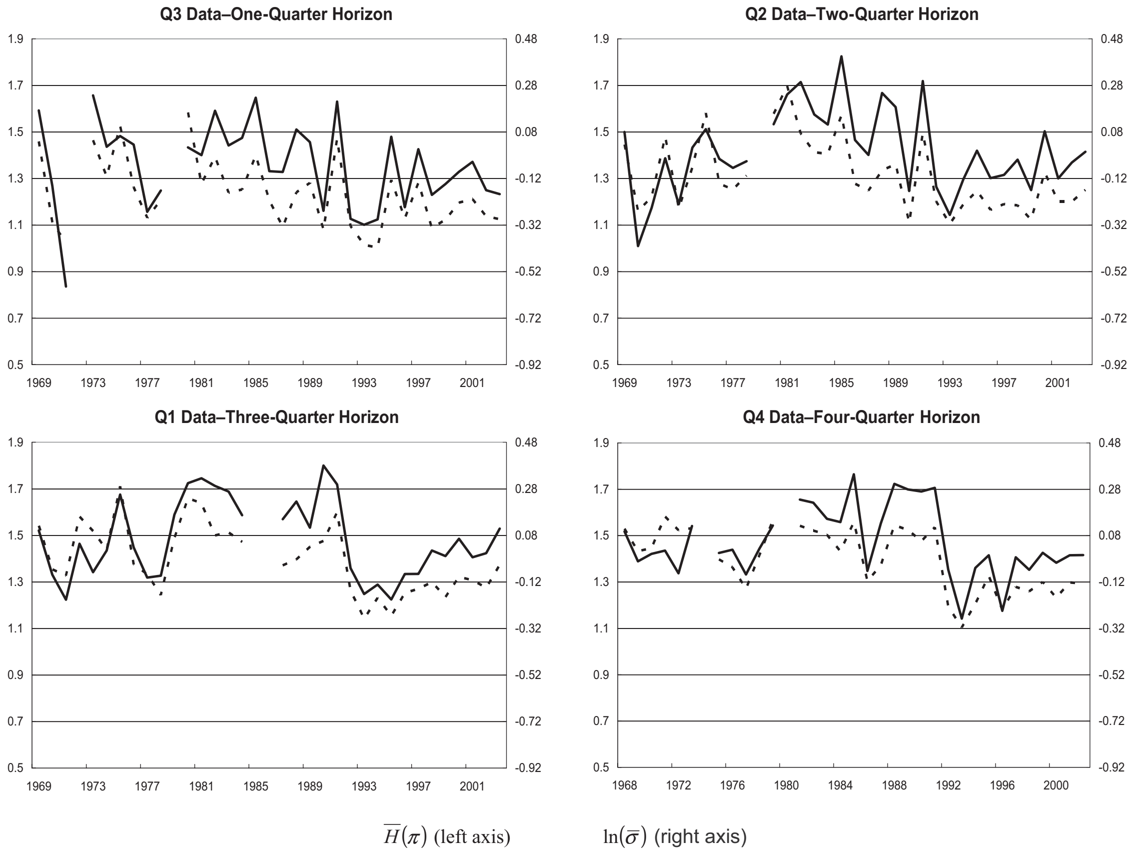
[15] Zarnowitz and Lambros (1987) adopted this same approach.

[16] Note that the explanatory variables will also not be identical across the forecast horizons. If this condition did not hold, then no gains in efficiency could be realized from the SUR estimator over the OLS estimator.

[17] We assume that the number of observations on each equation is the same. This accounts for the slight difference in the time subscripts for the data associated with horizons $h = 1, 2, 3$ ($t = 2, \ldots, T$) and horizon $h = 4$ ($t = 1, \ldots, T - 1$).

[18] The assumptions underlying the specification of Ω are broadly consistent with the data. Based on diagnostic checks of the OLS residuals of the individual regression equations, we decided it was not necessary to incorporate additional own- and cross-covariance terms into Ω.

