A Note on Uncertainty due to Infectious Diseases and Output Growth of the United States: A Mixed-Frequency Forecasting Experiment

Afees A. Salisu*, Rangan Gupta** and Riza Demirer***

Abstract

Utilizing a mixed data sampling (MIDAS) approach, we show that a daily newspaper based index of uncertainty associated with infectious diseases can be used to predict, both in- and out-of-sample, low-frequency movements of output growth for the United States (US). The predictability of monthly industrial production growth and quarterly real Gross Domestic Product (GDP) growth during the current period of heightened economic uncertainty due to the COVID-19 pandemic is likely to be of tremendous value to policymakers.

Keywords: Infectious Diseases Related Uncertainty, Output Growth, Forecast, Mixed-Frequency
JEL Codes: C22, C53, D80, E23, E32

1. Introduction

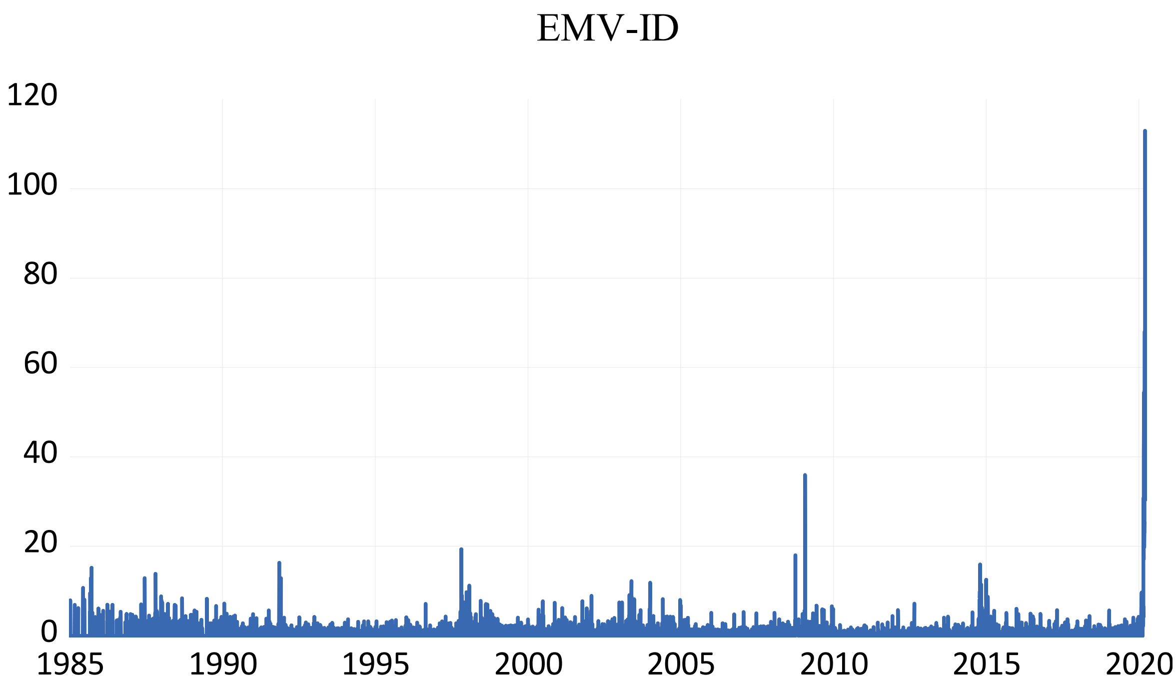
The role of various forms of uncertainty in forecasting output growth and recession has been examined in a growing number of studies, particularly following the "Great Recession" (see for example, Karnizova and Li, 2014; Balcilar et al., 2016; Junttila and Vataja, 2018; Segnon et al, 2018; Pierdzioch and Gupta, 2019). One argument to relate uncertainty to output highlights the effect of uncertainty on aggregate demand, which in turn slows down consumption and investing activity in the economy. Uncertainty can also lead to discount rate shocks, which lead to higher financing costs, hindering corporate investment, thus leading to a negative effect on output (Bernanke, 1983). Undoubtedly, the recent COVID-19 pandemic has triggered a massive spike in uncertainty associated with every aspect of human life ranging from health to livelihood, extending the impact of this health crisis to the overall economy. Given the uncertainty this health crisis created regarding economic fundamentals, the objective of our paper is to assess the role of uncertainty related to infectious diseases of various types (for example, MERS, SARS, Ebola,

