Research Program on Forecasting

What Does Forecaster Disagreement Tell Us about the State of the Economy?

Constantin Bürgi
St. Mary's College of Maryland
St. Mary's City, MD
crburgi@smcm.edu

Tara M. Sinclair
The George Washington University
Washington, DC
tsinc@gwu.edu

RPF Working Paper No. 2020-001
https://www2.gwu.edu/~forcpgm/2020-001.pdf

January 27, 2020

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
https://www2.gwu.edu/~forcpgm

What Does Forecaster Disagreement Tell Us about the State of the Economy?*

Constantin Bürgi and Tara M. Sinclair

January 27, 2020

Abstract

This paper shows in a simple model that the part of uncertainty measured by forecaster disagreement rises in advance of and during recessions. Subsequently, it is tested using the Survey of Professional Forecasters in a dynamic probit model. It is shown that increases in disagreement help predict recessions in an out of sample context for the US.

JEL: C22, C52, C53, E17, E37

Keywords: Expectations, SPF, Uncertainty, Dynamic Probit

---

*Bürgi: St. Mary's College of Maryland, 47645 College Drive, St. Mary's City, MD, 20686, crburgi@smcm.edu; Sinclair: The George Washington University, 2115 G Street, NW, # 340, Washington, DC, 20052, tsinc@gwu.edu. We would like to thank an anonymous referee, Herman Stekler, James Foster, Jane Ryngaert, and Pavel Potuzak, as well as the participants in the IAES conference, the SAGE seminar, and the the Futures Past: Economic Forecasting in the 20th and 21st Century workshop for their valuable comments and support.

1 Introduction

Professional forecasters have difficulties predicting recessions as defined by the NBER (see Zarnowitz (1986), Fintzen and Stekler (1999) and Sinclair et al. (2010)). Bürgi (2016) shows that information on disagreement can improve the in sample recession prediction one quarter out in a dynamic probit model.[1] This note extends the approach by creating a theoretical foundation and testing its out of sample performance for real GDP growth using forecasts from the US Survey of Professional Forecasters (SPF). Disagreement is measured by the square root of the modified coefficient of variance (mCV) as shown in Figure 1. We show that increases in forecaster disagreement provide useful information for predicting recessions.

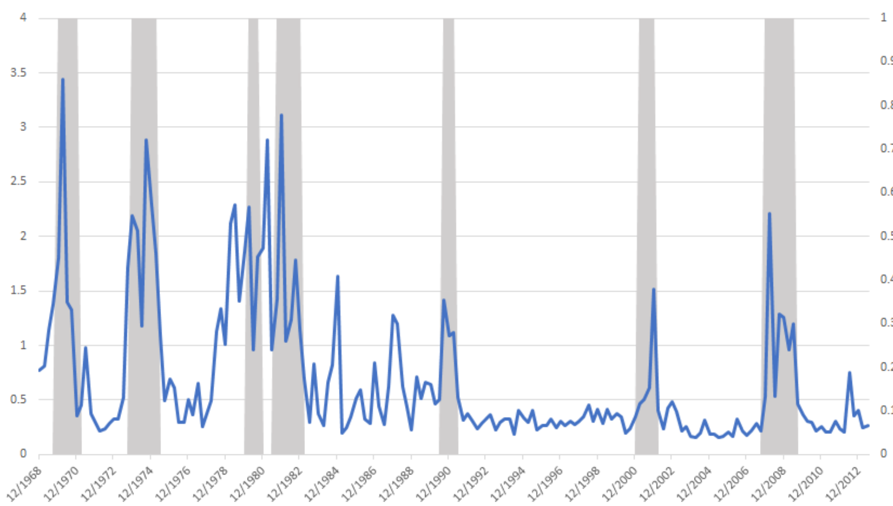

Figure 1: Modified coefficient of variation for one quarter ahead GDP forecast with recession shading for the forecasted period. The modified coefficient of variance is defined as $mCV = \frac{\sigma}{\max(1, abs(\mu))}$, where σ is the standard deviation across forecasters and μ is the simple average of forecasts.

[1] Driver et al. (2013) showed that disagreement can improve the prediction of underlying variables. Patton and Timmermann (2010) and Döpke and Fritsche (2006) both find that disagreement rises in recessions. For further information on disagreement and uncertainty, see Batchelor and Dua (1993), Bomberger (1996), Lahiri and Sheng (2010) and Ozturk and Sheng (2016).