FIGURE 1.—MEASURES OF AVERAGE UNCERTAINTY



where ρ_{mn} is the estimated correlation between the OLS residuals of the $i = 4$ equations and T is the number of observations in each equation. The tests statistic is distributed asymptotically as a chi square random variable with $i(i - 1)/2$ degrees of freedom under the null hypothesis of zero correlation between the disturbance terms.

IV. Empirical Results

A. *Measures of Expected Inflation, Disagreement, and Uncertainty*

Figures 1 to 3 present the time profiles for the measures of uncertainty, disagreement, and expected inflation used in the empirical analysis.[19] Abstracting from differences in the mean of the series, figure 1 shows that the behavior of the two uncertainty measures is qualitatively similar. As expected, they tend to reflect a greater dispersion of intrapersonal probabilistic beliefs as the forecast horizon increases, although there is a surprising slight decline at the $h = 4$ quarter horizon. In addition, both measures depict a decline in inflation uncertainty starting around 1990. On the other hand, the evidence that inflation uncertainty was higher during the 1970s than in the 1980s appears mixed.

With regard to the disagreement measure, figure 2 indicates a greater diversity of opinion about expected inflation during the earlier part of the sample period. Dispersion of point forecasts also tends to increase as the forecast horizon lengthens, with the cross-sectional standard deviation of the point forecast characterized by occasional spikes in disagreement. When we examine the measures of expected inflation in figure 3, however, we observe that the series display a high degree of conformity and are practically indistinguishable from each other. The two inflation expectations series display the same pronounced rise and subsequent decline as actual inflation during the course of the sample period.

B. *Estimated Relationships*

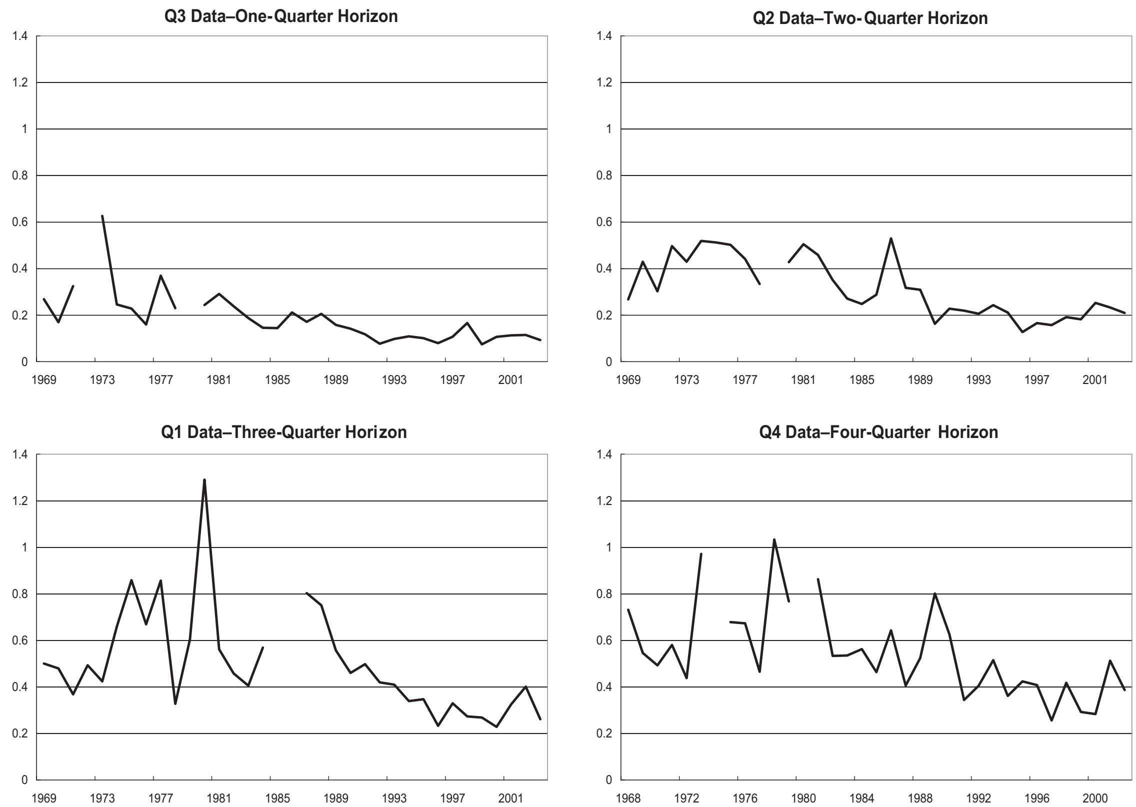
The sample covers the surveys conducted from 1968:Q4 through 2003:Q3, so that the values on the realized annual rate of inflation cover the periods 1968–1969 through 2002–2003. We begin by examining correlations and goodness-of-fit measures from the OLS estimation of equation (4) for the various relationships. The results are reported in the second and fifth columns in table 1.[20] Because we subsequently address the issue of estimation efficiency, we defer for the moment from any discussion of statistical significance and instead focus on the economic significance of the relationships.