* Corresponding author. Centre for Econometrics & Applied Research, Ibadan, Nigeria, and Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. Email: aa.salisu@cear.org.ng.
** Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. Email: rangan.gupta@up.ac.za.
*** Department of Economics and Finance, Southern Illinois University Edwardsville, Edwardsville, IL 62026-1102, USA. E-mail address: rdemire@siue.edu.

H5N1, H1N1, and of course Coronavirus) in predicting the path of output (industrial production) growth in the US. Clearly, this is an issue of importance for not only policy makers for fiscal and monetary policy implementation, but also corporations in their future planning decisions.

In this process, a necessary first step is to quantify the uncertainty related to infectious diseases in a way that it would act as suitable input into a statistical forecasting model for output growth. In this regard, we use the recently developed newspaper-based index of Baker et al., (2020), which tracks equity market volatility (EMV), in particular the movements in the Chicago Board Options Exchange (CBOE)'s Volatility Index (VIX), due to infectious diseases. This index, available at a daily frequency, is plotted in Figure 1 to highlight the massive increase in uncertainty over the recent months. With the uncertainty index available in daily frequency and given that industrial production is only available monthly, we use a mixed data sampling (MIDAS) approach in a forecasting setting in order to predict the future path of U.S. output growth over the period of January, 1985 to March, 2020. To the best of our knowledge, this is the first study to forecast U.S. industrial production growth using a measure of uncertainty associated with infectious diseases based on a mixed-frequency approach.

Figure 1. Daily Equity Markey Volatility (EMV) due to Infectious Diseases (ID): EMV-ID



Our findings yield significant evidence of both in- and out-of-sample predictability, suggesting that incorporating uncertainty due to infectious diseases in the predictive model can successfully improve the predictive accuracy of economic growth models compared to the historical average model benchmark. The findings are robust across the long and short horizons, suggesting that policymakers can use the proposed framework in their efforts to nowcast output growth at a high frequency, and design appropriate policy responses. The remainder of the paper is organized as follows. Section 2 lays out the data and the econometric model, while Section 3 presents the empirical results. Section 4 concludes the paper.

2. Data and Methodology

Our empirical analysis involves two variables namely, the year-on-year growth rate of the seasonally adjusted industrial production (IPG) for the U.S. and the recently developed measure of equity market uncertainty due to infectious diseases (EMV-ID). The industrial production data is obtained from the FRED database of the Federal Reserve Bank of St. Louis. Developed by Baker et al., (2020) via textual analysis, the daily infectious disease equity market volatility tracker is a newspaper article based index that measures the component of stock market uncertainty due to infectious diseases. This index is available at the daily frequency from January 1985 at http://policyuncertainty.com/infectious_EMV.html. To construct the EMV-ID, Baker et al., (2020) utilize four sets of terms namely, (i) E: economic, economy, financial; (ii) M: "stock market", equity, equities, "Standard and Poors"; (iii) V: volatility, volatile, uncertain, uncertainty, risk, risky; (iv) ID: epidemic, pandemic, virus, flu, disease, coronavirus, mers, sars, ebola, H5N1, H1N1. Tracing across approximately 3,000 U.S. newspapers, they obtain daily counts of newspaper articles that contain at least one term in each of E, M, V, and ID. First, scaling the raw EMV-ID counts by the count of all articles in the same day, the authors then multiplicatively rescale the resulting series to match the level of the VIX, by using the overall equity market volatility (EMV) index, and finally scaling the EMV-ID index to reflect the ratio of the EMV-ID articles to total EMV articles. Based on data availability of the two variables under consideration, our analysis covers the effective sample of January 1986 to March 2020.