2 Model

To show how a model with disagreement can predict recessions better than the simple average of forecasts alone, assume forecaster j forms rational expectations for period $t+1$ as follows:

$$F_{jt,t+1} = ax_{jt} + (1-a)F_{jt-1,t+1} \tag{1}$$

That is, the forecast $F_{jt,t+1}$ for period $t+1$ made in period t is the weighted average of the forecast for period $t+1$ made in the previous period $F_{jt-1,t+1}$ and some noisy new information x_{jt}. The weight a is the Bates and Granger (1969) optimal inverse variance weight. This approach modifies the Sims (2003) noisy information model by having a fixed event instead of a fixed horizon.

Now assume that part of the signal x_{jt} received in period t is that there will be a recession in period $t+1$. This signal can create large disagreement in the following two cases: First, agents might all update their prediction, but might receive quite different information regarding the severity of the recession. This means that some predict barely negative growth, while others predict a deep recession. In the second case, there is rigidity as in Mankiw and Reis (2002) and not all agents update their forecast. Even if all agents that update receive the same information, there will be large disagreement overall due to the ones that did not update. The ones that update their prediction will predict negative growth, while the ones that do not will still predict positive growth. These cases are also broadly in line with Baker et al. (2018). In both cases the simple average will decline somewhat, but it might not be clear if a recession is expected or not. Disagreement however will become substantially larger when recessions are anticipated by at least some forecasters than in normal periods when forecasters tend to bunch together (e.g. see Bürgi and Sinclair (2017)).

3 Estimation

We estimate a dynamic probit model similar to Dueker (1997) and Proaño and Theobald (2014):

$$\phi_t = \alpha + \beta y_{t-1} + \gamma X_{t-1} + \varepsilon_t, \tag{2}$$

where ϕ_t is the NBER recession dummy that takes value one if there is a recession and value zero otherwise, y_{t-1} is the real GDP growth rate for period $t-1$ that was available at the end of the second month of quarter t (real time data), and X_{t-1} are additional variables included in the regression.[2]

We begin with some in sample analysis. We first consider a simple Probit model which includes the lagged GDP growth rate, our disagreement measure, mCV, and the mean of the SPF forecast.[3] All variables are available from Q4 1969 through Q2 2013, our sample thus has 179 observations. Table 1 reports the results. The key finding is that the disagreement adds statistically significant information above and beyond both past GDP growth and the mean of the SPF forecast and has the expected sign.

Next, three models are compared in sample: no X_{t-1}, X_{t-1} being equal to mCV, and X_{t-1} being equal to the mean SPF forecast. In addition to these models, a naïve model is added that never predicts a recession ($\phi_t = 0$ for all t).

Since we are focused on a binary outcome, the performance of these models is assessed on the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate for all possible probability thresholds (for background on the ROC, see Berge and Jordà (2011).[4] The area is 0.5 if the probabilities are a coin toss and

[2]For real time data we use the second release which is released at the end of the second month of the quarter. This release has a slight informational advantage compared to the SPF forecast made earlier that month.

[3]Results are similar if we use the median of the SPF, see Mboup et al. (2018). We also explored models with CPI inflation and the unemployment rate. For inflation we did not find a cyclical pattern in disagreement, consistent with the inconsistent cyclicality of inflation in this period. For the unemployment rate we found a weaker, but similar, pattern in disagreement to what we report for GDP growth.

[4]Proaño and Theobald (2014) instead use three measures that are commonly used for continuous outcomes,

Table 1: Single model probit results in sample

Full sample (N = 179)	Coef.	Std. Err.
Constant	-0.98**	0.39
Past GDP growth	-0.03	0.04
Mean SPF forecast	-0.23**	0.10
Disagreement (mCV)	0.75***	0.24

*, **, *** significant at the 10%, 5%, and 1% level respectively

the area is 1 if there is a perfect fit. For most model comparisons, the ROC curves intersect, which means that none of the models is dominant for all probability thresholds. That is, one model might be better if 30% is the threshold probability for recessions and another model might be better at an 80% threshold. Thus, a larger area under the curve does not necessarily reflect an improvement at all thresholds, but can still be a guideline as to which model is better.

