# Strategy for a low-cost uncertainty index: A method for sparse data environments[†]

Steven P. Cassou[1] and Jesús Vázquez[2]

November 2026

**ABSTRACT:** This paper describes a low-cost strategy for generating an uncertainty index. The method is useful in sparse data settings. Several investigations are undertaken which illustrate its value. First, it is shown to work well in rich data settings. Using U.S. data from 1960 onward, it is shown that a variety of simple uncertainty measures can be generated which exhibit properties, such as high peak uncertainty, high correlation and goodness of fit, that aligns well with indices that require significantly higher data requirements, such as the approach proposed by Jurado et al. (2015). Second, an illustration using sparse historical U.S. data shows that this approach can generate sound uncertainty measures when rich data sources are unavailable. Historical uncertainty measures offer valuable insights into the evolution of the business cycle. They also enable comparisons of uncertainty during temporally dispersed economic events.

JEL classification: C32, D80, E32

Keywords: Uncertainty index, Sparse data environment, Markov regime-switching VAR, Conditional expectations with GARCH errors.

---

*1* Steven P. Cassou: scassou@ksu.edu. Affiliation: Department of Economics, 327 Waters Hall, Kansas State University, Manhattan, KS, 66506, U.S.A. ORCID: 0000-0002-3724-7571.

*2* Jesús Vázquez: jesus.vazquez@ehu.eus. Affiliation: University of the Basque Country (EHU), Departamento de Análisis Económico, Bilbao, Spain. ORCID: 0000-0003-4788-4838.

†We thank Luis Herrera and Carlos Maldonado for useful discussions on this topic. We would also like to thank Daniel Levy, Xavier Raurich, Juan Rubio-Ramírez and the participants of the 2025 Society for Economic Measurement (SEM) Conference for their comments on a preliminary version of this paper. This research was supported by the grant PID2024-158011NB-I00 funded by MICIU/AEI/10.13039/501100011033, and by "ERDF/EU", and the grant IT1879-26 funded by the Basque Government.

# 1 Introduction

Although economic uncertainty is not observable, it appears to be a critical factor in economic performance and the business cycle. Over the past decade, several seminal articles (e.g., Jurado et al. 2015; Rossi and Sekhposyan 2015; Baker et al. 2016; Scotti 2016) have proposed methods of proxy measurement for economic uncertainty. These methods work well when there is lots of data or specialized data available for the computations. For instance, the Jurado et al. (2015) method uses over three hundred data series to compute an uncertainty index from 1960 to the present while the Baker et al. (2016) method requires comprehensive searches in various news sources to compute an uncertainty index from 1985 to the present. On the other hand, the Rossi and Sekhposyan (2015) method uses specialized data from professional forecasters and the Scotti (2016) method uses specialized real-time data. These rich data or specialized data environments mean that these methods cannot be applied in other settings featuring sparse data. In this paper, we describe a simple method for generating an uncertainty index, show that it works well in rich data environments, and then show how to use it in a *historical* setting in which data is sparse.

We propose a simple algorithm that can be used in sparse data environments, but begin by showing that the algorithm is able to reproduce desirable properties of rich-data uncertainty measures in rich data settings. This demonstration is intended to answer the question of whether the algorithm can only perform well in sparse data settings. We show this is not the case, and that the algorithm does well even in rich data environments. For this demonstration we generate several simple uncertainty indices using data from 1960, which is the period over which the Jurado, Ludvigson and Ng (JLN) index is available. The various simple indices studied were guided by the set of variables seen in simple New-Keynesian models, but the approach can be readily applied to other sets of variables. In addition to showing the ability to reproduce many of the attributes of rich-data uncertainty measures, it is shown that adding variables found in medium-scale New-Keynesian models does not improve performance, further demonstrating the need for only a few variables to make this algorithm work. Once we have established that the algorithm can produce indices that posses many of the properties of data-rich uncertainty measures over periods in which

data is essentially unlimited, we investigate historical periods as far back as the late 18th century in which data is much more sparse, and show that this approach is able to generate good uncertainty measures in these circumstances as well, thus demonstrating an ability unattainable for rich-data approaches.

These simple uncertainty indices are mostly computed using just two economic variables that primarily characterize the economic outlook: Gross domestic product (GDP) and inflation.[1] The final indices are found by combining several individual volatility indices. One group of volatility measures included in all of our indices are simple analogues of the measures used to compute the JLN index. Here the conditional forecast error standard deviations using univariate Autoregressive Integrated Moving Average (ARIMA) forecasting models with Generalized Autoregressive Conditional Heteroscedasticity (GARCH) errors are computed. These simple conditional forecast error standard deviations are then combined with one of several alternative measures that provide information about the state of the economy. One of these other indices is the National Bureau of Economic Research (NBER) index for the state of the economy. Another is a constructed probability index of the economy being in a regime featuring high volatility and, in most cases, severe recession. This index uses a two-state Markov switching vector autoregression (MS-VAR). A third index is the rolling observed standard deviation of GDP growth. One attraction of including a state of the economy variable in an uncertainty index is that it captures important asymmetries in the economy. Although using conditional forecast error standard deviations (as employed in the JLN and the parts of the simple indices proposed herein) are a valid way of quantifying uncertainty, they are symmetric and do not address the asymmetry in perceived uncertainty. For example, positive income surprises are generally considered much less uncertain than negative ones, which are likely to necessitate further adjustments to cope with uncertainty. In other words, the economy is more uncertain when it is performing poorly.[2] By combining forecast error uncertainty indices

[1] One exception is an index that also uses the National Bureau of Economic Research recession indicator series.

[2] Building on this insight, Rossi and Sekhposyan (2015) introduced a new way to measure macroeconomic uncertainty, based on forecast error distributions rather than forecast error variances. As their index is based on distribution position rather than squared error, it can be split into an "upside" uncertainty index (unexpectedly good news) and a "downside" uncertainty index (unexpectedly bad news), which a variance-based measure cannot distinguish. They apply the method to real GDP growth forecasts from the Survey of Professional Forecasters

with measures of the state of the economy we are able to introduce some degree of asymmetry into our indices.

Like the JLN uncertainty index, the simple uncertainty indices proposed here are based on ex-post, full-sample estimation. This contrasts with recent studies (e.g., Rossi and Sekhposyan 2015; Scotti 2016; Jo and Sekkel 2019; and Xu and Rogers 2025) that use real-time macroeconomic or professional forecaster data. While real-time indexes are ideal for capturing policy reactions, indices based on historical data are essential for understanding the role of uncertainty in long-term economic cycles. Thus, historical data enables researchers to define a benchmark uncertainty measure across different decades, centuries, and economic regimes. Moreover, historical data helps contextualize rare shocks. Without historical indices, it is impossible to assess the uncertainty surrounding modern events, such as the 2008 financial crisis or the Covid-19 pandemic, relative to historical events of a similar magnitude, such as the Great Depression.[3]

We conducted several empirical analyses to investigate how well our simple uncertainty indices reproduce the behavior of other well-known uncertainty indices during rich data periods. We start with a simple plot and show our simple measures exhibit high peak uncertainty during periods similar to those indicated by the JLN index. Additionally, we show our indices are highly correlated with the JLN index and regressions show our indices fit the JLN measure well. As further evidence for the value of simple indices, we also investigate whether additional variables, such as the policy interest rate or other variables seen in medium-scale New Keynesian models, could improve the performance, and found that they did not.

After demonstrating an ability of our indices to reproduce behavior of the JLN in a rich data environment, we then use our low-cost strategy to calculate measures of U.S. uncertainty dating back to the late 18th century. Very little data is available from this period, but Williamson (2026)

(SPF), using both nowcasts and four-quarter-ahead forecasts.

[3]Furthermore, unlike real-time data, historical data smooths out nowcasting noise, capturing true structural and macroeconomic uncertainty rather than temporary sentiment swings. Xu and Rogers (2025) have shown that recalculating the JLN index using strictly real-time recursive/rolling windows yields quantitatively different paths (e.g., lower peaks during certain historical crises). Moreover, Jo and Sekkel (2019) have shown that choosing different data releases (i.e., vintages) impacts the estimates and the size of the spikes in the estimated uncertainty series.

has constructed annual GDP and GDP-deflator data since then, which can be used to calculate two of the three simple indices. The one based on the NBER recession indicator is not able to go that far back, but can be calculated back to 1871. Our analysis shows a clear trend of less frequent recessions and uncertainty spikes since the period after World War II, which becomes clearer when viewed from a long-term historical perspective using our simple indices. In addition, the historical uncertainty indices allow one to compare uncertainty during significant economic events that occur during very different periods of time.

The rest of the paper is organized as follows. Section 2 describes the strategy for generating a low-cost uncertainty index. Section 3 describes a simple Markov-switching measure of uncertainty which is used in one of our simple low-cost indices. In Section 4, three low-cost indices using just two (three for the index using the NBER recession indicator series) macroeconomic variables is presented, and an empirical comparison of these indices with other widely used uncertainty indices including the JLN index is undertaken. Section 5 presents a robustness analysis of the low-cost strategy across several dimensions. One demonstration in the robustness section is to show that additional variables, such as the policy rate, or other medium-scale New Keynesian variables are unnecessary. Section 6 studies historical data and demonstrates the ability of this simple approach to generate uncertainty indices that are unattainable for rich-data approaches. Section 7 concludes.

## 2 A low-cost strategy

As noted in Jurado et al. (2015), uncertainty is the conditional volatility of the purely unforeseeable component of the future value of a series. As part of their approach, they start by computing the uncertainty for a particular variable, $y_{jt}$, by the expression

$$U_{jt}^{y}(h) = \sqrt{E\left[\left(y_{jt+h} - E\left[y_{jt+h} \mid I_t\right]\right)^2 \mid I_t\right]},$$

where $h$ is a horizon for which one is interested in measuring uncertainty and $I_t$ is the information set at date $t$. Later, this variable uncertainty is aggregated across many variables to give a general

uncertainty measure which they denote by

$$U_t^y(h) = plim_{N_y \to \infty} \sum_{j=1}^{N_y} w_j U_{jt}^y(h) \equiv E_w\left[U_{jt}^y(h)\right].$$

There are two things that make the calculation of this uncertainty measure computationally burdensome. First, they allow the information set $I_t$ to be as broad as possible. As they put it, they use factors from a large set of predictors $\{X_{it}\}, i = 1, 2, ..., N$ whose span is as close to $I_t$ as possible. And second, they use $N_y$ different variable uncertainty values, and then let $N_y \to \infty$, which in practical terms means becomes very large.

We follow a similar approach for part of our indices, but suggest several simplifications which makes the calculation computationally less burdensome. Here we propose an uncertainty measure which consists of a weighted average of two types of uncertainty series. The first type are conditional forecast error standard deviations based on information sets consisting of only a variable's own history. Unlike Jurado et al. (2015), using only a variable's own history enormously simplifies the computation of the conditional forecast error standard deviations because it allows one to use ARIMA models with GARCH errors. ARIMA models have been shown to perform well for forecasting a variable, and methods for undertaking such computations are readily available in just about every statistical package. Using a GARCH model for the error process then allows one to compute the conditional forecast error standard deviation of the variable in a way that is again available in standard packages. The second type of series is a measure of the state of the economy. These state of the economy series also focus on simplicity. One state of the economy is simply the NBER recession indicator series, while another is the estimated probability of being in a high-volatility regime and a third computes the volatility of (the annual growth rate of) GDP. The most complicated of these state of the economy series is a two-state Markov regime-switching vector autoregressive (MS-VAR) model. Computation of a two-state MS-VAR is more involved than simple ARIMA models with GARCH errors, but not nearly as burdensome as the methods used by Jurado et al. (2015). We describe the MS-VAR model more thoroughly in the next section.

A further computational simplification can be obtained by keeping the number of variables used

in these computations small. In Section 4, we construct several simple uncertainty indices using mostly just two data series: real GDP and inflation. We show these simple uncertainty indices exhibit similar properties as the JLN index. The two variables used were in part guided by the early New Keynesian or Dynamic Stochastic General Equilibrium (DSGE) models which also include a policy interest rate. However, after some study summarized in Section 5, we eliminated the policy interest rate series since it did not add to the performance. Section 5 also considers a somewhat larger set of other variables which add to the simple uncertainty indices data for consumption, investment, hours worked, and the real wage. These other variables where investigated because they are the observables used to estimate the next generation of New Keynesian or DSGE models, as in Smets and Wouters (2007).[4] This investigation shows that these additional variables also do not add to the performance of the simple indices, further reinforcing our finding that simple indices can work well.