As shown, the variables display a positive association across all of the relationships. With regard to any systematic pattern to the

[19] The missing observations in figure 1 (as well as in subsequent figures) reflect the excluded survey dates.

[20] We recognize there is little difference in the information conveyed by the reported correlations and the $\bar{R}^2$'s, as the latter simply involves squaring the former and adjusting for degrees of freedom. Nevertheless, we report both statistics to allow for a basis of comparison to the results of other studies.

FIGURE 2.—MEASURES OF DISAGREEMENT



correlations, they do not behave in a monotonic fashion as the forecast horizon increases. For the relationships between disagreement and uncertainty, as well as between uncertainty and expected inflation, the correlations are highest at the $h = 3$ quarter horizon. For the relationship between disagreement and expected inflation, the correlations are highest at the $h = 2$ and 4 quarter horizons. However, several other notable findings emerge from the analysis.

One is immediately struck by the extremely low explanatory content of disagreement for movements in the moment-based measure of average uncertainty. With the exception of the regression associated

FIGURE 3.—MEASURES OF EXPECTED INFLATION

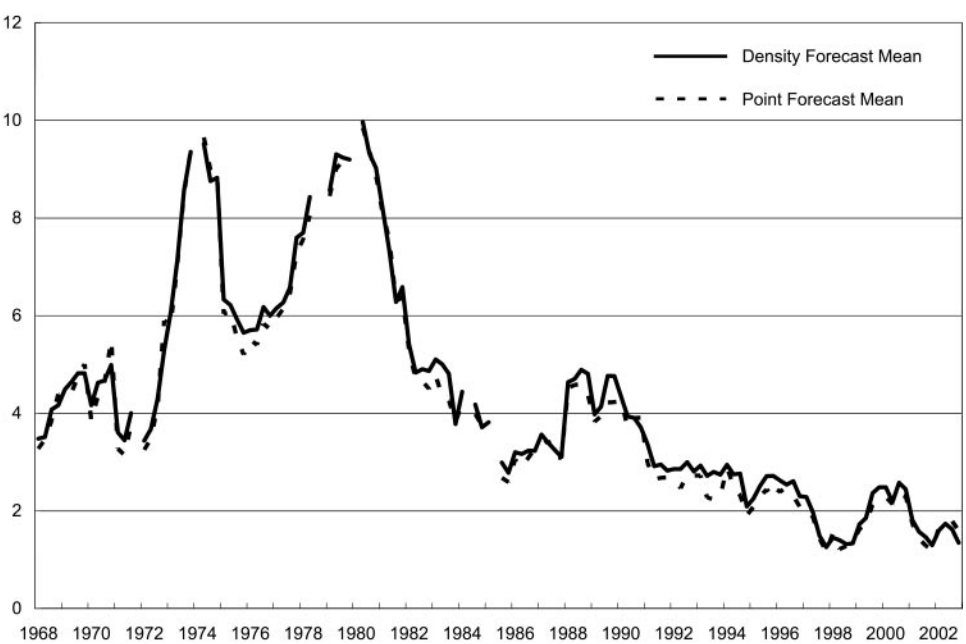


TABLE 1.—RELATIONSHIPS AMONG EXPECTED INFLATION, DISAGREEMENT, AND UNCERTAINTY

	Disagreement and Uncertainty					
	$\ln(\bar{\sigma}_{\phi(\pi)}) = a + b \times s_{f^e} + \varepsilon$			$\bar{H}(\pi) = a + b \times s_{f^e} + \varepsilon$		
	Correlations			Correlations		
Horizon/Quarter	$r\ (\bar{R}^2)$	Intercept	Slope	$r\ (\bar{R}^2)$	Intercept	Slope
$h = 1/Q3$	0.22/(0.02)	−0.144** (0.052)	0.458* (0.232)	0.37/(0.11)	1.164** (0.040)	0.434** (0.176)
$h = 2/Q2$	0.15/(−0.01)	−0.021 (0.063)	0.099 (0.177)	0.48/(0.21)	1.214** (0.049)	0.345** (0.138)
$h = 3/Q1$	0.44/(0.17)	−0.030 (0.053)	0.215* (0.094)	0.57/(0.30)	1.297* (0.042)	0.228** (0.075)