Table 2: Separate probit models ROC results in sample

Full sample (N = 179)	ROC Area
Disagreement (mCV)	0.87
Mean SPF forecast	0.88
Past GDP growth alone	0.75
Naïve no recession	0.50

Table 2 reports the in sample performance of the different models. In sample, the best model is the one which includes the mean of the SPF forecasts, closely followed by the model

namely root mean squared error (RMSE), mean absolute error (MAE), and the Theil coefficient. For robustness we considered these measures as well and found that the results were broadly consistent with our findings based on ROC.

which includes our disagreement measure, mCV. There is a clear gap between these two models and the models without X_{t-1}.

Table 3: Out of sample results (Estimation through Q4 1987)

Post 1987 performance (N=103)	ROC Area
Disagreement (mCV)	0.84
Mean SPF forecast	0.83
Past GDP growth alone	0.73
Naïve no recession	0.50

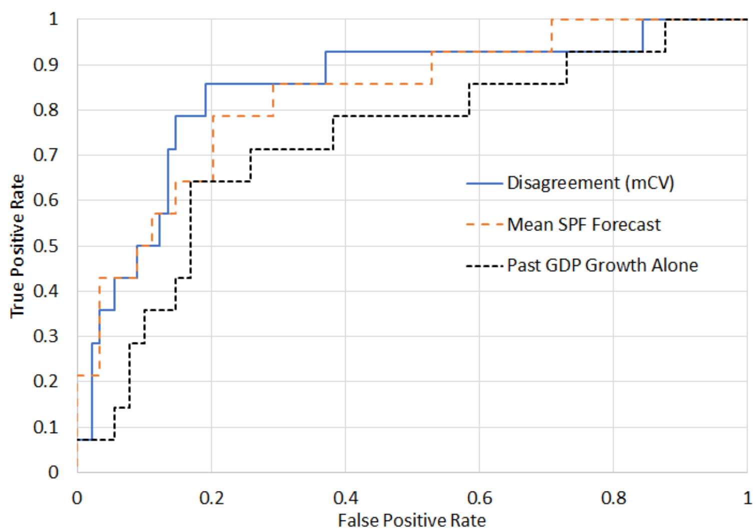

Figure 2: ROC Curves for the three models starting 1987

What is important, however, is how the models perform out of sample (OOS), not in sample. We therefore perform two different (pseudo) OOS exercises.[5] For our first OOS

[5]This analysis is pseudo out of sample in the sense that we are performing this analysis after the events occurred, but we estimated the model only based on data through Q4 of 1987 and 1995 respectively. Also note that if both the mean SPF forecast and disagreement are included, the out of sample results are indistinguish-

exercise shown in Table 3 and Figure 2, the same models as in Table 2 are estimated again using data only through Q4 1987. Subsequently, the models are used to predict the recession probabilities from Q1 1988 onwards, which includes the next three recessions (1990, 2001 and 2008). The mCV model performs best overall, but only slightly better than the mean SPF forecasts.

Table 4: Out of sample results (Estimation through Q4 1995)

Post 1995 performance (N=71)	ROC Area
Disagreement (mCV)	0.86
Mean SPF Forecast	0.84
Past GDP growth alone	0.73
Naïve no recession	0.50

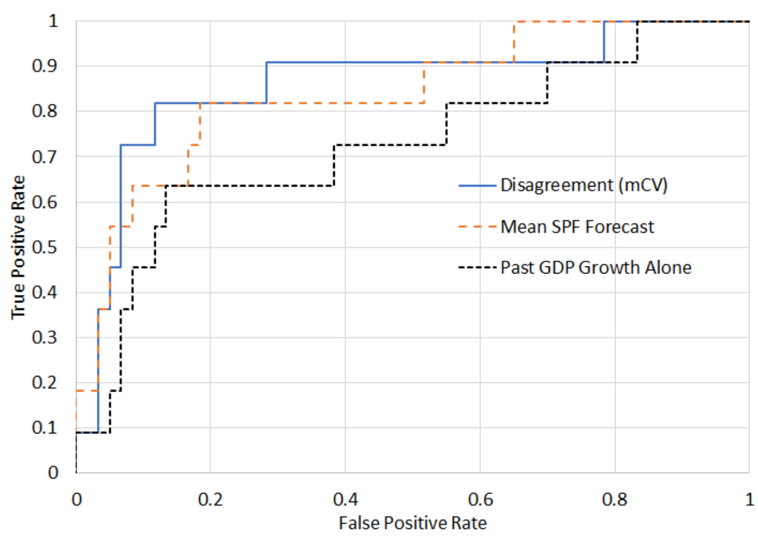

Figure 3: ROC Curves for the three models starting 1995

For our second OOS exercise the models are estimated over a longer period, through Q4 of

able from the ones of disagreement only.