# 3 Measuring uncertainty with a simple MS-VAR

Here we describe a second-order MS-VAR model that includes a linear-quadratic detrended GDP and the inflation rate associated with the implicit GDP deflator. Formally, this VAR can be written as follows:

$$Y_t = \mu(s_t) + A_1(s_t) Y_{t-1} + A_2(s_t) Y_{t-2} + [\Omega(s_t)]^{\frac{1}{2}} \varepsilon_t, \tag{1}$$

where $Y_t = \left[\pi_t, G\tilde{D}P_t\right]'$, $\varepsilon_t \sim N(0, I)$, $\pi_t = 400 \times ln(P_t/P_{t-1})$ denotes the annualized inflation rate based on the GDP deflator $P_t$, and $G\tilde{D}P_t$ is the stationary component of GDP found by removing a linear-quadratic trend.[5] The state variable $s_t$ takes values 1 or 2 and follows a first-order two-state Markov process with probabilities of remaining in a particular state of $prob(s_t = 1 \mid s_{t-1} = 1) = p_{11}$ and $prob(s_t = 2 \mid s_{t-1} = 2) = p_{22}$ and transition probabilities be-

[4] As discussed below, we also explored incorporating financial variables, like the S&P 500 Volatility Index (VIX) and credit spreads.

[5] This stationary component of GDP is obtained as the residual from an ordinary least squares regression of the log of GDP on a constant, linear, and quadratic time trend.

tween the two states of $prob\left(s_t = 2 \mid s_{t-1} = 1\right) = 1 - p_{11}$ and $prob\left(s_t = 1 \mid s_{t-1} = 2\right) = 1 - p_{22}$. We estimate the Cholesky decomposition $\Sigma\left(s_t\right)$ of $\Omega\left(s_t\right)$ by $\Omega\left(s_t\right) = \Sigma\left(s_t\right)\left[\Sigma\left(s_t\right)\right]'$. Part 1 of the supplementary appendix provides a brief description of the maximum-likelihood estimation procedure used to estimate this MS-VAR system. Full estimation results are shown in Part 2 of the supplementary appendix.

Three conclusions emerge from a comparison of the parameter estimates across states. First, the conditional covariances are significantly larger in the second state than in the first, with moduli for the eigenvalues of $\Omega\left(1\right)$ and $\Omega\left(2\right)$ given by (0.5378, 0.2560) and (0.8089, 5.1750), respectively. Based on this finding, we can refer to state 1 as the low volatility state and state 2 as the highly volatility state. Second, state 1 is more persistent than state 2 because the estimated value of $p_{11}$ is significantly larger than $p_{22}$. This fits one's intuition that low volatility states, often associated with periods of economic stability, last longer than high volatility states, such as economic recessions. Finally, the persistence of the system in each state can be measured through the moduli of the *highest* eigenvalue for the companion matrix when writing the VAR(2) in equation (1) as a VAR(1). Formally, the companion matrix is given by

$$F(s_t) = \begin{pmatrix} A_1(s_t) & A_2(s_t) \\ I_{2\times 2} & 0_{2\times 2} \end{pmatrix}.$$

The highest eigenvalues of the companion matrices $F(1)$ and $F(2)$ are 0.9522 and 0.9099, respectively, showing that the time series are somewhat more persistent in the first state than in state 2.

Figure 1 shows the allocation of time periods based on the smoothed probabilities associated with the high volatility state (state 2) as well as the National Bureau of Economic Research (NBER) recession periods indicated with gray colored regions. This figure shows that the 1970s and the first-half of the 1980s are mostly allocated to state 2, featuring high volatility around the recurrent oil crises. However, the figure also shows that starting around the mid-1980s, most periods are assigned to the low volatility state (state 1). During this post mid-1980s period, there

are a few jumps from state 1 to state 2 during recession periods, with jumps back to state 1 after the recessions end. This post mid-1980s period shows that the smoothed probability assignment almost exclusively links the highly volatile state with severe recessions, including the Great Recession and the recession caused by the Covid-19 pandemic, but it is not linked with milder recessions of the early 1990's and early 2000's. However, one exception to this link between severe recessions and high state 2 probabilities is the last few years of the sample which show high probabilities of state 2 despite no recession. This period was characterized by a rapid economic recovery from the Covid-19 downturn, an outbreak of inflation driven by supply chain shortages and excessive stimulus put in place to combat the Covid-19 downturn, as well as a significant escalation of the war in Ukraine.

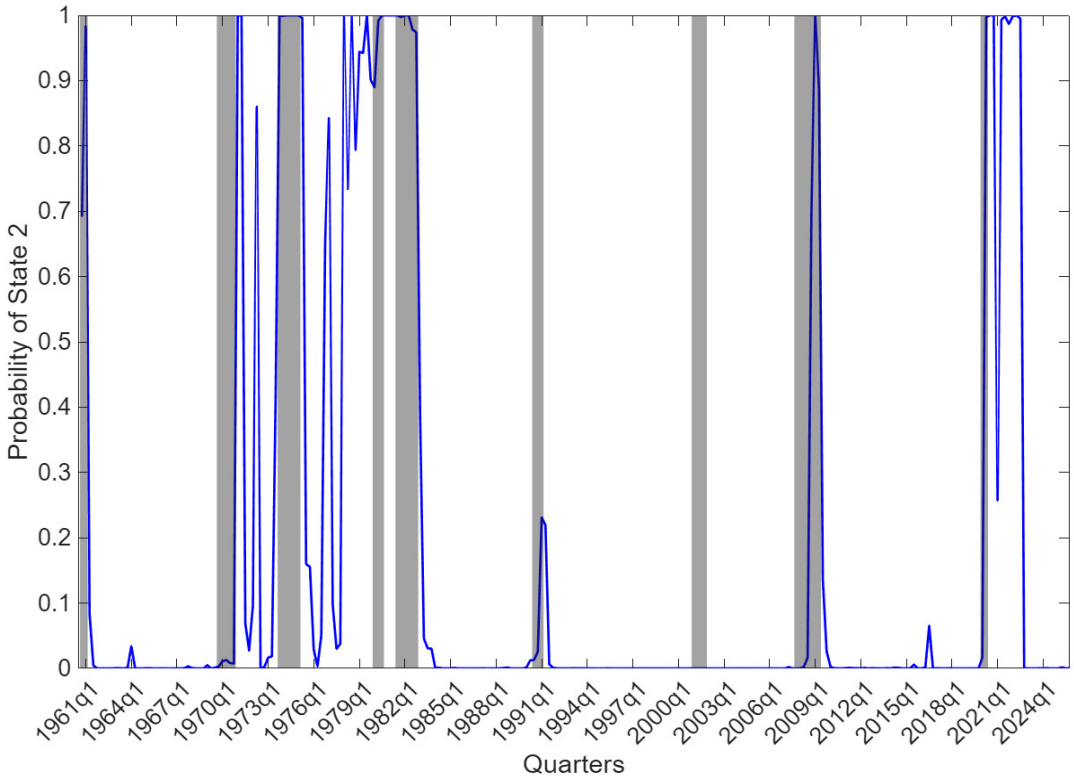



Figure 1. Probability of the high volatility state.

Since high uncertainty episodes are associated with severe recessionary periods and high macroeconomic volatility, the *probability* of being in a state characterized by these two features makes a

good candidate for a measure of uncertainty as well as a good component in the construction of a simple uncertainty index.

# 4 A comparison of several simple uncertainty indices with known indices

In this section we evaluate how several simple uncertainty indices compare to a few well known indices over recent data periods. This comparison is undertaken to show the simple indices have good performance characteristics during a period in which there is a lot known about the economy, i.e., a period featuring a data-rich environment in the U.S. The first subsection focuses on comparisons with the JLN index in several basic ways, while the second expands these basic comparisons to the Economic Policy Uncertainty (EPU) index of Baker et al. (2016).

## 4.1 A comparison to the JLN index

In this section, we construct three simple uncertainty indices and compare them to the JLN index. We refer to them as simple uncertainty indices because an important feature of their construction is simplicity. These simple indices combine a measure of the state of the economy, with the conditional standard deviations of one-quarter-ahead forecast errors for two variables: GDP and inflation.[6] What distinguishes the three uncertainty indices is the measure for the state of the economy, with each using a different approach for generating the measure. The three measures for the state of the economy are all computationally simple. The first measure is the NBER recession indicator series. The second is a two-state MS-VAR, as described in Section 3, and the third is the rolling standard deviation of GDP growth computed from the four most recent quarterly GDP growth values. In addition, the conditional standard deviations of the forecast

[6]The choice of GDP and inflation was guided by the early small-scale New Keynesian models which had just these two variables along with a policy rate variable. Early investigations, summarized in Section 5, also used the policy interest rate, but found it to not improve on a model with just GDP and inflation alone, so the policy rate is not used in our baseline indices. Furthermore, for illustrative purposes, we only considered one-quarter-ahead forecast errors here, but our algorithm can be extended to consider longer forecasting horizons.

errors for each variable are computationally simple, using only own data in ARIMA forecasting models with GARCH(1,1) errors. The structure of the various ARIMA models are based on the Bayesian Information Criterion (BIC).[7] So, rather than using a large number (around three hundred) of variables with complicated computations for generating conditional forecast error standard deviations, as in Jurado et al. (2015), two of these simple uncertainty indices use only two variables, while the third index adds the NBER recession series, thus using a third variable. Finally, the three component indices of each index are aggregated to form each simple uncertainty index with a focus on simplicity. The three component series are first standardized to have a mean of zero and a variance of one. Then, the final index gives equal weight to each part. For comparability, the JLN index is also standardized to have a mean of zero and a variance of one over the sample period considered.

We start by looking at a plot of the various series. Figure 2 shows a plot of the JLN measure (solid black line), a simple index using the NBER recession index for the state variable (green doted line), a simple index using a two variable MS-VAR recession probably index (blue long dashed line) and a simple index using the GDP growth volatility index (red short dashed line).[8] One of the binding features that impacted the range of the figure, as well as some of the statistical analysis in various tables below, is that the JLN index range begins in the early 1960's. The figure shows that our simple uncertainty indices closely tracks the JLN measure, and, most importantly, captures the high uncertainty spikes occurring around recessions.

[7] We considered six alternative forecasting models for each variable. For variables showing a trend, we considered the following ARIMA processes: (0, 1, 0); (1, 1, 0); (0, 1, 1); (2, 1, 0); (2, 1, 1); and (2, 1, 2). For variables without a trend, we considered the following ARIMAs: (0, 0, 0); (1, 0, 0); (0, 0, 1); (2, 0, 0); (2, 0, 1); and (2, 0, 2). Following the BIC, the model used to forecast GDP was an ARIMA(0,1,0) and to forecast inflation was an ARIMA(2,0,0).

[8] For this plot and the other statistical analysis below we use the real JLN index with a forecasting horizon of one quarter ahead.

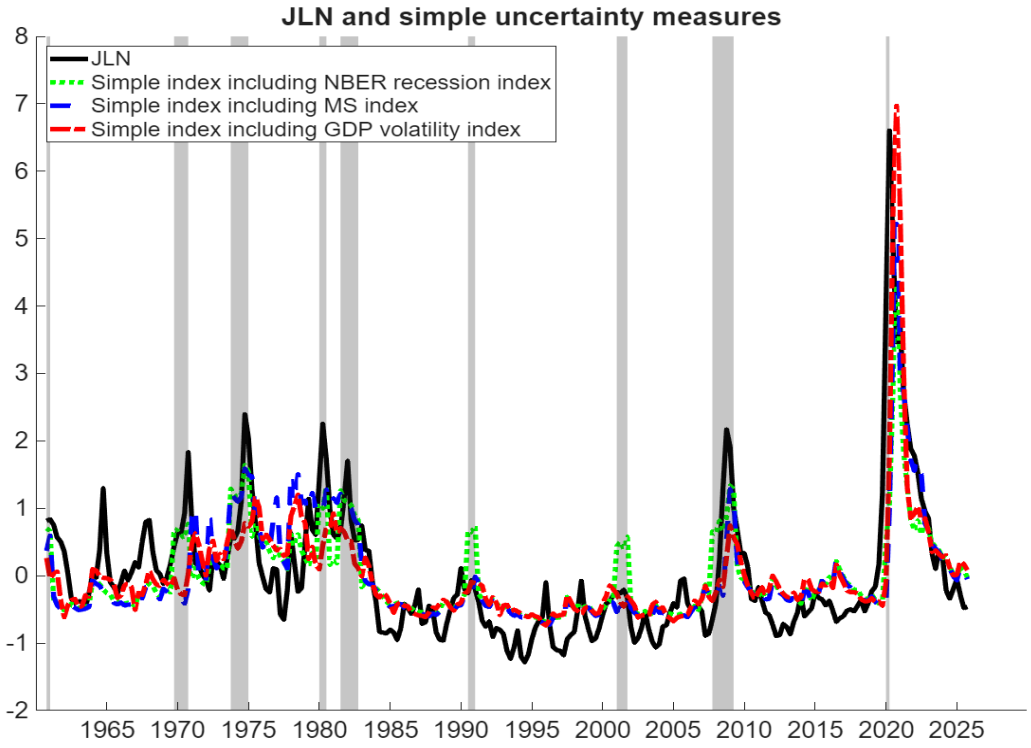



Figure 2. JLN and simple uncertainty measures. Sample period 1960:4-2025:4.

