Uncertainty and the Business Cycle

Freie Wissenschaftliche Arbeit
für die
Masterprüfung in Economics and Finance
an der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Eberhard Karls Universität
Tübingen

Eingereicht bei:
Prof. Dr. Thomas Dimpfl

Eingereicht von:
Oliver Buchholz

Abgabetermin: 02.11.2020

Contents

List of Tables

List of Figures

1 Introduction

What is the role of economic uncertainty and how does it interfere with macroeconomic fluctuations, that is, with the business cycle? This question has concerned economists for decades and incisive events such as the global financial crisis or the ongoing COVID-19 pandemic—to mention only the most prominent and recent examples—have revealed the relevance of finding conclusive answers. There is no doubt that uncertainty is of particular importance in economics, as it crucially affects human preferences and behavior. However, scientific knowledge about the sources and effects of uncertainty within the economy is not in line with the importance of the subject. Even worse, it diminishes as the level of aggregation increases: While, for instance, the theory of individual choice under uncertainty is fleshed out rather well and nowadays forms part of every serious textbook in microeconomic theory (e.g. Mas-Colell et al. 1995, Ch. 6), it is far from clear whether on an aggregate, macroeconomic level, business cycles are driven by uncertainty shocks or vice versa.

Thus far, theoretical work was not able to elucidate this realm and, as a consequence, it is an empirical and hence econometric task to improve the situation with evidence concerning uncertainty and the business cycle (Ludvigson et al. 2020, p. 5). Yet uncertainty is an ambiguous concept that is difficult to grasp empirically, for any empirical analysis relies on data and thus presupposes that the concepts involved can be conceived of as quantities that can be measured in one way or the other. However, a concept such as 'uncertainty' allows for a wide range of interpretations and not surprisingly, there is no consensus as to the correct way of measuring it in empirical work.

In addition to the issue of measurement, any econometric analysis that seeks to identify causal relations between macroeconomic quantities is faced with the problem of endogeneity. In a nutshell, the problem arises because aggregate economic quantities like, for instance, industrial production or employment, are likely to be driven by a myriad of other factors such that there is no truly exogenous variation in macroeconomic quantities that might be exploited for causal inference (Nakamura and Steinsson 2018, p. 59). Thus, although economics undergoing a "credibility revolution" (Angrist and Pischke 2010) due to the fact that methodological advances have been facilitating a careful empirical identification of causal effects from observational data, this revolution mostly concerns the field of applied microeconomics, while econometric tools for causal inference in macroeconomic settings are still unavailable.

The econometric challenges of measuring uncertainty and identifying true causal effects led to a situation in which the literature consists of a variety of contributions that diverge along several dimensions, be it with respect to the way in which uncertainty is measured, the method of causal inference that is applied, or, as a consequence, with respect to the findings made. Clearly, given this situation, a comparison of different results is not straightforward, since they differ in too many aspects for which one would have to control. This is the point of departure for the investigation at hand: The modest contribution I would like to achieve is to allow for a comparison of results obtained from different econometric methods for causal inference. In order to achieve this goal, I employ a strategy inspired by the *ceteris paribus* interpretation of regression analysis that is well known to any econometrician: While using different econometric methods, I leave everything else equal, that is, I rely on the same measure(s) of uncertainty and on the same macroeconomic quantities throughout all empirical analyses. This strategy enables a comparison of different econometric methods and the results they bring about and avoids the additional complication introduced by different data to which the methods are applied.

Mirroring the strategy that I just outlined, the remainder of this text is organized as follows: In the second chapter, I trace the current state of the theoretical and empirical literature on uncertainty and the business cycle. In the third chapter, I provide a theoretical overview over three different econometric methods for causal inference, namely testing for Granger-causality, structural vector autoregressive analysis, and invariant causal prediction. Then, in the fourth chapter, I apply these methods separately to a set of macroeconomic data to shed light on the causal relation between uncertainty and the business cycle. Finally, in chapter five, I conclude with a brief summary of the preceding discussion.

2 Related Literature

As stated at the beginning of this paper, research on uncertainty and its relation to business cycle fluctuations just began to intensify over the course of the last years. The contributions that exist so far can be organized along two main dimensions, the first being their research approach, that is, whether it is a theoretical or empirical investigation, and the second being their findings, that is, whether causality running from uncertainty to the business cycle or vice versa is established. It should be noted, however, that both dimensions are continuous, in the sense that there are studies that combine theoretical and empirical approaches and there are others that find evidence for both directions of causality.