$h = 4/Q4$	0.29/(0.05)	0.002 (0.067)	0.115 (0.115)	0.49/(0.22)	1.281** (0.056)	0.234* (0.097)
	Uncertainty and Expected Inflation					
	$\ln(\bar{\sigma}_{\phi(\pi)}) = a + b \times \bar{\phi}^e(\pi) + \varepsilon$			$\bar{H}(\pi) = a + b \times \bar{\phi}^e(\pi) + \varepsilon$		
	Correlations			Correlations		
Horizon/Quarter	$r(\bar{R}^2)$	Intercept	Slope	$r(\bar{R}^2)$	Intercept	Slope
$h = 1/Q3$	0.22/(0.02)	−0.149* (0.062)	0.020 (0.012)	0.51/(0.24)	1.111** (0.044)	0.032** (0.009)
$h = 2/Q2$	0.29/(0.06)	−0.100 (0.061)	0.024* (0.012)	0.65/(0.41)	1.133** (0.042)	0.043** (0.008)
$h = 3/Q1$	0.48/(0.21)	−0.059 (0.049)	0.031** (0.009)	0.66/(0.42)	1.246** (0.036)	0.038** (0.006)
$h = 4/Q4$	0.35/(0.10)	−0.048 (0.055)	0.027* (0.012)	0.52/(0.26)	1.241** (0.044)	0.041** (0.009)
	Disagreement and Expected Inflation					
	$\ln(s_{f^e}) = a + b \times \bar{f}^e + \varepsilon$			$s_{f^e} = a + b \times \bar{f}^e + \varepsilon$		
	Correlations			Correlations		
Horizon/Quarter	$r(\bar{R}^2)$	Intercept	Slope	$r(\bar{R}^2)$	Intercept	Slope
$h = 1/Q3$	0.70/(0.47)	−2.393** (0.122)	0.139** (0.025)	0.58/(0.31)	0.078* (0.030)	0.025** (0.006)
$h = 2/Q2$	0.72/(0.51)	−1.768** (0.098)	0.129** (0.020)	0.74/(0.53)	0.145** (0.029)	0.041** (0.006)
$h = 3/Q1$	0.68/(0.44)	−1.250** (0.102)	0.110** (0.020)	0.65/(0.41)	0.240** (0.058)	0.059** (0.012)
$h = 4/Q4$	0.74/(0.53)	−1.179** (0.092)	0.130** (0.021)	0.75/(0.55)	0.246** (0.050)	0.076** (0.011)

Note: One-tailed test for statistical significance of slope coefficient. H_0: $\lambda b = 0$, λH_1: $\lambda b > 0$. ** Significant at 1% level. * Significant at 5% level.

with the $h = 3$ quarter horizon, disagreement accounts for 5% or less of the variation in uncertainty.[21] The results are qualitatively similar when we turn to the link between expected inflation and the moment-based measure of uncertainty in the middle panel. The lack of any meaningful co-movement between the variables suggests that the issue of the statistical significance of these relationships is largely irrelevant for the remaining analysis.