Next, we undertook some formal statistical comparisons between these simple indices and the JLN index. These results are shown in Table 1. Row 1 provides information about the period in which the statistics were tabulated and various features of each simple index that distinguish it from the others. Rows 2 and 3 show the constant term coefficient and its standard error. Here, all the constant terms have an estimated coefficient of zero because the JLN index and the various component series of each simple index were normalized to have a mean of zero. Rows 4, 6 and 8 show coefficients for the state variables which are unique to each index, while rows 5, 7 and 9 show the 5%-95% confidence intervals for the state variable coefficients. Rows 10 and 12 show the coefficients for GDP-UM and Inflation-UM while 11 and 13 show the 5%-95% confidence intervals for these variables. We denote the individual series with an additional UM notation for uncertainty

measure in order to distinguish them from their fundamental data series. These two variables are in common to each model, so all three model columns have entries, unlike the state variable rows. Rows 4, 6, 8, 10 and 12 show two terms in each cell, with the first showing the coefficients obtained from the regression with a constant and the three individual uncertainty measures entering in each specification, and the second showing the Shapley R-squared contributions for each regressor in italics.[9] Row 14 shows the *adjusted* R-squared coefficient of this regression, while Row 15, called 'Adjusted R-squared for simple OLS', shows the adjusted R-squared of a regression of JLN on a constant and each simple index alone. Row 16 shows the correlation coefficient between the JLN and the simple uncertainty indices. Row 17 shows the standard deviation of a twenty-year rolling-window correlation between the JLN index and the various simple indices along with Row 18 which shows the chi-square 95% confidence interval for the rolling correlation.

A few conclusions emerge from this table. First, with the exception of the GDP-UM coefficient in the GDP realized volatility index, all three variables other than the constant are significantly different from zero in each simple uncertainty index regression.[10] Second, each of the three measures for the state of the economy which enters in one of the simple indices show a relatively high Shapley R-squared contribution, showing that each of these individual measures play an important role in the corresponding simple index in fitting the JLN index. Third, each of the three uncertainty indices has a correlation with the JLN index near 0.7, which is similar to the 0.76 value found by Carriero et al. (2018) using a much more complicated model than any of these simple indices.[11] Fourth, as indicated by the rolling standard deviation correlations, there is some variability in the correlation between the simple indices and the JLN index, with the NBER recession dummy index showing the most stable correlation.[12] Overall, Table 1 shows that the three simple uncertainty

[9] Note that by construction the sum of the Shapley R-squared contributions for each regressor equal the *unadjusted* R-squared of the regression.

[10] The reason that the GDP-UM coefficient in the GDP realized volatility index is insignificant is that the GDP-UM series is very similar to the state variable, GDP growth volatility, so the two variables provide similar information.

[11] The value of 0.76 should be viewed as a broad reference, as Carriero et al. (2018) considered a slightly shorter sample period (1959–2014) and different econometric approaches.

[12] Rolling correlation plots included in Figure A.1 of Part 3 of the supplementary appendix also support this result. These plots show that the relatively low correlations occur during the Great Moderation period of the 1990s, which was also a period of time that Carriero et al. (2018) found differences with the JLN index.

indices perform well despite their simplicity.

Table 1. Regression results of JLN on the simple uncertainty indices

| | Simple index based on | | |
|---|---|---|---|
| Period: 1960:4-2025:4 | MS-UM | Recession dummy | GDP realized volatility |
| Constant | 0.0000 | 0.0000 | -0.0000 |
| | (-0.0833, 0.0833) | (-0.0795, 0.0795) | (-0.0816, 0.0816) |
| MS-UM | 0.4028/*0.2269* | | |
| | (0.2282, 0.5774) | | |
| Recession dummy | | 0.3804/*0.1592* | |
| | | (0.2353, 0.5254) | |
| GDP realized volatility | | | 0.6637/*0.2739* |
| | | | (0.2222, 1.1052) |
| GDP-UM | 0.3664/*0.1914* | 0.4396/*0.2379* | -0.1012/*0.1319* |
| | (0.1774 0.5553) | (0.2318, 0.6475) | (-0.4521, 0.2497) |
| Inflation-UM | 0.1683/*0.1204* | 0.3475/*0.1828* | 0.3094/*0.1510* |
| | (0.0158, 0.3209) | (0.2561, 0.4389) | (0.1786, 0.4403) |
| Adjusted R-squared | 0.5315 | 0.5733 | 0.5498 |
| Adjusted R-squared for simple OLS | 0.5196 | 0.5734 | 0.4499 |
| Correlation with JLN | 0.7234 | 0.7594 | 0.6739 |
| Std. deviation of rolling correlation | 0.1601 | 0.1086 | 0.1862 |
| | (0.1452, 0.1785) | (0.0985, 0.1211) | (0.1688, 0.2076) |

Notes: The last two rows show the standard deviation and the 95% confidence interval of the twenty-year rolling-window correlation between the JLN index and the corresponding index.

## 4.2 Further comparisons with the EPU index

This section adds additional comparisons between the three simple indices, the JLN index, and the EPU index.[13] Figure 3 adds the EPU index to the series shown in Figure 2, while Table 2 shows the statistical results. One noteworthy difference in both Figure 3 and Table 2 is the time period, which is 1985:1-2025:4. The starting date is different, simply because the EPU index starts in the first quarter of 1985. Looking at Figure 3, one can see that the EPU index is generally more volatile than the other four indices and it becomes quite different from them beginning in 2025. Given this different behavior, we do not expect the various indices to statistically match the EPU index very well.

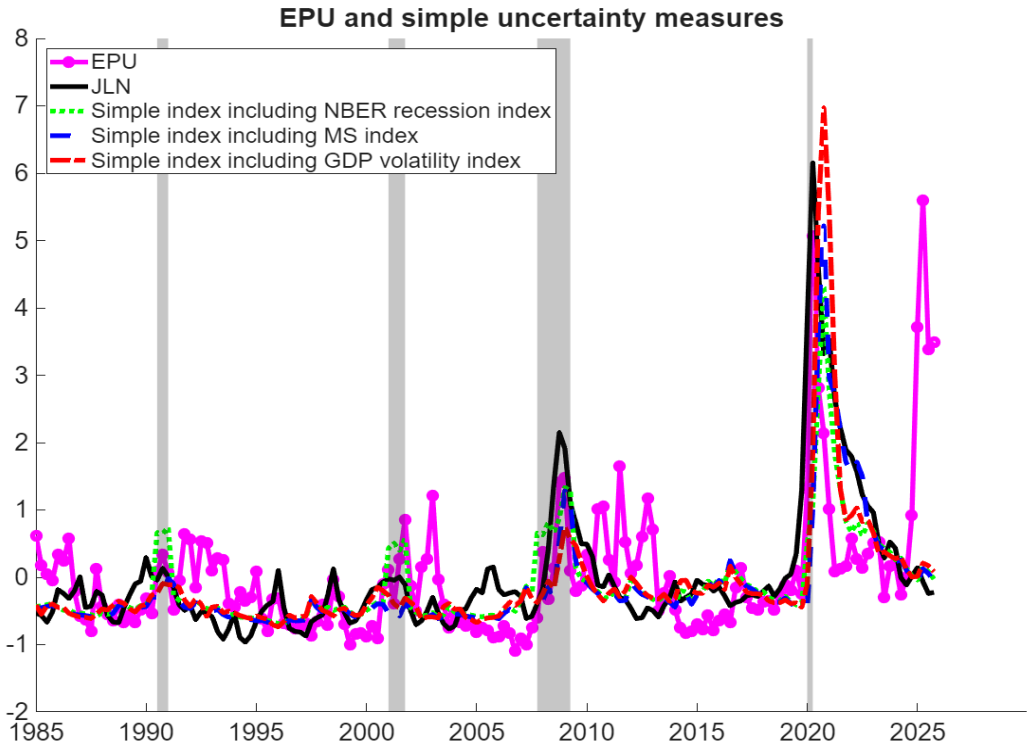



Figure 3. EPU, JLN and simple uncertainty measures. Sample period 1985:1-2025:4.

[13] Baker et al. (2016) consider the frequency with which keywords related to economic uncertainty appear in major U.S. newspapers. In particular, their index of economic policy uncertainty calculates the frequency with which articles in ten major U.S. newspapers contain the following trio of terms: "economic" or "economy"; "uncertain" or "uncertainty"; and one or more of "Congress," "deficit," "Federal Reserve," "legislation," "regulation," or "White House." The EPU measure is updated regularly and is available on the St. Louis Federal Reserve's website.

Turning to Table 2, we first note that the table has been constructed to follow a similar structure as Table 1, so we will not describe its structure here. However, one noteworthy difference is that Table 2 includes the JLN index in the statistical analysis. Comparing Tables 1 and 2, it is clear that it is harder to reproduce the EPU dynamics with the various simple uncertainty indices. For instance, all the regressions have much lower adjusted R-squared values than in Table 1, and several show coefficients that are not significantly different from zero. The JLN index does fit the EPU index somewhat better than our uncertainty indices in terms of the adjusted R-squared and the correlation with the EPU index.[14] However, the rolling standard deviation of the correlation between these two indices is considerably larger, indicating at least some performance differences across the twenty-year rolling windows. That the simple uncertainty indices fit the JLN index better than the EPU index is not surprising since the simple uncertainty indices and the JLN index use to some extent very similar construction methods, while the EPU measure does not make use of standard statistical data series at all, and instead is constructed from major newspaper sources.

[14]Note, the regression for the EPU on JLN has only a constant term and a single independent variable, so the Adjusted R-squared and the Adjusted R-squared for simple OLS are identical.

Table 2. Regression results of EPU on JLN and the simple uncertainty indices

| | | Simple index based on | | |
|---|---|---|---|---|
| Period: 1985:1-2025:4 | JLN | MS-UM | Recession dummy | GDP realized volatility |
| Constant | 0.0000 | 0.0000 | 0.0000 | -0.0000 |
| | (-0.1342, 0.1342) | (-0.1425, 0.1425) | (-0.1397, 0.1397) | (-0.1358, 0.1358) |
| JLN | 0.4968/*0.2468* | | | |
| | (0.3230, 0.6706) | | | |
| MS-UM | | 0.1463/*0.0541* | | |
| | | (-0.2147, 0.5074) | | |
| Recession dummy | | | 0.2136/*0.0453* | |
| | | | (0.0306, 0.3966) | |
| GDP realized volatility | | | | 0.8096/*0.1162* |
| | | | | (0.0500, 1.5693) |
| GDP-UM | | 0.1920/*0.0622* | 0.2452/*0.0850* | -0.5254/*0.0625* |
| | | (0.0487, 0.3354) | (0.0924, 0.3980) | (-1.1912, 0.1405) |
| Inflation-UM | | 0.1448/*0.0451* | 0.2080/*0.0646* | 0.2767/*0.0603* |
| | | (-0.0752, 0.3648) | (0.0735, 0.3425) | (0.1350, 0.4183) |
| Adjusted R-squared | 0.2375 | 0.1404 | 0.1747 | 0.2199 |
| Adj. R-sqrd. for simple OLS | 0.2375 | 0.1504 | 0.1845 | 0.1691 |
| Correlation with EPU | 0.4968 | 0.4010 | 0.4410 | 0.4234 |
| Std. dev. rolling correlation | 0.2134 | 0.1227 | 0.0783 | 0.1292 |
| | (0.1854, 0.2514) | (0.1066, 0.1446) | (0.0681, 0.0923) | (0.1122, 0.1521) |

Notes: The last two rows show the standard deviation and the 95% confidence interval of the twenty-year rolling-window correlation between the EPU index and the corresponding index.

# 5 Robustness

This section is divided into two parts. The first part investigates two richer simple indices than those studied in Section 4, while the second part offers commentary on other robustness exercises.