Since theory often provides the starting point to form empirically testable hypotheses, it seems reasonable to review this strand of the literature first. It dates back to the work of Bernanke (1983) and Hassler (1996) who were the first authors explicitly analyzing the role of uncertainty in economic fluctuations. Both of them argue that uncertainty affects the business cycle through the investment channel: For an agent who is faced with an investment decision, higher uncertainty increases the "returns to waiting for information" (Bernanke 1983, p. 85) and hence discourages investment, thereby reducing the demand for investment goods as well as for durables at least in the short run (Hassler 1996, p. 1135).

Recent theoretical work picks up this line of argumentation according to which changes in uncertainty lead to business cycle fluctuations. The seminal paper in this context is Bloom (2009), who finds that a shock in macroeconomic uncertainty leads to a rapid drop in aggregate output and employment (Bloom 2009, p. 623). Similar to earlier contributions, Bloom (2009, p. 674) stresses the importance of investment activity that slows down whenever uncertainty is high and hence transmits changes in uncertainty to changes in overall economic activity, leading to a countercyclical behavior of uncertainty. Bloom et al. (2018, p. 1062) confirm this finding and conclude, that recessions are modelled most realistically by combining a negative first-moment and a positive second-moment shock, that is, by a decrease in the level of the series at hand and a simultaneous increase in its volatility. Furthermore, Christiano et al. (2014) find that shocks to uncertainty, to which they refer as "risk shocks", are in fact a major driver for fluctuations in output and other aggregate variables. Interestingly, however, they are the first authors who explicitly state their crucial assumption of variations in uncertainty being purely exogenous and admit that "[p]resumably, in reality there is a large endogenous component to risk shocks" (Christiano et al. 2014, p. 63). In other words, they hint at the fact that causality does not necessarily run from an exogenous shock in uncertainty to fluctuations in the business cycle, but rather the other way around, such that uncertainty is determined endogenously by changes in economic activity.

For instance, van Nieuwerburgh and Veldkamp (2006) theoretically explore learning about productivity of economic agents as one mechanism leading to endogenous uncertainty. They argue, that the aggregate technology and hence productivity of an economy is unobservable to the agents and information about it is only revealed via production: When production is high, a higher amount of information about the level of productivity is revealed, leading to more precise estimates about the true productivity than in times of low production when there is less information and, consequently, more noise in the estimates (van Nieuwerburgh and Veldkamp 2006, p. 754). Uncertainty as captured by the amount of noise in the agents' estimates is therefore determined endogenously by fluctuations in the business cycle that lead to high or low amounts of production which in turn reveal high or low amounts of information about the level of productivity in the economy.

Saijo (2017) argues in a similar way, modelling agents such that they learn about unobservable economic fundamentals, that is, the level of technology or productivity, via capital accumulation. In recessions, investment decreases and the process of capital accumulation slows down which impedes learning about fundamentals and increases uncertainty. Conversely, in a boom, investment is high, capital accumulation is accelerated and the information produced decreases uncertainty about fundamentals. Thus, uncertainty is determined endogenously and exhibits changes that behave countercyclically over the business cycle (Saijo 2017, p. 2). It should be noted, however, that the author extends this reasoning by stating that changes in uncertainty, though endogenous, amplify business cycles and give rise to a multiplier effect that clearly blurs any causal relationships (Saijo 2017, p. 21).

Finally, Fajgelbaum et al. (2017) provide another contribution to this part of the theoretical literature that emphasizes learning about economic fundamentals via production or investment as the major channel that translates variations in the business cycle into changes in uncertainty.

Another theoretical justification for endogenous uncertainty is given by Bachmann and Moscarini (2012). They explore the hypothesis that firms face imperfect information about the price elasticity of demand for their products and only learn about it by examining their sales volumes. Naturally, learning from sales volumes is facilitated by high economic activity when sales are high and create much information. Conversely, learning is impaired after a first-moment shock to the real economy that reduces the level of economic activity and hence demand. In the latter situation, firms need to “try harder to learn their demand curves” (Bachmann and Moscarini 2012, p. 2) and according to the authors they do so by varying their prices considerably, such that the sales generated are more variable and informative (ibid.). This price-experimenting behavior, however, increases the cross-sectional dispersion in price changes which the authors consider as their measure of uncertainty.