1995, and their performance compared for the 2001 and 2008 recessions. The results become stronger as shown in Table 4 and Figure 3.

In both cases, the ROC curves with disagreement or the SPF mean dominate the model with just past GDP growth. The disagreement model dominates the GDP growth model more. No direct comparison between the disagreement and SPF mean models is possible, as the ROC curves cross.

4 Conclusion

This paper shows that the part of uncertainty measured by disagreement in the Survey of Professional Forecasters can aid the prediction of recessions in an out of sample dynamic probit model. Indeed, while disagreement does not dominate the average forecasts at all probability thresholds, it dominates the past GDP growth model by more than the average when business cycles since 1987 or 1996 are predicted.

The results presented in this paper leave several questions for future research. First, we are focused here on uncertainty about the future state of the economy, but other forms of uncertainty, such as those discussed in Bloom (2009) may also be a source of change in the state of the economy themselves. Learning more about how different types of uncertainty interact might improve forecasting models further. Second, we presented two potential reasons for why disagreement rises around recessions, and we would like to identify which one is the main driver. Finally, while disagreement is predictive of recessions, it was not compared to other approaches. Future research might compare it to other measures like the slope of the yield curve.

References

Baker, S. R., McElroy, T., and Sheng, X. S. (2018). Expectation formation following large unexpected shocks.

Batchelor, R. and Dua, P. (1993). Survey vs arch measures of inflation uncertainty. *Oxford Bulletin of Economics and Statistics*, 55(3):341–353.

Bates, J. and Granger, C. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Berge, T. J. and Jordà, Ò. (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, 3(2):246–277.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, 28(3):381–392.

Bürgi, C. (2016). What Do We Lose When We Average Expectations? Working Papers 2016-013, The George Washington University, Department of Economics, Research Program on Forecasting.

Bürgi, C. and Sinclair, T. M. (2017). A nonparametric approach to identifying a subset of forecasters that outperforms the simple average. *Empirical Economics*, 53(1):101–115.

Döpke, J. and Fritsche, U. (2006). When do forecasters disagree? an assessment of german growth and inflation forecast dispersion. *International Journal of Forecasting*, 22(1):125–135.

Driver, C., Trapani, L., and Urga, G. (2013). On the use of cross-sectional measures of forecast uncertainty. *International Journal of Forecasting*, 29(3):367–377.

Dueker, M. (1997). Strengthening the case for the yield curve as a predictor of U.S. recessions. *Federal Reserve Bank of St. Louis Review*, 79(2):41–51.

Fintzen, D. and Stekler, H. (1999). Why did forecasters fail to predict the 1990 recession? *International Journal of Forecasting*, 15(3):309–323.

Lahiri, K. and Sheng, X. (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics*, 25(4):514–538.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *Quarterly Journal of Economics*, 117(4).

Mboup, F., Wurtzel, A., et al. (2018). Battle of the forecasts: Mean vs. median as the survey of professional forecastersâ consensus. *Research Brief*, (Q4).

Ozturk, E. and Sheng, X. S. (2016). Measuring global and country-specific unvertainty. *Working Paper*.

Patton, A. J. and Timmermann, A. (2010). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, 57(7):803–820.

Proaño, C. R. and Theobald, T. (2014). Predicting recessions with a composite real-time dynamic probit model. *International Journal of Forecasting*, 30(4):898–917.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665 – 690. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

Sinclair, T. M., Joutz, F., and Stekler, H. (2010). Can the fed predict the state of the economy? *Economics Letters*, 108(1):28 – 32.

Zarnowitz, V. (1986). The record and improvability of economic forecasting. Working Paper 2099, National Bureau of Economic Research.