## 5.1 Other variables in the simple model

Many good uncertainty measures can be generated using this simple algorithm. We focused on the three discussed in Section 4 in part, because they are simple, but also because these are useful for the application discussed in Section 6, which generates an uncertainty index into the distant past. One attraction of this simple strategy, is that it can be used in sparse data settings, so an illustration into the distant past shows the value of the strategy. However, before undertaking that application, we now investigate whether the simple indices can be improved with more data. In this subsection we consider two richer indices to see whether additional variables improve the simple indices. One of these richer indices is a three-variable index that also includes a policy rate variable, which is the third variable included in the early New-Keynesian models, and the other includes the seven variables used in the medium-scale New Keynesian models estimated by Smets and Wouters (2007).[15] Also, as discussed below, and in Part 4 of the supplementary appendix, we further investigated whether financial variables can improve the simple indices because these financial variables are part of the next generation of medium-scale New Keynesian models that address financial frictions (e.g., Bernanke et al. 1999; Gertler and Karadi 2011; Christiano et al. 2014). Here we show that the first two sets of additional variables do not add much to the three simple indices described in Section 4 and we simply discuss the financial variable results.

[15] These seven variables are the policy interest rate, consumption, investment, hours worked and real wages, as well as GDP and inflation.

Table 3. Regression results of JLN on the alternative simple uncertainty indices with the policy rate

| | | Simple index based on | |
|---|---|---|---|
| Period: 1960:4-2025:4 | MS-UM | Recession dummy | GDP realized volatility |
| Constant | 0.0000 | 0.0000 | -0.0000 |
| | (-0.0844, 0.0844) | (-0.0796, 0.0796) | (-0.0817, 0.0817) |
| MS-UM | 0.3510/*0.1744* | | |
| | (0.1737, 0.5282) | | |
| Recession dummy | | 0.3732/*0.1410* | |
| | | (0.2053, 0.5410) | |
| GDP realized volatility | | | 0.6370/*0.2554* |
| | | | (0.1256, 1.1485) |
| GDP-UM | 0.4041/*0.2065* | 0.4409/*0.2367* | -0.0769/*0.1306* |
| | (0.2088, 0.5994) | (0.2332, 0.6486) | (-0.4858, 0.3321) |
| Inflation-UM | 0.1792/*0.1129* | 0.3401/*0.1649* | 0.2959/*0.1347* |
| | (0.0328, 0.3256) | (0.2585, 0.4217) | (0.1781, 0.4137) |
| Policy rate-UM | 0.0370/*0.0343* | 0.0209/*0.0377* | 0.0426/*0.0375* |
| | (-0.0572, 0.1312) | (-0.0731, 0.1150) | (-0.0955, 0.1807) |
| Adjusted R-squared | 0.5189 | 0.5720 | 0.5494 |
| Adjust. R-squared for simple OLS | 0.4618 | 0.4975 | 0.5038 |
| Correlation with JLN | 0.6826 | 0.7081 | 0.7125 |
| Std. dev. rolling correlation | 0.1552 | 0.1108 | 0.1776 |
| | (0.1407, 0.1730) | (0.1005, 0.1235) | (0.1610, 0.1980) |

Notes: The last two rows show the standard deviation and the 95% confidence interval of the twenty-year rolling-window correlation between the JLN index and the corresponding index.

Table 3 shows the results for three uncertainty indices which use all three of the variables from the small-scale New Keynesian models. To be clear, these three uncertainty indices are generated

using the one quarter ahead forecast error standard deviations for output, inflation and the policy interest rate (i.e. the effective federal funds rate) along with one of the three state of the economy variables used earlier.[16] Here, the MS-UM index is also modified to include these three variables, rather than the two variable model used earlier.

To understand whether the policy interest rate series improves the performance of the three uncertainty indices one should compare Table 3 with Table 1. Start by looking at the various state variable coefficients and the coefficients on GDP-UM and Inflation-UM, along with their respective confidence intervals. With the exception of the GDP-UM coefficient in the GDP realized volatility index, these coefficients are always significant in each simple index, indicating they belong in the various indices. Next look at the coefficients on Policy rate-UM and their associated confidence intervals. The confidence intervals show that the policy interest rate coefficients are insignificantly different from zero for all three indices, indicating that they do not improve upon the ability to match the JLN index. Next, look at the rows for the Adjusted R-squared (for the regression with a constant and four independent variables), the Adjusted R-squared for simple OLS (for the regression with a constant and one simple index), and the correlation with the JLN index. For the first two simple indices (MS-UM and Recession dummy), the values in Table 3 are marginally smaller than the values in Table 1, while for the GDP growth volatility index, these statistics are marginally larger. Because the Policy rate-UM series is always insignificant and has only marginal effects on the ability of an index to match the JLN index we left it out of the baseline simple indices in Section 4.

Next consider simple uncertainty indices which include all seven of the variables from the medium-scale New Keynesian models. As with the simple uncertainty indices, each of the additional variables used an ARIMA model with GARCH(1,1) errors to generate the individual one step ahead conditional forecast error standard deviations.[17] Here, we did not expand the variables in the MS-VAR index and simply used the three variable MS-VAR index used to generate the

[16] Following the BIC selection approach, the policy interest rate forecasting model is an ARIMA(2,0,1).

[17] Following the BIC selection criterion for the additional variables, the forecasting models are as follows: consumption (ARIMA(0,1,0)), investment (ARIMA(2,1,0)), hours worked (ARIMA(1,0,0)), and real wage (ARIMA(0,1,0)).

series in Table 4 above.

Start by looking at a plot of the various series. Figure 4 shows a plot of the JLN measure (solid black line), a medium-scale index using the NBER recession index for the state variable (green doted line), a medium-scale index using a three variable MS-VAR recession probably index (blue long dashed line) and a medium-scale index with GDP volatility index (red short dashed line). Like Figure 2, Figure 4 shows that all three medium-scale indices track the JLN index over time. This shows that there are many good uncertainty indices one might consider. So to make a choice about which variables to include we consider various statistical measurements.

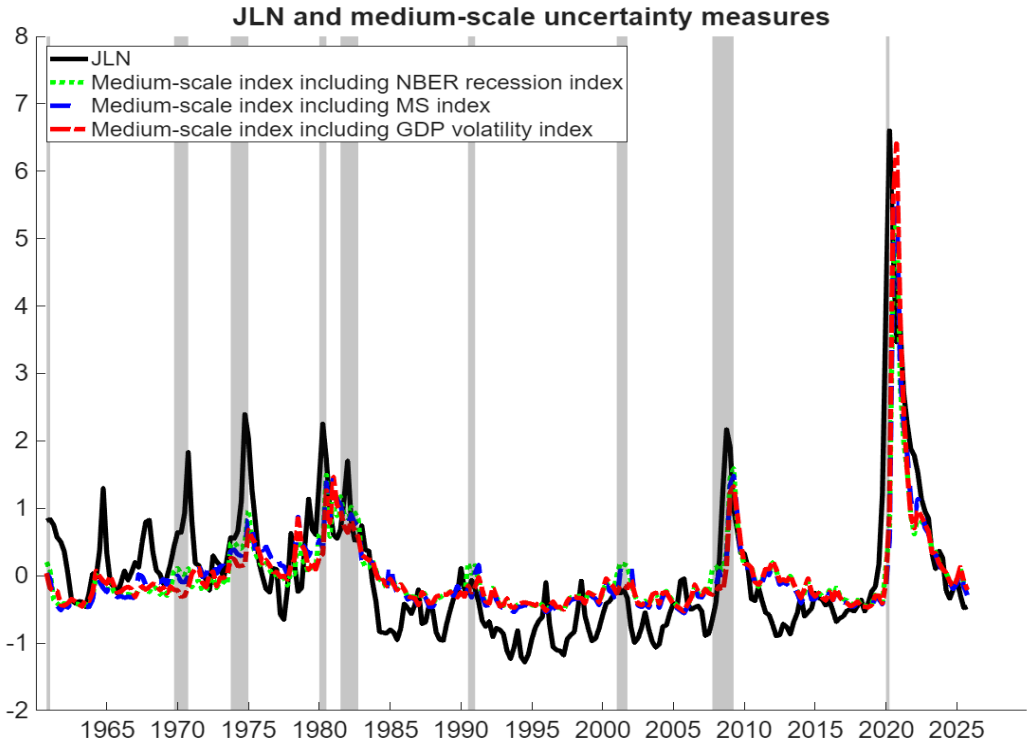



Figure 4. JLN and medium-scale uncertainty measures. Sample period 1960:4-2025:4.

Table 4 shows statistical measurements that allows one to compare the many different candidate simple indices. This table has a similar format as Tables 1 and 3, only now the regressions have a constant and eight independent variables. Scanning across the various state of the economy index coefficients, one sees they are always significant. Next scanning the various forecast error

standard deviation coefficients, one sees that Inflation-UM is always significant, GDP-UM, Policy rate-UM, Consumption-UM, Investment-UM, Wage-UM are never significant, and mixed results for Hours-UM.[18] One also sees that the Adjusted R-squared (for the regression with a constant and eight independent variables) are in some cases higher than the analogues in Table 1, while the Adjusted R-squared for simple OLS (for the regression with a constant and one index), and the correlation with the JLN index are all smaller. That the Adjusted R-squared is larger in few cases is not surprising since there are five additional variables. The fact that the variable GDP-UM is never significant is somewhat bothersome, but we attributed this to that series having similar information content as some of the other series in the regression.

Overall, we can conclude that including more data into the computation of an uncertainty index can improve the R-squared in some cases and perhaps suggest different variables one might want to include in the index.[19] But these differences are marginal and the simple uncertainty indices seen in Section 4 are acceptable and can achieve good performance without a high cost. Furthermore, the simple indices in Section 4 are particularly attractive in sparse data settings and are useful in the historical illustration in Section 6, when many of these other variables are not available too far into the past.

[18] A reader might wonder if the lack of contribution of the policy rate-UM to explaining the JLN index is because the sample period includes the Great Recession and the Covid-19 pandemic, during which the federal funds rate effectively hit its zero lower bound. In preliminary estimation attempts, we also considered the shadow rate suggested by Wu and Xia (2016) to address this issue. The shadow rate is equal to the federal funds rate when the zero lower bound is not binding. However, it is negative in order to account for the implementation of unconventional policy tools when the federal funds rate approaches the zero lower bound. This occurred during the Great Recession and its aftermath, from 2009 to 2015, as well as during the 2020-2022 pandemic. The estimation results are quantitatively the same when using the shadow rate instead of the federal funds rate.

[19] Among the other models we considered in preliminary attempts to expand our simple indices, we considered including popular measures of economic and financial uncertainty. These measures include the University of Michigan Consumer Sentiment Index, the S&P 500 Volatility Index (VIX) for the 30 days following the measurement date, and spread forecast error conditional volatility. The spread is defined as the Moody's seasoned Baa corporate bond yield relative to the 10-year Treasury constant maturity yield (BAA10Y). As an alternative to this measure of uncertainty, we further considered the conditional volatility of excess bond premium (EBP) forecast errors, where the EBP was estimated by Gilchrist and Zakrajšek (2012). These preliminary estimation results suggest that extending the simple indices with these measures only improves the fit of the JLN slightly. As shown in Table A.2 in Part 4 of the supplementary appendix, this does not mean that these other uncertainty indicators are not informative for fitting the JLN index. They simply do not provide additional information.