There is an increase in the strength of these same relationships when the entropy-based measure of uncertainty is used in the regressions. For example, the correlation between expected inflation and the entropy-based measure of uncertainty exceeds 0.6 at the $h = 2$ and $h = 3$ quarter horizons, with the other correlations being of moderate size. For the disagreement measure, however, the magnitude of the correlations would still suggest a weak relationship with the entropy-based measure of uncertainty. The evidence indicates that disagreement is not a good proxy for uncertainty. Moreover, this conclusion is robust to concerns that measured disagreement may be unduly influenced by outliers in the sample of point forecasts.[22]

As shown in the bottom panel of table 1, the evidence is much less ambiguous and quite favorable about a positive co-movement between disagreement and expected inflation. The correlations and regression fit indicate a fairly strong association between the variables that is not only robust across the forecast horizons, but also to the use of a nontransformed value of the dependent variable. It is also interesting to note that of the three relationships examined in the paper, the link between disagreement and expected inflation has received the least attention by researchers.

To address the issue of statistical significance in the relationships, we applied the Breusch-Pagan (1980) testing procedure to the OLS

[21] Giordani and Söderlind (2003) focus their analysis on the data for the $h = 3$ quarter horizon.

[22] Following Giordani and Söderlind (2003) and Boero, Smith, and Wallis (2008), we also used the quasi-standard deviation of the point forecasts as a measure of disagreement. The results using this alternative measure of disagreement were very similar to those in table 1 in terms of economic and statistical significance.

residuals within each system of four equations. In all cases, the values of the test statistic strongly rejected the assumption of an absence of correlation between the equations' disturbance terms associated within the same inflation target.[23] Consequently we proceeded to reestimate the relationships using the SUR method.

The estimates of the parameters and the corresponding standard errors for the various relationships are reported in the relevant columns in table 1.[24] Because the definition of an $\bar{R}^2$ statistic is not obvious in the case of SUR estimation, we do not report any goodness-of-fit statistic. Moreover, because most researchers typically posit a positive relationship between the variables, we conduct a one-tailed test for statistical significance. The conclusions, however, generally will not depend on the choice of a one- versus two-tailed test for statistical significance.

As shown, the findings typically document a statistically significant positive association between the variables in the relationships. Not surprisingly, the qualitative features of the results parallel those from the previous analysis in terms of forecast horizon and data construct. That is, the statistical significance of the relationships between disagreement and uncertainty, as well as between expected inflation and uncertainty, is less robust using the moment-based measure of uncertainty than the entropy-based measure. Moreover, the relationship between disagreement and expected inflation remains highly statistically significant.

Our findings concerning the estimated relationship between disagreement and uncertainty are closer to those reported by Zarnowitz and Lambros (1987) and contrast sharply with the conclusions of Giordani and Söderlind (2003). With regard to the latter study, their results likely differ because of their use of a smoother measure of uncertainty, as well as a measure of disagreement that pertains to the density forecast data.[25] On the other hand, our findings concerning the relationship between disagreement and expected inflation are much stronger than those of Zarnowitz and Lambros (1987) in terms of economic and statistical significance. This is a consequence of the longer sample period used in our analysis.[26]

V. Conclusion

Our study uses matched point and density forecasts of inflation from the Survey of Professional Forecasters to revisit questions concerning the co-movements between expected inflation at the aggregate level, the degree of disagreement among individual inflation forecasts, and the level of average inflation uncertainty. We attempt to improve on previous studies in terms of the construction of the measures used for the empirical analysis as well as the statistical methods used to assess the nature of the relationships.

The results reveal that the alternative data constructs for inflation uncertainty lead to different associations with expected inflation and disagreement. The relationship between expected inflation and uncertainty displays little economic significance when we employ a moment-based measure of uncertainty. On the other hand, we find an economically and statistically significant relationship between expected inflation and the entropy-based measure of uncertainty. The lack of robustness of these results leads us to conclude that the empirical relevance of one of the posited channels of effect of expected inflation on real activity remains an open question.