Given the availability of the data in different frequencies, our empirical analysis adopts the autoregressive distributed lag mixed data sampling [ADL-MIDAS] model suitable for economic analyses involving a lower frequency dependent variable (monthly IPG) and a higher frequency predictor (daily EMV-ID). The MIDAS framework utilizes variables at their available frequencies, allowing us to circumvent possible information loss due to aggregation and/or dis-aggregation of variables (see Salisu and Ogbonna, 2019). The construction of the model follows the argument that higher level of uncertainty due to wide spread infections/diseases (EMV-ID) can depress real sector activity (IPG), either from a cash flow channel that affects aggregate demand or from a discount rate channel that affects investment. The model is specified with the lag structure p_{pdd}^{M} *and* q_{pdr}^{D} for the predicted and predictor variables, respectively (see equation [1a]). For robustness checks, we consider an alternative measure of output growth by using quarterly real Gross Domestic Product (GDP) growth (see equation [1b]). Likewise, we also use a newspaper-based measure of economic sentiment (ES) as an alternative to EMV-ID. The model is specified as:

$$pdd_{t+1}^{M} = \lambda + \sum_{i=0}^{p_{pdd}^{M}-1} \alpha_i pdd_{t-i}^{M} + \beta \sum_{i=0}^{q_{pdr}^{D}-1} \cdot \sum_{j=0}^{N_D-1} w_{i+j*N_D}(\phi^D) pdr_{N_D-j,t-i}^{D} + \xi_{t+1} \qquad [1a]$$

$$pdd_{t+1}^{Q} = \lambda + \sum_{i=0}^{p_{pdd}^{Q}-1} \alpha_i pdd_{t-i}^{Q} + \beta \sum_{i=0}^{q_{pdr}^{D}-1} \cdot \sum_{j=0}^{N_D-1} w_{i+j*N_D}(\phi^D) pdr_{N_D-j,t-i}^{D} + \xi_{t+1} \qquad [1b]$$

where pdd is the predicted variable (IPG or growth of real GDP (GDPG)); pdr is the predictor variable (EMV-ID or ES); N_D is the number of days in a given month or quarter, as the case may be; D, M and Q are used to respectively represent daily, monthly and quarterly data frequencies. Similarly, p_{pdd}^{M} and p_{pdd}^{Q} respectively denote the lags of the low frequency (monthly and quarterly) variables, while q_{pdr}^{D} denotes the lags for the high frequency predictor.[1] Finally, w and ϕ represent the weighting function of the polynomial vector and the normalized exponential Almon lag polynomial parameters (He and Lin, 2018), respectively, while the stochastic disturbance term that is independently and identically distributed is denoted by ξ_t. In the equations, λ is the constant term, while α and β are the coefficients of the lagged predicted and the predictor variables, respectively, with the latter indicating the stance of predictability or otherwise. Consequently, the

[1] Note that equation (1a) incorporates monthly and daily, while equation (1b) involves quarterly and daily data frequencies.

null of non-predictability, that is $H_0 : \beta = 0$, is tested, with the rejection of the null implying that the news index has predictability characteristics for output growth.

For the purpose of forecast evaluation of the ADL-MIDAS growth model, we also consider an alternative (usually the baseline) predictive model for output growth using the historical average model, which is also the restricted model for equations [1a] and [1b]. Forecast measures such as the root mean square error [RMSE] and the Clark and West (2007) [C-W] statistics are employed. The latter is particularly important as it helps to determine the statistical significance of the difference between the forecast errors of the two nested (restricted and unrestricted) models. We conduct only the out-of-sample forecast given that we had earlier shown predictability in the in-sample period. Three out-of-sample forecast horizons - 6, 12 and 24 months, are considered for robustness, based on a recursive estimation on a 50% in- and out-of-sample split of the data.