Table 4. Regression results of JLN on the medium-scale uncertainty indices

| | | Simple index based on | |
|---|---|---|---|
| Period: 1960:4-2025:4 | MS-UM | Recession dummy | GDP realized volatility |
| Constant | 0.0000 | 0.0000 | 0.0000 |
| | (-0.0841, 0.0841) | (-0.0790, 0.0790) | (-0.0810, 0.0810) |
| MS-UM | 0.3861/*0.1534* | | |
| | (0.1924, 0.5798) | | |
| Recession dummy | | 0.4065/*0.1430* | |
| | | (0.2257, 0.5872) | |
| GDP realized volatility | | | 0.6823/*0.1926* |
| | | | (0.1750, 1.1896) |
| GDP-UM | 0.1616/*0.0790* | 0.2631/*0.0888* | -0.4197/*0.0714* |
| | (-0.1357, 0.4590) | (-0.0468, 0.5731) | (-0.9405, 0.1010) |
| Inflation-UM | 0.1450/*0.0845* | 0.3108/*0.1202* | 0.2831/*0.1078* |
| | (0.0047, 0.2854) | (0.2400, 0.3815) | (0.1723, 0.3939) |
| Policy rate-UM | 0.0356/*0.0299* | 0.0221/*0.0320* | 0.0264/*0.0321* |
| | (-0.0461, 0.1173) | (-0.0653, 0.1096) | (-0.1008, 0.1535) |
| Consumption-UM | 0.1602/*0.0614* | 0.1674/*0.0656* | 0.2051/*0.0528* |
| | (-0.0814, 0.4019) | (-0.0857, 0.4205) | (-0.0129, 0.4231) |
| Investment-UM | -0.0170/*0.0552* | -0.0829/*0.0579* | 0.0557/*0.0511* |
| | (-0.2465, 0.2124) | (-0.3071, 0.1413) | (-0.1406, 0.2520) |
| Hours-UM | 0.0795/*0.0406* | 0.1118/*0.0501* | 0.0663/*0.0374* |
| | (-0.0384, 0.1974) | (0.0094, 0.2142) | (-0.0622, 0.1947) |
| Wage-UM | 0.0980/*0.0350* | 0.0588/*0.0350* | 0.0664/*0.0272* |
| | (-0.0620, 0.2579) | (-0.0921, 0.2097) | (-0.0831, 0.2159) |

Table 4. (*Continued*)

| Period: 1960:4-2025:4 | MS-UM | Simple index based on Recession dummy | GDP realized volatility |
|---|---|---|---|
| Adjusted R-squared | 0.5225 | 0.5781 | 0.5571 |
| Adjust. R-squared for simple OLS | 0.4803 | 0.4968 | 0.3866 |
| Correlation with JLN | 0.6959 | 0.7076 | 0.6256 |
| Std. dev. rolling correlation | 0.1421 | 0.1096 | 0.1584 |
| | (0.1289, 0.1585) | (0.0994, 0.1222) | (0.1436 0.1766) |

Notes: The last two rows show the standard deviation and the 95% confidence interval of the twenty-year rolling-window correlation between the JLN index and the corresponding index.

## 5.2 Other robustness exercises

The rest of this Robustness section examines additional sensitivity exercises that may be useful in this and other applications.[20]

### MS-VAR models

As noted above, we used a two-variable MS-VAR model with two lags as the baseline specification. In this model, the stationary component of GDP was obtained by removing its linear-quadratic time trend. Additionally, rather than using the linear-quadratic detrended component of GDP to estimate the baseline stationary MS-VAR, we examined alternative definitions of the cyclical component of GDP. Specifically, we considered alternative cyclical components of GDP obtained by: (i) using the growth rate; (ii) removing the Hodrick-Prescott trend (Hodrick and Prescott 1997); (iii) removing the Hamilton trend (Hamilton 2018); (iv) removing a variant of the Hamilton trend that assumes a random walk on GDP. None of these showed any improvement.

[20]For the sake of brevity, we will not present the full set of results from these additional sensitivity analyses, but Part 4 of the supplementary appendix shows the results for a few robustness exercises.

Additionally, we considered other MS-VAR specifications, including (i) a three-state MS-VAR, (ii) one-lag MS-VARs, and (iii) three-variable MS-VARs. Figure 1 shows that the baseline specification clearly identifies the high-volatility state with severe recessionary periods, which are typically characterized by high uncertainty. However, identifying the high-volatility state in the other MS specifications is generally much noisier than what is shown in Figure 1, making it more difficult to visualize the link between the high-volatility state and specific recessions. Nevertheless, these methods could be useful in other applications.

**ARIMA forecasting models**

As noted above, we have used the forecasting models selected by BIC to compute the forecast errors for each variable. However, other selection information criteria, such as the Akaike Information Criterion (AIC) and the Root Mean Squared Error (RMSE) criterion, could also be appropriate for some applications. In this application, however, the performance of the simple uncertainty indices remains robust when other ARIMA models are used.

Alternatively, one could use all the forecasting models for each variable and then use a principal component approach to reduce the number of factors in the computation of the uncertainty index. For example, a simple uncertainty index could be computed using all six alternative ARIMA models, discussed above in footnote 7, for each of the two variables. This allows one to compute twelve alternative forecasting errors and twelve corresponding conditional variances. Then, a principal component analysis can be performed to reduce the dimension of the full set of regressors. This includes the entire set of conditional variances, as well as the corresponding state variable considered in each of the three simple indices. Finally, one can compute an index as a weighted average of a group of small factors, where a factor's weight reflects its proportion of the total variability in the full set of regressors. As shown in Part 4 of the supplementary appendix, this approach resulted in a worse fit for the JLN measure in this application. However, it could be useful for other applications.

**High-frequency uncertainty movements**

Comparing alternative uncertainty measures can be biased due to small, short-lived fluctuations in these measures caused by noise (rather than new signals of uncertainty) in the estimation of forecast errors. One way to address this issue is to remove the high-frequency fluctuations of the uncertainty indices by generating smooth indices. For example, indices can be computed using a symmetric moving average window that includes four quarterly observations. Using this approach, the R-squared statistic for the simple index based on the MS-UM index in the JLN regression increases to 0.7083 (from 0.5196), and the correlation increases to 0.8429 (from 0.7234) when both the JLN and the simple index are smoothed. These increases are similar for the other two simple indices when smoothed. For instance, the R-squared statistic for the smoothed simple index based on the recession dummy increases to 0.7437 (from 0.5734), and the correlation increases to 0.8635 (from 0.7594).[21] While smoothing is useful for increasing the correlation between the JLN and the simple indices suggested here, it could also be useful for other applications. However, smoothing results in shorter spikes and somewhat less clear timing of these spikes.

**Forecast errors of variables in levels or in growth rates?**

As previously mentioned, models that forecast variables with long-run trends, such as the GDP series used in the simple indices, use alternative ARIMA specifications, and the log levels of the forecast errors are used to compute the associated conditional standard deviation. Alternatively, one can compute the forecast errors of the variable's growth rate and calculate the corresponding conditional standard deviation. Note that these two approaches may produce different uncertainty indices because the former includes a trend forecast error component, which the latter ignores. In this application, however, the results are similar for both approaches.

[21]Since applying a symmetric four-quarter moving average to both series before correlating them can artificially increase the correlation between almost any two reasonably persistent series, we conducted the following placebo test. We generated two independent AR(1) series with a high persistence of 0.88, similar to the first-order autoregression coefficients obtained from the simple uncertainty indices. Then, we computed the correlation between the unsmoothed and smoothed time series. The correlation gain from smoothing was about 0.0528, which is approximately 50% lower than the correlation gain obtained from smoothing our simple indices.

# 6 An illustration: Historical simple uncertainty indices

Up to this point, this paper has shown that it is possible to construct simple uncertainty indices from just a few variables and these simple computational methods do a good job of matching the far more complicated JLN index during the post 1960 period. In this section we show that this approach is particularly useful in sparse data environments where data rich indices cannot be computed due to a lack of data. Here we focus on historical annual data for the U.S. dating as far back as 1790. Very little data is available back to this date, but Williamson (2026) constructed annual GDP data and GDP-deflator data since then. To investigate this historical period, we first interpolate annual to quarterly series for real GDP and the GDP-deflator for the entire period from 1790 to 2025. Next, this constructed quarterly data is used to compute the same three simple uncertainty indices described earlier. This data allowed us to compute the MS-UM index and the GDP volatility index over the full 235-year sample, but the NBER recession dummy uncertainty index is constrained to only dates after 1871 when that NBER recession dummy index begins. The results of these calculations can be seen in Figure 5, which plots the same series seen in Figure 2. One noticeable difference is that the JLN index (solid black line) only begins in 1960 and is missing for roughly three quarters of the dates, and the NBER recession dummy uncertainty index is only available back to 1871, when the NBER recession index begins. Figure 5 shows that the various indices show very similar patterns during periods in which there is overlap.

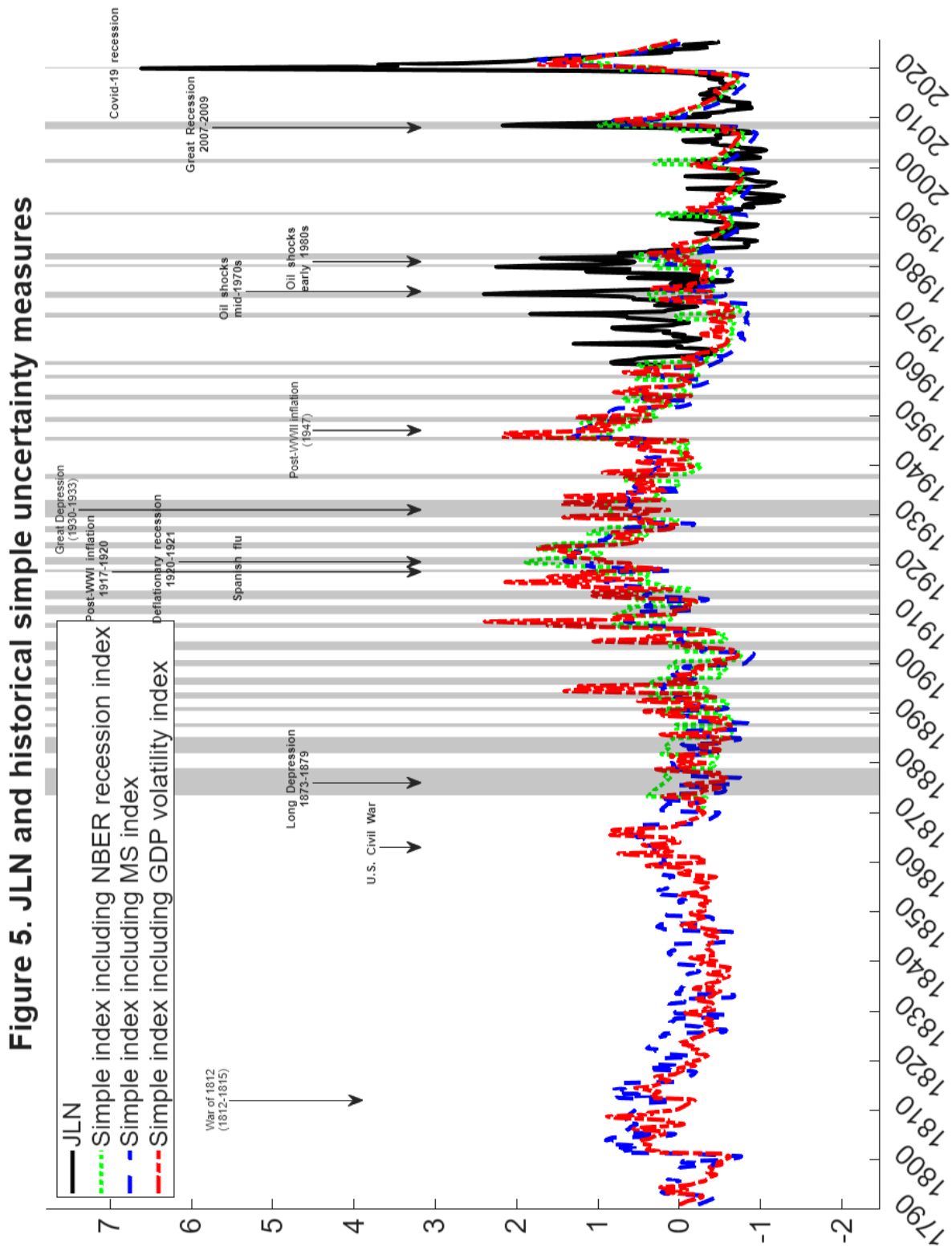

Figure 5. JLN and historical simple uncertainty measures
JLN
Simple index including NBER recession index
Simple index including MS index
Simple index including GDP volatility index
War of 1812 (1812-1815)
U.S. Civil War
Long Depression 1873-1879
Post-WWI inflation 1917-1920
Deflationary recession 1920-1921
Spanish flu
Great Depression (1930-1933)
Post-WWII inflation (1947)
Oil shocks mid-1970s
Oil shocks early 1980s
Great Recession 2007-2009
Covid-19 recession
7
6
5
4
3
2
1
0
-1
-2
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020

Table 5. Regression results of JLN on the alternative simple uncertainty indices using historical data

| | | Simple index based on | |
|---|---|---|---|
| Period: 1960:4-2025:4 | MS-UM | Recession dummy | GDP realized volatility |
| Constant | 0.0000 | 0.0000 | -0.0000 |
| | (-0.0886, 0.0886) | (-0.0913, 0.0913) | (-0.0917, 0.0917) |
| MS-UM | 0.4879/*0.2763* | | |
| | (0.3071, 0.6686) | | |
| Recession dummy | | 0.3870/*0.1600* | |
| | | (0.2362, 0.5379) | |
| GDP realized volatility | | | 0.3970/*0.2005* |
| | | | (0.2189, 0.5751) |
| GDP-UM | 0.1723/*0.1074* | 0.3809/*0.1713* | 0.2920/*0.1377* |
| | (0.0671, 0.2774) | (0.2215, 0.5404) | (0.1749, 0.4092) |
| Inflation-UM | 0.1989/*0.0943* | 0.2339/*0.1140* | 0.2153/*0.1029* |
| | (0.0681, 0.3297) | (0.1183, 0.3494) | (0.0756, 0.3550) |
| Adjusted R-squared | 0.4699 | 0.4367 | 0.4323 |
| Adjusted R-squared small-scale UM | 0.4383 | 0.4311 | 0.4351 |
| Correlation JLN with small-scale UM | 0.6653 | 0.6599 | 0.6629 |
| Std. deviation of rolling correlation | 0.2262 | 0.1169 | 0.2208 |
| | (0.2051, 0.2521) | (0.1060, 0.1303) | (0.2003, 0.2462) |

Notes: The last two rows show the standard deviation and the 95% confidence interval of the twenty-year rolling-window correlation between the JLN index and the corresponding index.