With regard to the relationship between uncertainty and disagreement, we again find a stronger association between the variables using the entropy-based measure of uncertainty. Nevertheless, the correlations are generally too weak across both uncertainty measures to support the use of disagreement as a proxy for inflation uncertainty, thereby raising questions about the validity of previous empirical findings based on this practice.[27] Further support for this concern is provided by Boero, Smith, and Wallis (2008) who examine point and density forecast data from the Bank of England's Survey of External Forecasters and also find a weak correlation between measures of disagreement and uncertainty.

In contrast to the previous results, the evidence is much clearer in terms of the relationship between disagreement and expected inflation. Specifically, we find that more diversity among respondents' point predictions of inflation coincides with increases in expected inflation, with the link between the variables displaying both economic and statistical significance. While we are cautious about the interpretation and implications of these findings at the aggregate level for specific models of expectations formation, we acknowledge that the positive co-movement between disagreement and expected inflation appears to be an important feature of predictive behavior and an issue warranting greater attention on the part of researchers.[28]

[23] The values of the test statistic rejected the null hypothesis at the 1% significance level. We do not report these results to conserve space.

[24] There is an unequal number of observations across the equations. However, we ignored this difference and calculated the own- and cross-covariances using all available observations.

[25] Specifically, Giordani and Söderlind (2003) use normal approximation methods to the individual SPF histograms to derive their uncertainty measure.

[26] Zarnowitz and Lambros (1987) base their findings on a sample that runs from 1986:Q4–1981:Q2.

[27] An extensive literature has used forecast dispersion measures from surveys of inflation expectations as a proxy for inflation uncertainty. Zarnowitz and Lambros (1987) and Giordani and Söderlind (2003) contain references to various studies that have sought to determine the effect of inflation uncertainty on macroeconomic and financial variables such as output growth, unemployment, nominal interest rates, and labor contract durations.

[28] This finding initially might be viewed as corroborating evidence in support of the sticky information model of Mankiw and Reis (2002). In a related paper, however, Rich and Tracy (2004) argue that another implication of the sticky information model is that there should be no persistent differences across SPF respondents in their forecast behavior. When we examine the SPF inflation data at the individual level, we strongly reject the model's prediction that there are no significant fixed effects associated with either the respondents' ex ante forecast uncertainty or their ex post forecast accuracy.

REFERENCES

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis, "Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters," *Economic Journal* 118 (2008), 1107–1127.

Breusch, Trevor S., and Adrian R. Pagan, "The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics," *Review of Economic Studies* 47 (1980), 239–253.

Engelberg, Joseph, Charles F. Manski, and Jared Williams, "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters," *Journal of Business and Economic Statistics* 27 (2009), 30–41.

Giordani, Paul, and Paul Söderlind, "Inflation Forecast Uncertainty," *European Economic Review* 47 (2003), 1037–1059.

Harris, B. "Entropy." In *Encyclopedia of Statistical Science* (Hoboken, NJ: Wiley, 2006).

Mankiw, N. Gregory, and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics* 117 (2002), 1295–1328.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers, "Disagreement about Inflation Expectations" (vol. 18, pp. 209–248), in *NBER Macroeconomics Annual 2003* (Cambridge, MA: MIT Press, 2003).

Rich, Robert, and Joseph Tracy, " 'Sticky Information' and Heterogeneity in Forecast Behavior: Further Implications and Evidence from the Survey of Professional Forecasters," Federal Reserve Bank of New York unpublished paper (2004).

——— "The Relationship between Expected Inflation, Disagreement and Uncertainty: Evidence from Matched Point and Density Forecasts," Federal Reserve Bank of New York staff reports no. 253 (July 2006).

Wallis, Kenneth F., "Forecast Uncertainty, Its Representation and Evaluation," IMS Singapore tutorial lectures (May 3–6 2004).

——— "Combining Density and Interval Forecasts: A Modest Proposal," *Oxford Bulletin of Economics and Statistics* 67 (2005 Supplement), 983–994.

——— "A Note on the Calculation of Entropy from Histograms," University of Warwick unpublished paper (2006).

Zarnowitz, Victor, and Louis A. Lambros, "Consensus and Uncertainty in Economic Prediction," *Journal of Political Economy* 95 (1987), 591–621.

Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* 57 (1962), 348–368.