3. Empirical Findings

We begin the analysis by examining the in-sample predictive power of the daily EMV-ID index for monthly output growth using the predictive ADL-MIDAS model that incorporates a mixed data sampling in the conventional autoregressive distributed lag model framework. Table 1 presents the estimated coefficients for the Almon polynomial weights as well as the slope coefficient of the predictor variable under the monthly frequency. We find that, while the Almon polynomial weights are statistically significant, the estimated slope coefficient is negative. This indicates that the overall effect of daily uncertainty due to infectious diseases on monthly industrial production growth is indeed negative and linearly increasing in the lags. This inference is found to be valid given that both Almon polynomial coefficients reported in the table are statistically significant, since we require at least one of the Almon polynomial coefficients to be statistically significant for the impact of the predictor variable to be valid. We also observe that about 10 daily EMV-ID observations are required to adequately predict output growth on a monthly basis.

Table 1. In-Sample Predictability Results

Predicted	Predictor	Almon Polynomial Weights		Slope Coefficient Estimate
		PDL01	PDL02	
IPG	EMV-ID	-0.0521** [0.0204]	0.0067** [0.0033]	-0.0454 {-10}

Note: This table presents findings for the in-sample predictive power of the daily EMV-ID index for monthly output growth using the predictive ADL-MIDAS model. Figures in each cell are the estimated coefficients and corresponding standard error in square brackets. PDL01 and PDL02 are respectively the first and second Almon polynomial distributed lag (PDL) weights. The statistical significance of the estimated coefficients at 1%, 5% and 10% levels are indicated by ***, ** and *, respectively. Under the column labelled "Slope Coefficient Estimate", the figures in brace bracket "{}" indicate the number of the high frequency EMV-ID proxy lags employed in predicting the low output growth proxy.

Table 2. Forecast Evaluation

Model	Predicted	Predictor	RMSE			CLARK AND WEST		
			$h=6$	$h=12$	$h=24$	$h=6$	$h=12$	$h=24$
Panel A: Monthly Frequency								
ADL-MIDAS	IPG	EMV-ID	0.0372	0.0369	0.0364	1.54E-04*** [1.49E-05]	1.52E-04*** [1.47E-05]	1.48E-04*** [1.44E-05]
		ES	0.0394	0.0391	0.0386	7.12E-04*** [1.05E-04]	7.03E-04*** [1.04E-04]	6.85E-04*** [1.01E-04]
Historical Average			0.0411	0.0408	0.0403	-	-	-
Panel B: Quarterly Frequency								
ADL-MIDAS	GDPG	EMV-ID	0.0308	0.0315	0.0312	8.96E-04*** [1.77E-04]	8.83E-04*** [1.74E-04]	8.57E-04*** [1.69E-04]
		ES	0.0306	0.0313	0.0310	9.61E-04*** [1.32E-04]	9.48E-04*** [1.31E-04]	9.22E-04*** [1.28E-04]
Historical Average			0.0372	0.0379	0.0377	-	-	-

Note: This table presents the out-of-sample forecast performance of the ADL-MIDAS model in comparison with the benchmark model – the historical average. Panel A reports the findings for equation (1a) that incorporates monthly and daily data, while Panel B reports the estimates for equation (1b) that involves quarterly and daily data data. Under the column labelled Clark and West, the figures in each cell are the estimated constants and corresponding standard error in square brackets. The statistical significance of the estimated coefficients at 1%, 5% and 10% levels are indicated by ***, ** and *, respectively.

In addition to the predictability analysis, we further evaluate the forecast performance of our predictive model in comparison with the benchmark model – the historical average. The results for equation [1a] in Panel A show that the ADL-MIDAS model yields lower RMSE values, consistently across all the forecast horizons considered, compared to the benchmark model. This stance is upheld by the C-W test as the estimated constant is found to be positive and statistically significant across the three out-of-sample forecast horizons. Combined with the evidence of in-sample predictability, the consistent out-of-sample outperformance of our predictive model shows that incorporating EMV-ID in the predictive model for industrial output growth can successfully improve the predictive accuracy of EMV-ID-based model compared to the benchmark model. Furthermore, the consistency of out-performance across the long and short horizons indicates that the results are insensitive to the forecast horizon considered. This is indeed an important consideration given the variety in the forecasting horizons employed by policy makers, corporations and investors.