Before getting into what the indices tell us about these historical periods, we next conduct some statistical analysis to compare the interpolated series with the earlier non-interpolated quarterly

data series to see how they perform relative to each other. We start with Table 5 which computes the same set of statistics as in Table 1, only here the table is based on the new interpolated data series. As the table indicates, only the portion of these series from 1960:4 are used in the statistical analysis. A comparison of Tables 5 and 1 reveals that the R-squared and correlation statistics in Table 5 are smaller than those in Table 1 but still fairly similar. This suggests that the performance of the series using interpolated and non-interpolated quarterly data is similar.[22]

Next consider Table 6 which shows cross-correlations between the three interpolated data indices over two periods of time. The upper panel focuses on the period from 1871 to 2025, which is the period in which data is available to construct the NBER recession dummy simple index, while the lower panel focuses on the period from 1791 to 2025, which is the period in which data is available to construct the other two simple indices. Table 6 shows that the indices are highly correlated with each other over both periods.

Having shown that the historical data simple indices have similar performance characteristics as the JLN index and the simple indices computed with modern data during the post 1960 period, we now investigate what these different indices tell us about the economy during various historical time periods. Some of these comparisons are within the range of observations for the rich data indices, and for those comparisons, we highlight differences in the stories told by the different indices. On the other hand, some of these comparisons are made during periods when rich data series are unavailable. For those periods, we simply note what the historical simple indices tell us about the economy.

[22]There are two notable differences between Tables 1 and 5 in the GDP realized volatility model. First, the estimated coefficient of realized GDP volatility is half as high in Table 5 as in Table 1. Second, the estimated GDP-UM coefficient is significant in Table 5 but not in Table 1. Additionally, the significant GDP realized volatility coefficient in Table 1 (0.6637) is nearly equal to the sum of the significant GDP realized volatility and GDP-UM coefficients in Table 5 (0.6890). We interpret these results as a consequence of the interpolated quarterly time series being smoother than the original quarterly data. The smoother, less noisy interpolated data allows us to distinguish the unpredictable component of GDP volatility, captured by GDP-UM, from the predictable sources of volatility captured by the realized GDP volatility time series.

Table 6. Cross-correlations of the alternative simple indices using historical data

| | | Simple index based on | |
|---|---|---|---|
| | | Recession | GDP |
| Period 1871-2025 | MS-UM | dummy | volatility |
| Simple index based on MS-UM | 1.0 | | |
| Simple index based on recession dummy | 0.7916 | 1.0 | |
| Simple index based on GDP volatility | 0.9012 | 0.7407 | 1.0 |
| | | | |
| Period 1791-2025 | | | |
| Simple index based on MS-UM | 1.0 | | |
| Simple index based on GDP volatility | 0.8027 | | 1.0 |

We begin with a comparison of business cycles. Figure 5 clearly shows that NBER recessions tend to be shorter and less frequent during recent years. From a historical perspective, this is good news. Not only has the U.S. economy grown over the last two centuries, but turbulent times featuring recessions have also tended to be shorter and less frequent during recent years. Figure 5 also shows a close link between recessions and the historical simple indices. Spikes in the historical simple indices occur consistently around recession periods. Therefore, one may conclude that historical uncertainty spikes are also shorter and less frequent. This favorable trend of less frequent recessions and uncertainty spikes is evident in the post-WWII period and becomes clearer when viewed from a long-term historical perspective.

Next, let us compare two similar, yet different events. Here we compare the Covid-19 pandemic and the Spanish flu.[23] The spike in uncertainty resulting from the Covid-19 pandemic is historically high for all four plotted measures. This spike was notably higher than the one associated with the Spanish flu. Why is that? The answer is beyond the scope of this paper, but one possible reason

[23]The Spanish flu occurred in the U.S. from March 1918 through spring 1919 in three distinct waves. The second wave, which occurred in the fall of 1918, was the deadliest. Additionally, observe that the arrows corresponding to Post-WWI inflation (1917–1920) and the Spanish flu overlap in Figure 5.

is differences in (mis)information. While news of the spread of Covid-19 from China was widely disseminated on social media, radio, TV, and the internet, information about the Spanish flu was much more limited and misleading.[24]

Finally, let us rank the uncertainty of several dissimilar historical events. Figure 5 shows the spikes in the simple indices around the Great Depression are a bit larger than the Great Recession, and both spikes are considerably larger than those around the oil crises of the 1970s and early 1980s. On the other hand, the JLN index ranks the oil crisis of the 1970s and Great Recession as equally uncertain. Alternatively, Figure 3 shows that the EPU index spikes around the beginning of 2025, which none of the other indices show. That spike in the EPU is similar in magnitude to the Covid-19 spike, and is likely due to the significant changes implemented during the early part of the second term of the Trump administration, with the introduction of tariffs and other geopolitical turmoil. While it is easy to understand why the EPU index spikes, whether that spike reveals an economic downturn of similar magnitude as the Covid-19 recession is yet to be seen. The historical indices suggest that this is not the case.

# 7 Conclusion

This paper proposes a low-cost strategy for creating simple measures of economic uncertainty. Unlike recent measures of uncertainty that incorporate massive or specialized data sets (Jurado et al. 2015; Rossi and Sekhposyan 2015; Baker et al. 2016; Scotti 2016), our measure uses a small set of widely available data variables, which summarize the economic outlook. Similar to the approach of Jurado et al. (2015), our approach calculates the conditional forecast error standard deviations of a set of variables in the construction of the uncertainty index. However, our approach has several notable differences in the conditional forecast error computations. First, our conditioning

[24]Many historical studies have found that the early Spanish flu outbreaks were in military camps during WWI, such as in Étaples, France, in 1916–1917 and Fort Riley, Kansas, in early 1918. However, these outbreaks were hidden by the countries involved in the conflict, such as the U.S., the U.K., France, and Germany, because they enforced wartime press censorship to hide the severity of the illness and protect military morale. However, Spain was neutral during WWI, so its press could freely report on the widespread illness, including when King Alfonso XIII fell ill. This made it appear to the rest of the world as though Spain was the epicenter of the outbreak, hence the name "Spanish flu," even though that has been shown to be untrue with time.

set only uses own data information in the computation of the conditional forecast error standard deviations, and second, the number of variables ultimately used to construct the index is much smaller. Furthermore, our low-cost strategy differs in another notable way. The strategy combines the conditional forecast error standard deviations of a small set of variables with information about the state of the economy. In the applications studied here, three different state of the economy uncertainty measures are investigated, but other states of the economy could also be used in different applications.

It is shown that the simple uncertainty indices behave similarly to the JLN index using recent U.S. data, illustrating that, despite its simplicity, this approach is a promising tool for measuring uncertainty even in rich data settings. However, where this approach really has merit is in sparse data settings in which the massive data methods are not feasible. It is shown how to apply this low-cost strategy using historical data for the U.S. back to the late 18th century. This construction showed that the nature of the business cycle has changed over time, with longer cycles and less uncertainty volatility during recent years. The construction also allows one to compare different economic events from different dates easily. Furthermore, these methods could be useful in other sparse information settings. For instance, data collection is often more limited in other countries than in the U.S., which makes studying the role of uncertainty in those economies difficult. With the low-cost strategy described here, it is possible that uncertainty issues can be investigated in many international settings that heretofore where unavailable for analysis.

# References

Baker, S. R., Bloom, N. and Davis, S. J.: 2016, Measuring economic policy uncertainty, *Quarterly Journal of Economics* **131**(3), 1593–1636.

Bernanke, B. S., Gertler, M. and Gilchrist, S.: 1999, The financial accelerator in a quantitative business cycle framework, *Handbook of Macroeconomics* **1**, 1341–1393.

Carriero, A., Clark, T. E. and Marcellino, M.: 2018, Measuring uncertainty and its impact on the economy, *Review of Economics and Statistics* **100**, 799–815.

Christiano, L. J., Motto, R. and Rostagno, M.: 2014, Risk shocks, *American Economic Review* **104**(1), 27–65.

Gertler, M. and Karadi, P.: 2011, A model of unconventional monetary policy, *Journal of Monetary Economics* **58**(1), 17–34.

Gilchrist, S. and Zakrajšek, E.: 2012, Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4), 1692–1720.

Hamilton, J. D.: 1994, *Time Series Analysis*, Princeton University Press.

Hamilton, J. D.: 2018, Why you should never use the Hodrick-Prescott filter, *Review of Economics and Statistics* **100**(5), 831–843.

Hodrick, R. J. and Prescott, E. C.: 1997, Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit and Banking* **29**(1), 1–16.

Jo, S. and Sekkel, R.: 2019, Macroeconomic uncertainty through the lens of professional forecasters, *Journal of Business & Economic Statistics* **37**(3), 436–446.

Jurado, K., Ludvigson, S. and Ng, S.: 2015, Measuring uncertainty, *American Economic Review* **105**(3), 1177–1216.

Rossi, B. and Sekhposyan, T.: 2015, Macroeconomic uncertainty indices based on nowcast and forecast error distributions, *American Economic Review: Papers & Proceedings* **105**(5), 650–655.

Scotti, C.: 2016, Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises, *Journal of Monetary Economics* **82**, 1–19.

Smets, F. and Wouters, R.: 2007, Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review* **97**(3), 586–606.

Williamson, S. H.: 2026, The annual real and nominal gdp for the united states, 1790 - 1928, *Measuring Worth* .

Xu, J. and Rogers, J.: 2025, How well does uncertainty forecast economic activity?, *Journal of Money, Credit and Banking* **57**, 645–662.

# Supplementary Appendix (not intended for publication)

## Part 1

Following closely Hamilton (1994), this part of the appendix briefly describes the recursive algorithm used in the maximum likelihood estimation procedure. We limit our attention to the two-state case considered in this paper. Let $\theta$ denote the vector of parameters. Let $\hat{\xi}_{t/t}$ denote the $2 \times 1$ vector containing the econometric inference about the values of $s_t$ $(= 1, 2)$ based on data obtained through date $t$ and conditional on a given value for $\theta$. Finally, let $\hat{\xi}_{t+1/t}$ denote the $2 \times 1$ vector containing the one-period forecast about the values of $s_{t+1}$ $(= 1, 2)$ based on data obtained through date $t$. As described in Hamilton (1994), the optimal inference and the one-period forecast can be solved recursively from the following two equations:

$$\hat{\xi}_{t/t} = \frac{\hat{\xi}_{t/t-1} \bigodot \eta_t}{\mathbf{1}' \left( \hat{\xi}_{t/t-1} \bigodot \eta_t \right)},$$

$$\hat{\xi}_{t+1/t} = P \hat{\xi}_{t/t},$$

where $\bigodot$ denotes element-by-element multiplication, $\mathbf{1}$ denotes a $2 \times 1$ vector of 1s, $P$ is the $2 \times 2$ transition probability matrix, and $\eta_t$ is a $2 \times 1$ vector containing the two conditional densities, one for each state:

$$\eta_t = \begin{bmatrix} \eta_t(1) \\ \eta_t(2) \end{bmatrix},$$

where

$$\eta_t(s_t) = \frac{1}{2\pi \left| \Omega(s_t) \right|^{\frac{1}{2}}} exp \left\{ -\frac{1}{2} \left[ Y_t - \mu(s_t) - A(s_t) Y_{t-1} \right]' \left[ \Omega(s_t) \right]^{-1} \left[ Y_t - \mu(s_t) - A(s_t) Y_{t-1} \right] \right\}.$$

The log likelihood function $\mathcal{L}(\theta)$ for the data set evaluated at a value of $\theta$ can be calculated using

the previous algorithm from the following expression

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} ln\left[1'\left(\hat{\xi}_{t/t-1} \bigodot \eta_t\right)\right].$$

The value of $\theta$ that maximizes $\mathcal{L}(\theta)$ is found using the maximum likelihood routine programmed in GAUSS. The Newton-Raphson numerical algorithm is typically used for the update of the Hessian at each iteration of the maximization procedure.