The discussion of the theoretical literature, that is continued at greater length in Fernández-Villaverde and Guerrón-Quintana (2020), reveals that there is no consensus as to the mechanism that links uncertainty and the business cycle and not even as to the direction of causality between both concepts. In fact, there are serious arguments for both directions that need to be considered, either putting weight on the return to waiting that *increases* or on the ease of learning about fundamentals that *decreases* whenever uncertainty is on the rise. This insight leads Ludvigson et al. (2020, p. 5) to the conclusion that

> “[p]ut simply, the body of theoretical work does not provide precise identifying restrictions for empirical work. Instead, what the literature presents is a wide range of theoretical predictions about the relationship between uncertainty and real economic activity that are also ambiguous about the sign of the relationship. The absence of a theoretical consensus on this relationship, along with the sheer number of theories and limited body of evidence on the structural elements of specific models, underscores the extent to which the question of cause and effect is fundamentally an empirical one.”

The strand of the literature that undertakes this empirical investigation is closest to my approach taken in the paper at hand, which is why I will discuss it below. Yet, to allow for an empirical investigation, data is essential in the first place and although this may sound like a platitude in general, it is a non-trivial issue in the context of uncertainty—in fact, as mentioned above, there are several contributions that propose different ways to measure uncertainty. The article by Bloom (2014) and the recent working paper by Baker et al. (2020) provide a comprehensive overview on that literature. For instance, Baker et al. (2020, pp. 3) mention the following measures for economic uncertainty: Stock market volatility as captured by either implied or realized volatility, newspaper-based measures, business expectation surveys, forecaster disagreement and statistical forecast uncertainty. Newspaper-based measures are developed by Baker et al. (2016) for the case of economic policy uncertainty (EPU) and by Baker et al. (2019) for the case of equity market volatility (EMV). Both of them rely on the idea to capture uncertainty by an index that is based on the frequency with which articles that contain certain keywords appear in leading newspapers. I will discuss these indices in greater detail in section 4.1 below.

Jurado et al. (2015), on the other hand, question the use of both stock market volatility and newspaper-based indices, since, according to their view, they represent proxies for uncertainty that might exhibit a rather low correlation with the unobserved process that generates true uncertainty. Hence, they propose another measure of uncertainty that relies on the notion of predictability: Uncertainty, they argue, is not about volatility *per se*, but about the question "whether the economy has become more or less *predictable*" (Jurado et al. 2015, p. 1178). This is why, for their measure, the authors first compute forecasts for the series at hand which they remove subsequently from the series to get rid of all predictable components and consider the remaining, inherently unpredictable volatility only afterwards.

Now, what does empirical research make of the theoretical results and the data? The survey of this literature reveals that—not very surprisingly—both directions of causality, that is, some measure of uncertainty influencing the business cycle and vice versa, could already be identified. I provide an overview of the empirical literature in table 1, where each line corresponds to one contribution to the literature, the first column reports the authors and the year of that given contribution, the second column indicates the way in which uncertainty is measured, the third column contains information regarding the econometric methodology that the authors employ and the last two columns indicate by means of ticks (✓), whether evidence for a causal effect running from the business cycle to uncertainty (BC $\Rightarrow$ UC) or vice versa is identified.

The table reveals that the strongest evidence seems to exist for causality running from uncertainty to fluctuations in the business cycle, since all but one article report this finding (table 1, last column). This is an important insight, especially because the result seems robust to a variety of methodological approaches and different measures of uncertainty. For instance, Baker and Bloom (2013) consider exogenous events such as natural disasters or terrorist attacks in the context of a panel regression with instrumental variables, while Baker et al. (2020) additionally use realized and implied stock market volatility, EPU and EMV indices as well as business expectation surveys within the same modelling framework. Baker et al.

(2016), Berger et al. (2020) and Jurado et al. (2015), on the other hand, employ SVAR models along with the popular yet problematic recursive identification scheme, using realized and implied volatility as measured by the Chicago Board Options Exchange's VIX or their own index as a measure of uncertainty. A slight modification with the same finding is given in Carriero et al. (2020), who use Bayesian methods for estimation and exploit heteroskedastic data to achieve identification of their SVAR model.[1]

Apart from contributions that find a causal effect of uncertainty on the business cycle, there are others that do not reach such a clear conclusion, but rather find evidence for both directions of causality instead. Examples are the articles by Bachmann et al. (2013), at the same time one of the first empirical investigations of the topic, Cesa-Bianchi et al. (2019) and Ludvigson et al. (2020). Finally, there is only the article by Carriero et al. (2018), appropriately entitled "Endogenous Uncertainty", that reports an effect of macroeconomic dynamics on uncertainty (table 1, penultimate column).