As additional robustness checks, we first redo the analysis using a daily newspaper based index of economic sentiment (ES) of the US, developed by Shapiro et al., (2020).[2] Buckman et al. (2020) show that this index of economic sentiment can accurately track the recent downturn of market sentiment due to the outbreak of the COVID-19 virus. The ES index is a high frequency measure of economic sentiment based on lexical analysis of economics-related news articles derived from 16 major U.S. newspapers. The decision to use the ES as a matter of comparison to the EMV-ID is to accommodate for the fact that uncertainty (related to infectious diseases and other economic and financial events) tends to impact output via economic sentiment that drives aggregate demand and investment decisions, as noted earlier. This index starts in January 1980 and hence our forecasting analysis for industrial production growth now starts from this period until March 2020. Moreover, using EMV-ID and ES, we also forecast the year-on-year growth rate of real GDP (GDPG) for the U.S. over the quarterly period of 1986:Q1 (or 1981:Q1) to 2020:Q1, using the data obtained from the FRED database. As shown in Panel B, ES is found to also accurately predict IPG relative to the historical average in a statistically significant manner, with the same also holding for GDPG under both EMV-ID and ES. Accordingly, the robustness checks confirm the predictive role of uncertainty or sentiment proxies due to infectious diseases over future economic growth forecasts.

4. Conclusion

This paper utilizes the MIDAS approach to show that a daily newspaper-based metric of uncertainty associated with infectious diseases can be used to accurately forecast the future path of low frequency measures of output growth, such as monthly industrial production growth and

[2] The data is available publicly from: https://www.frbsf.org/economic-research/indicators-data/daily-news-sentiment-index/.

quarterly real GDP growth. Given the unprecedented economic uncertainty experienced by all dimensions of the economy due to the current COVID-19 pandemic, our results should be of tremendous value to policymakers in their efforts to nowcast output growth at a high frequency, and design appropriate policy responses to reduce the negative impact of the coronavirus, and all possible such outbreaks in the future.

References

Baker, S.R., Bloom, N., Davis, S.J., Terry, S.J. (2020). Covid-Induced Economic Uncertainty. NBER Working Paper No. 26983.

Balcilar, M., Gupta, R., and Segnon, M. (2016). The Role of economic policy uncertainty in predicting U.S. recessions: A mixed-frequency Markov-switching vector autoregressive approach. Economics: The Open-Access, Open-Assessment E-Journal, 10, 1-20.

Bernanke, Ben S. (1983). Irreversibility, Uncertainty, and Cyclical Investment. The Quarterly Journal of Economics, 98(1), 85-106.

Buckman, S.R., Shapiro, A.H., Sudhof, M., and Wilson, D.J. (2020). News Sentiment in the Time of COVID-19. FRBSF Economic Letter, 2020-08 (April 6).

Clark, T.E., and West, K.D., (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138, 291–311.

He, Y. and Lin, B. (2018). Forecasting China's total energy demand and its structure using ADL-MIDAS model. Energy 151, 420-429.

Jung, A. (2017). Forecasting broad money velocity. North American Journal of Economics and Finance, 42, 421-32.

Junttila, J., and Vataja, J. (2018). Economic policy uncertainty effects for forecasting future real economic activity. Economic Systems, 42(4), 569-583.

Karnizova, L. and Li, J.C. (2014). Economic policy uncertainty, financial markets and probability of US recessions. Economics Letters, 125, 261-265.

Pierdzioch, C., and Gupta, R. (2019). Uncertainty and Forecasts of U.S. Recessions. Studies in Nonlinear Dynamics & Econometrics. DOI: https://doi.org/10.1515/snde-2018-0083.

Salisu, A.A. and Ogbonna, A.E. (2019). Another look at the energy-growth nexus: New insights from MIDAS regressions, Energy, 174, 69-84.

Shapiro, A.H., Sudhof, M., and Wilson, D.J. (2020). Measuring News Sentiment. Federal Reserve Bank of San Francisco Working Paper No. 2017-01.

Segnon, M., Gupta, R., Bekiros, S., and Wohar, M.E. (2018). Forecasting US GNP growth: The role of uncertainty. Journal of Forecasting, 37(5), 541-559.