## Part 2

Table A.1. shows the MS-VAR estimation results.

Table A.1. MS-VAR estimation results

| Parameter | Estimate | Std. error | Parameter | Estimate | Std. error |
|---|---|---|---|---|---|
| $\mu_1(1)$ | 0.5132 | 0.1407 | $\mu_1(2)$ | 1.8537 | 0.7121 |
| $\mu_2(1)$ | 0.1399 | 0.0702 | $\mu_2(2)$ | -1.0057 | 0.8874 |
| $a_{1,11}(1)$ | 0.5650 | 0.0806 | $a_{1,11}(2)$ | 0.5938 | 0.1823 |
| $a_{1,12}(1)$ | 0.1967 | 0.0950 | $a_{1,12}(2)$ | 0.0706 | 0.2458 |
| $a_{1,21}(1)$ | 0.0114 | 0.0369 | $a_{1,21}(2)$ | -0.3634 | 0.3048 |
| $a_{1,22}(1)$ | 1.1231 | 0.0664 | $a_{1,22}(2)$ | 0.7109 | 0.3180 |
| $a_{2,11}(1)$ | 0.1835 | 0.0690 | $a_{2,11}(2)$ | 0.2224 | 0.1808 |
| $a_{2,12}(1)$ | -0.1361 | 0.1123 | $a_{2,12}(2)$ | -0.0212 | 0.2414 |
| $a_{2,21}(1)$ | -0.0150 | 0.0363 | $a_{2,21}(2)$ | 0.3884 | 0.3643 |
| $a_{2,22}(1)$ | -0.1616 | 0.0703 | $a_{2,22}(2)$ | 0.1428 | 0.3619 |
| $\sigma_{11}(1)$ | 0.7322 | 0.0545 | $\sigma_{11}(2)$ | 1.5503 | 0.2409 |
| $\sigma_{12}(1)$ | -0.0292 | 0.0442 | $\sigma_{12}(2)$ | 1.3560 | 0.4436 |
| $\sigma_{22}(1)$ | 0.5067 | 0.0283 | $\sigma_{22}(2)$ | 1.3198 | 0.1927 |
| $p_{11}$ | 0.9842 | 0.0121 | $p_{22}$ | 0.9252 | 0.0251 |
| $L$ | -1073.3702 | | | | |

Notes: $\mu_i(s_t)$ denotes a generic element of vector $\mu(s_t)$ , $a_{l,ij}(s_t)$ denotes a generic element of matrix $A_l(s_t)$,

$\sigma_{ij}(s_t)$ denotes a generic element of $\Sigma(s_t)$, and $L$ is the value of the estimated log likelihood function.

The implied estimates of the covariance matrices for the two states are given by

$$\Omega(1) = \begin{bmatrix} 0.5361 & -0.0214 \\ -0.0214 & 0.2576 \end{bmatrix}, \Omega(2) = \begin{bmatrix} 2.4033 & 2.1022 \\ 2.1022 & 3.5806 \end{bmatrix}.$$

## Part 3

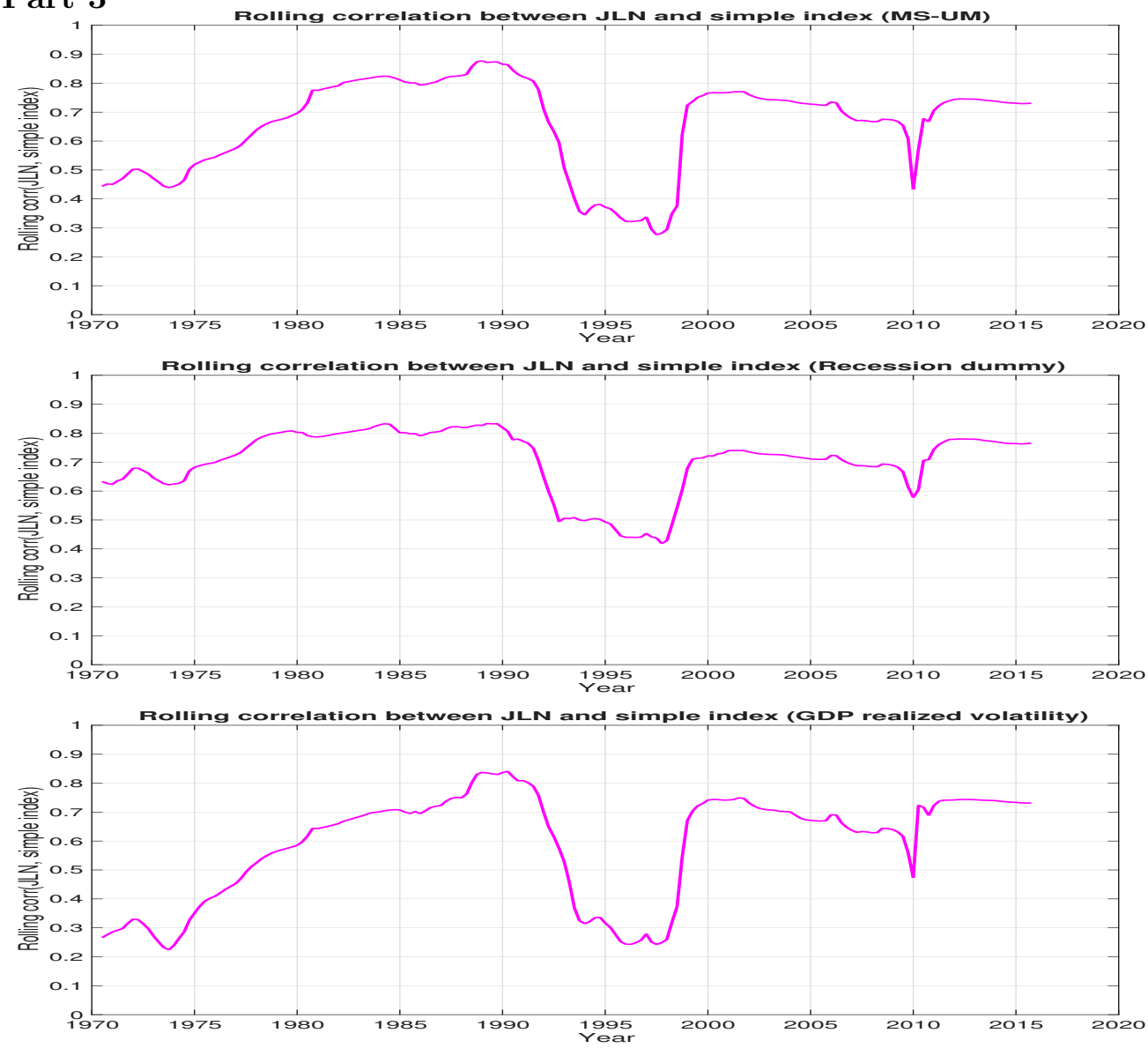



Figure A.1. Rolling correlation between JLN and historical simple uncertainty measures.

Figure A.1 show the 20-year rolling correlation of JLN index with each of the three simple uncer-

tainty indices. These plots show that the relatively low correlations mainly occur during the Great Moderation period of the 1990s, which was also a period of time that Carriero et al. (2018) found differences with the JLN index.

## Part 4

### Other measures of economic and financial uncertainty

In preliminary attempts in looking for a simple uncertainty index, we also considered various popular measures of economic and financial uncertainty using a shorter sample since the available starting observation can be different (for instance, the VIX time series starts in 1990:1).[25] Table A.2 shows the results of the regression analysis of the JLN index against a constant and these measures of economic and financial uncertainty, considered individually. The indices studied here are the University of Michigan Consumer Sentiment Index (CSI); the conditional volatility of the CSI index forecast error (CSI-UM); the S&P 500 volatility index for the 30 days following the measurement date (VIX); the VIX index forecast error conditional volatility (VIX-UM); the spread forecast error conditional volatility (BAA10Y-UM), where the spread is defined as the Moody's seasoned Baa corporate bond yield relative to the 10-year Treasury constant maturity yield; the excess bond premium (EBP) forecast error conditional volatility, where the EBP is estimated by Gilchrist and Zakrajšek (2012); and the probability that the U.S. economy will enter a recession within the next 12 months, also estimated by Gilchrist and Zakrajšek (2012). As noted in footnote 19 and shown in Table A.2, these series contain information related to uncertainty. However, when used with the constructed series in the simple indices, this information does not improve performance.

[25] The last observation of the dataset used in these preliminary exercises is 2023:1.

Table A.2. Regression results of JLN on other popular uncertainty indices (1990:4-2023:1)

| Variable | Constant | Coefficient | Adjusted R-squared | Correlation with JLN |
|---|---|---|---|---|
| CSI | 0.0000 | -0.3819 | 0.1325 | -0.3819 |
| | (-0.1604, 0.1604) | (-0.5259, -0.2378) | | |
| CSI-UM | 0.0000 | 0.4250 | 0.1678 | 0.4250 |
| | (-0.1571, 0.1571) | (0.1854 0.6646) | | |
| VIX | 0.0000 | 0.4474 | 0.1877 | 0.4474 |
| | (-0.1552, 0.1552) | (0.2471, 0.6477) | | |
| VIX-UM | 0.0000 | 0.3121 | 0.0833 | 0.3121 |
| | (-0.1649, 0.1649) | (0.0499, 0.5744) | | |
| BAA10Y-UM | 0.0000 | 0.3240 | 0.0910 | 0.3240 |
| | (-0.1642, 0.1642) | (0.1851, 0.4630) | | |
| EBP | 0.0000 | 0.1584 | 0.0099 | 0.1584 |
| | (-0.1713, 0.1713) | (-0.0179, 0.3346) | | |
| EBP-UM | 0.0000 | 0.2806 | 0.0643 | 0.2806 |
| | (-0.1666, 0.1666) | (0.1511, 0.4101) | | |
| Prob. of Recession | 0.0000 | 0.1414 | 0.0047 | 0.1414 |
| | (-0.1718, 0.1718) | (-0.0312, 0.3140) | | |

**Other MS-specifications and sample periods**

Figure A.2 shows the probability of a high-volatility state for the following models: (i) the two-variable MS-VAR (2), which is used in one of the historical simple indices; (ii) the two-variable MS-VAR (2) used in the corresponding simple index in Table 1, computed for the sample period of 1960:4 to 2025:4; and (iii) the three-variable MS-VAR (2), which includes the policy rate and is used in the corresponding simple index in Table 3, computed for the sample period of 1960:4 to 2025:4. As can be seen in the figure, the three probabilities largely overlap in the most recent sample. Thus, the probability obtained for the historical sample (black line) is difficult to distinguish from the others. Furthermore, the two probabilities based on the two- and three-variable MS models

(the blue and red lines, respectively) overlap significantly during the oil crisis recessions of the 1970s and early 1980s, as well as during the Great Recession and the Covid-19 recession. However, only the three-variable MS model associates the milder recessions of the early 1990s and 2000s with the high-volatility state.