GOVERNMENT OVERSIGHT OF PUBLIC UNIVERSITIES: ARE CENTRALIZED PERFORMANCE SCHEMES RELATED TO INCREASED QUANTITY OR QUALITY?

A. Abigail Payne and Joanne Roberts*

Abstract—Universities are engaged in many activities, primarily research and teaching. Many states have instituted performance measures that focus on evaluating a university's success in teaching. We suggest that multitasking may be important in this context, and we consider research outcomes after adoption. We find striking results that depend on university status. Research activity is higher at flagship institutions after the adoption of performance measures. Most of this increase in activity is with respect to the level of research funding and the number of articles produced. In contrast, research funding and the number of publications is dramatically lower at nonflagship institutions. There is some evidence that citations per publication at nonflagship institutions are higher after the adoption of performance standards. The evidence suggests that universities have become more specialized since the introduction of these programs.

I. Introduction

STATE governments have actively encouraged higher-quality teaching and better student experiences at public universities by using performance measures in their budgeting allocations. Are incentives targeted toward university teaching correlated with changes in university research activity? We find that research intensity changes after a state adopts these performance incentives. The change in intensity is correlated with the university's historical research stature.

Our study is focused on understanding whether the production of research is altered by the introduction of performance measures that capture only part of a university's activities devoted to teaching. If performance schemes generate stronger incentives for some activities than others, then these schemes may induce a distortion of incentives along the lines suggested by Holmstrom and Milgrom (1991, 1994). In this case, we should see an increase in the rewarded activities and a reduction in activities that are not rewarded. This reduction will be particularly severe if less intensely rewarded activities are hard to observe and measure. In this paper, we study research performance along three dimensions—research funding, number of publications, and number of citations per publication—to separate the changes in research quantity and quality after nonresearch-based performance schemes are adopted by a state government. This is interesting since quality is harder to observe and thus harder to encourage.[1]

We study two types of performance schemes adopted by state governments, both of which establish criteria for university evaluation. The first scheme links these criteria to an identified pot of funding, and the second uses these criteria in a more discretionary manner when allocating the state budget across public universities. These programs have been quite controversial.[2] Using survey data, Banta et al. (1996) have shown some concern about the measures used and the comparisons made across institutions with different missions.[3]

Our analysis exploits the heterogeneity in the time of adoption across states. We look at the relationship between these state-level policy adoptions and university-level outcomes. We control for many things affecting university governance at both the state and the university levels and study the within-university effects of the changes in the performance schemes on both the quality and quantity of

Received for publication November 12, 2004. Revision accepted for publication August 5, 2008.

* Payne: McMaster University; Roberts: University of Toronto and University of Calgary.

We thank Patti Tilson for invaluable research assistance. We received extremely helpful comments from the editor and journal referees, the participants of the NBER Higher Education Working Group, and seminar participants at Simon Fraser University, Brock University, McMaster University, University of Calgary, and Vanderbilt University. J.R. thanks SSHRC and CIFAR for research support. A.A.P. thanks the Andrew W. Mellon Foundation and SSHRC.

[1] As discussed by Dixit (2002) and Dewatripont, Jewitt, and Tirole (1999), multitasking problems can be particularly severe in the public sector. Our paper fits directly into this literature on the political control of public sector employees. Acemoglu, Kremer, and Mian (2008) consider the role of commitment in multitasking problems. If governments can commit to lower-powered incentive schemes, this can provide goods for which high-powered incentives are particularly detrimental.

[2] The concerns cover such things as the standards reflecting poor measures of quality (standardized testing that has no impact on student grades), the dependence on measures that are beyond university control (e.g., the state of the economy has a large impact on placement rates), and the use of measures that could result in the lowering of academic standards for students (e.g., inflating grades in an effort to increase student retention rates).

[3] It is clear that performance funding has been implemented in different ways in different states. Tennessee has been widely acknowledged as an example of a stable program. Burke and Mondaressi (2000) surveyed policymakers in an effort to uncover features that are important for program stability. In what they report of their survey evidence, implications for research were not mentioned. Our paper instead focuses solely on research outcomes.

The Review of Economics and Statistics, February 2010, 92(1): 207–212