**Figure A.2. Probability of a high-volatility state for alternative models/samples**

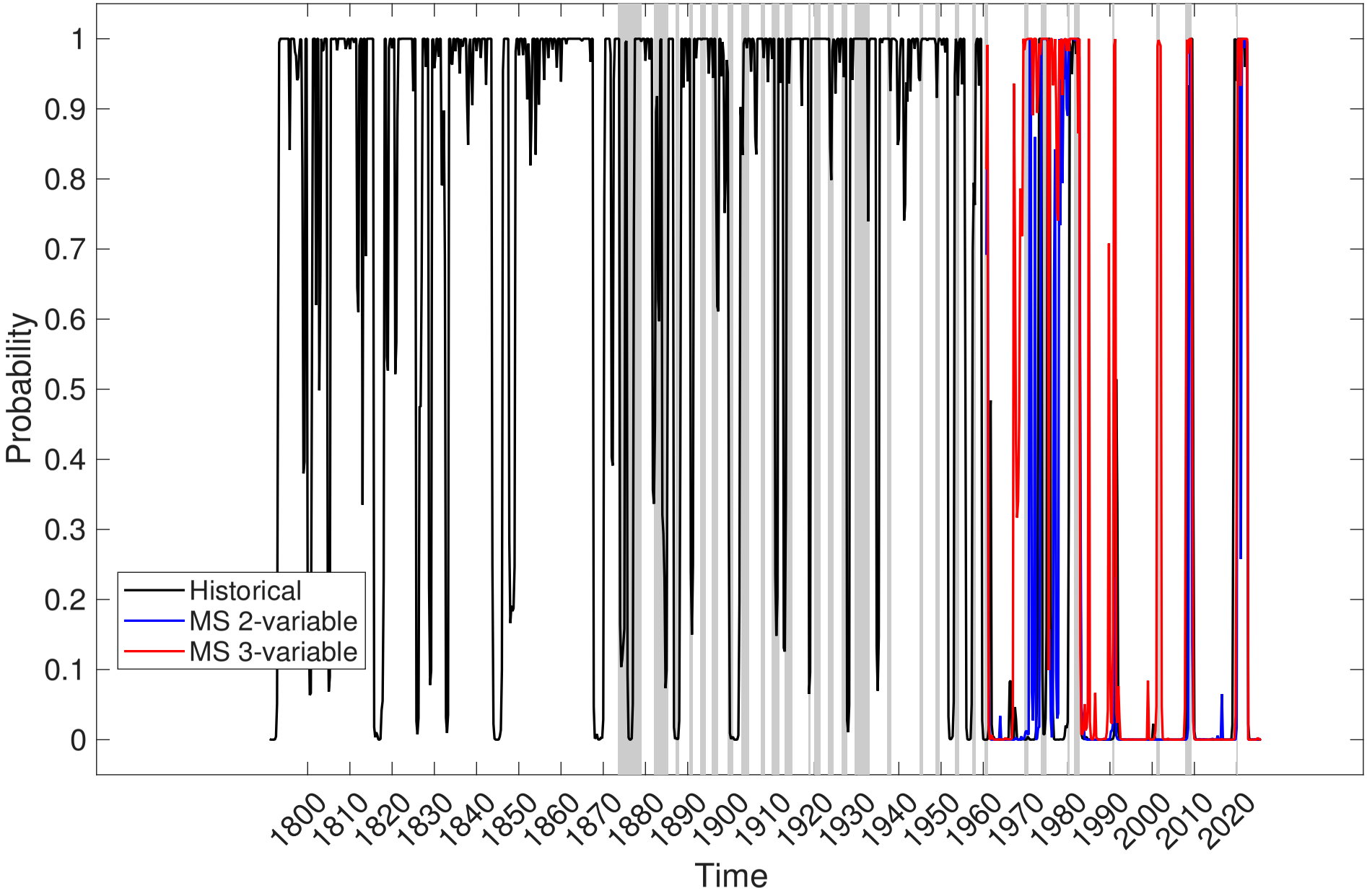


Table A.3. Regression results of JLN on alternative simple uncertainty indices using PC

| | Simple index based on | | |
|---|---|---|---|
| Period: 1960:4-2025:4 | MS-UM | Recession dummy | GDP realized volatility |
| Number of factors | 2 | 3 | 2 |
| Percentage of explained variance by PC | 91.9178 | 96.1338 | 94.2114 |
| Constant | 0.0000 | 0.0000 | -0.0000 |
| | (-0.0885, 0.0885) | (-0.0783, 0.0783) | (-0.0886, 0.0886) |
| First PC | 0.2438/*0.4734* | 0.2462/*0.4589* | 0.2415/*0.4756* |
| | (0.1979, 0.2897) | (0.2004, 0.2920) | (0.1888, 0.2941) |
| Second PC | 0.0283/*0.0032* | 0.0176/*0.0012* | -0.0010/*0.0000* |
| | (-0.0488, 0.1054) | -0.0525, 0.0877) | (-0.0791, 0.0771) |
| Third PC | | 0.3637/*0.1321* | |
| | | (0.2192, 0.5083) | |
| Adjusted R-squared | 0.4705 | 0.5858 | 0.4694 |
| Adjusted R-squared for simple OLS | 0.4413 | 0.4343 | 0.4169 |
| Correlation with JLN | 0.6675 | 0.6623 | 0.6491 |
| Std. deviation of rolling correlation | 0.1606 | 0.1439 | 0.1762 |
| | (0.1456, 0.1791) | (0.1305, 0.1605) | (0.1598, 0.1964) |

Notes: The last two rows show the standard deviation and the 95% confidence interval of the twenty-year rolling-window correlation between the JLN index and the corresponding index.

**Simple indices based on principal components**

This section presents the results of the regression analysis of JLN on alternative simple indices, which were calculated using the principal component (PC) approach described in Section 5.2. Table A.3 shows these results. Table A.3 is organized similarly to Table 1, with a few differences. First, the new second and third rows show the number of factors needed to explain at least 90% of the full set of regressors, as well as the percentage of variance these factors actually explain.

Second, the regressors are the factors obtained by the PC approach. Table A.3 clearly shows that the PC approach results in a worse fit for the JLN measure for the alternative simple indices. Thus, the adjusted R-squared statistics and the correlation coefficient with JLN are lower than those in Table 1.

## Part 5

The sources of US data considered in this paper are the following:[26]

- The real JLN uncertainty measure was downloaded from Professor Ludvigson's website on February 27, 2026.
- The EPU uncertainty measure was downloaded from the FRED dataset on October 2, 2026.
- The time series considered in the computation of the simple, small- and medium-scale uncertainty indices are the following (reference codes are written in parentheses):[27]

1. Real Gross Domestic Product (GDPC1): Billions of Chained 2017 Dollars, Quarterly, Seasonally Adjusted Annual Rate.
2. Gross Domestic Product Implicit Price Deflator (A191RI1Q225SBEA): Index 2017=100, Quarterly, Seasonally Adjusted.
3. Federal Funds Effective Rate (FEDFUNDS): Percent, Quarterly, Not Seasonally Adjusted.
4. Hourly Compensation for All Employed Persons (COMPNFB Nonfarm Business Sector): Index 2017=100, Quarterly, Seasonally Adjusted.
5. Employment Level (CE16OV ): Thousands of Persons, Quarterly, Seasonally Adjusted.
6. Population Level (CNP16OV): Thousands of Persons, Quarterly, Not Seasonally Adjusted.

[26] While an earlier version of the paper and some preliminary estimation exercises used data up to the first quarter of 2023, the current version extends the analysis to a sample period ending in the fourth quarter of 2025.

[27] This dataset was downloaded from FRED on September 10, 2026. Note that some of these time series are used to compute per capita and real variables as needed.

7. Average Weekly Hours Worked for All Employed Persons (PRS85006023, Nonfarm Business Sector): Index 2017=100, Quarterly, Seasonally Adjusted.

8. Personal Consumption Expenditures (PCEC), Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate.

9. Fixed Private Investment (FPI), Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate.

# 8 Data and program appendix

A zip file containing the data and programs used in this study is available from the authors upon request. It is possible this zip file will be put at a link connected to where the paper is eventually published.

## Contents of the Zip file

### Files for post-war WWII (1960-2025) analysis

7 Matlab program files with names 'x_uncertainty_measure_1960_2025.m', where x∈{GDP, inflation, FFrate, consumption, investment, hours, wage}. These programs generate conditional forecast error variance series for each of the data series. These are a first step program whose output is used to create 'x_uncertainty_measures_1960_2025.asc' and 'x_uncertainty_measures_1960_2025.des'.

GAUSS program '2MS2_2lags_1960-2025.gau' estimates the 2-variable MS-VAR(2) for the sample period 1960-2025. This is also a first step program which generates 'MS_smooth_uncert_1960_2025.asc' and 'MS_smooth_uncert_1960_2025.des'. These programs also contain other MS-UM obtained from alternative MS-VAR specifications, which are also considered in some robustness exercises.

Matlab program 'plot_Prob2_2var_MSVAR.m' makes Figure 1 in Section 3 of the paper.

'USdata_September_2026.asc': This is the raw data file which is loaded into the uncertainty measure and the MS switching programs. The file 'USdata_September_2026.des' contains the same time series as 'USdata_September_2026.asc', but adds a description of the time series.

7 data files 'x_uncertainty_measures_1960_2025.asc' which contain the conditional forecast error variance series obtained from 'x_uncertainty_measure_1960_2025.m'.

7 data files 'x_uncertainty_measures_1960_2025.des' which contain the conditional forecast error variance series from 'x_uncertainty_measure_1960_2025.m' along with some additional description, hence the .des extension.

Matlab program 'Uncert_Regressions_1960_2025.m': This is the final step program which reads in the 'y1_uncertainty_measures_1960_2025.asc', where y1∈{GDP, inflation} and the corresponding column of 'MS_smooth_uncert_1960_2025.asc', and carries out the statistical analysis shown in

Table 1 and makes Figure 2 in Section 4.1 of the paper. The alternative columns in Table 1 are obtained by running the Matlab program 'Uncert_Regressions_1960_2025.m' several times, with the corresponding lines (39–44) being activated and deactivated each time.

Matlab program 'Uncert_Regressions_1985_2025_EPU.m' carries out the statistical analysis corresponding to the EPU index shown in Table 2 and makes Figure 3 in Section 4.2 of the paper.

The alternative columns in Table 3 of Section 5.1 of the paper are obtained by running the Matlab program 'Uncert_Regressions_1960_2025.m' several times, with the corresponding lines (39–44) being activated and deactivated each time.

Matlab program 'Uncert_Regression_1960_2025_medium_scale.m': This is the final step program which reads in the 'y2_uncertainty_measures.asc', where y1∈{GDP, inflation, FFrate, consumption, investment, hours, wage} and the corresponding column of the 'MS_smooth_uncert_1960_2025.asc', and carries out the statistical analysis shown in Table 4 and makes Figure 4 in Section 5.1 of the paper.

**Files for historical sample (1790-2025) analysis**

2 Matlab program files with names 'x_uncertainty_measure_1790_2025.m', where x∈{GDP, inflation}. These programs generate conditional forecast error variance series for each of the historical data series. These are a first step program whose output is used to create 'x_uncertainty_measures_1790_2025.as and 'x_uncertainty_measures_1790_2025.des'.

GAUSS program '2MS2_2lags_1790-2025.gau' for the MS switching file. This is also a first step program which generates 'MS_historical_uncertainty_measures_1790_2025.asc' and 'MS_historical_uncertainty_r

'Historical_US_GDP_1790_2025_annual.asc': This is the raw historical data file constructed by Williamson (2026). The file 'Historical_US_GDP_1790_2025_annual.des' contains the same time series as 'Historical_US_GDP_1790_2025_annual.asc', but adds a description of the time series.

'Historical_US_GDP_1790_2025_quarterly.asc': This is the interpolated quarterly data obtained from the previous annual data by running the Matlab program 'Interpolating_annual_to_quarterly.m'. The file 'Historical_US_GDP_1790_2025_quarterly.des' contains the same time series as 'Historical_US_GDP_1790_2025_quarterly.asc', but adds a description of the time series.

2 data files 'x_uncertainty_measures_1790_2025.asc' which contain the conditional forecast error variance series from 'x_uncertainty_measure_1790_2025.m'.

2 data files 'x_uncertainty_measures_1790_2025.des' which contain the conditional forecast error variance series from 'x_uncertainty_measure_1790_2025.m' along with some additional description, hence the .des extension.

Matlab program 'Uncert_Regressions_1790_2025.m': This is the final step program which reads in the 'y1_uncertainty_measuresMS_smooth_uncert_1790_2025.asc', where y1∈{GDP, inflation} and the 'MS_historical_uncertainty_measures_1790_2025.asc', and carries out the statistical analysis shown in Tables 5-6, and makes Figure 5 in Section 6 of the paper. The alternative columns in Table 5 are obtained by running the Matlab program 'Uncert_Regressions_1790_2025.m' several times, with the corresponding lines (40–45) being activated and deactivated each time.

**Further comments for Matlab programs x-uncertainty_measure_1960_2025.m**

The first block of code contains alternative ARIMA forecasting models for each series. A line around line 260 selects which forecasting model to work with. The next block of code runs GARCH(1,1) on the forecast errors and computes the conditional forecast errors which are then transfered to 'x_uncertainty_measures_1960_2025.asc' and 'x_uncertainty_measures_1960_2025.des'. A similar comment applies to the files x-uncertainty_measure_1790_2025.m analyzing historical data.