As a consequence, it is fair to say that the empirical literature on uncertainty and the business cycle does not reach any conclusive results or even a consensus. As the discussion of this section shows, it conforms with the theoretical literature in this regard, although, clearly, this kind of conformity cannot be considered satisfactory. Consequently, the only consensus stressed by several authors seems to be the insight that more empirical research is required, or, put differently, "more empirical work on the effects of uncertainty would be valuable, particularly work which can identify clear causal relationships" (Bloom 2014, p. 168). This is the aim of the following sections.

[1] For a thorough treatment of SVAR models identified by heteroskedasticity that is beyond the scope of this paper, the reader is referred to Kilian and Lütkepohl (2018, ch. 14).

Table 1: Related empirical literature on the causal relationship between uncertainty and business cycles

Reference	Measure of UC	Methodology	Result	
			BC ⇒ UC	UC ⇒ BC
Bachmann et al. (2013)	Cross-sectional dispersion of ex post survey forecast errors	SVAR with recursive identification	✓	✓
Baker and Bloom (2013)	Stock market volatility, series of exogenous shocks (natural disasters, terrorist attacks, . . .)	Panel IV regression		✓
Baker et al. (2016)	EPU, VIX	Panel SVAR with recursive identification		✓
Baker et al. (2020)	Stock market volatility, EPU/EMV, business expectation surveys	Panel regression following Baker and Bloom (2013)		✓
Berger et al. (2020)	(Realized/Implied) Stock market volatility	SVAR with recursive identification		✓
Carriero et al. (2018)	VIX	SVAR identified by heteroskedasticity	✓	
Carriero et al. (2020)	BVAR-SV estimates	BVAR identified by heteroskedasticity		✓
Cesa-Bianchi et al. (2019)	Realized stock market volatility	Panel VAR	✓	✓
Jurado et al. (2015)	Construction of own index	SVAR with recursive identification		✓
Ludvigson et al. (2020)	Indices constructed following Jurado et al. (2015)	SVAR identified using shock-based constraints	✓	✓

Notes: "Uncertainty" is abbreviated as UC and "Business Cycle" as BC. The arrow "⇒" indicates the direction of causality that the authors of a given article identify.

3 Methods of Causal Inference in Macroeconometrics

Science, in particular when it is policy-relevant as is the case for most fields of economics, attempts to come up with answers to "what if"-questions: What would happen to the investment behavior if the tax rate were lowered? What would be the reaction of economic growth if monetary policy were tightened? And, finally, what would be the dynamic of uncertainty if economic activity were on the rise? The scientific gold standard to answer these kind of counterfactual questions is the so-called randomized controlled trial, an experiment that is carefully designed in advance and takes place in a well-controlled—hence its name—environment, free of any impacts that might alter the effect under investigation (Runge et al. 2019, p. 1).

Clearly, however, conducting experiments to answer macroeconomic questions like the ones I just mentioned is often infeasible, either because it is unethical or too costly (Pfister et al. 2019, p. 1264). As a consequence, methods of causal inference have to be employed that allow to detect causal relations from observational data. Although such methods have been developed, the detection of true causal effects within the overall economy remains a challenging task, since "identification in macroeconomics is difficult" (Nakamura and Steinsson 2018, p. 59). This means that it is difficult to come up with truly random and exogenous variation in a given quantity of interest. Consider, for instance, economic uncertainty: As we have seen in the section above, it might be driven by variation in the business cycle, but it might as well be the case that the news exhibit a non-negligible impact on the overall uncertainty—after all, that is the idea behind news-based uncertainty indices like the EPU whose details are discussed below. Yet the news themselves might also influence the business cycle thus acting as a hidden confounder that makes business cycle variation endogenous and, consequently, of little usefulness to explain economic uncertainty.

In this section, I review three econometric approaches that try to tackle the ambitious task of causal inference in macroeconomics from very different perspectives. I begin with the historically oldest methodology, that is, with the well-known notion of causality that was coined by Granger (1969). Next, I provide an outline of structural vector autoregressive models that allow for a different kind of causal inference and finally, I present the method of ICP that was originally developed by Peters et al. (2016) and extended to the case of multivariate time series by Pfister et al. (2019).

Since all approaches presented in this section rely on the framework of multiple time series, several pieces of notation arise repeatedly in the following discussion, which is why I mention them here to enhance notational efficiency and readability: If A is an arbitrary matrix, $a_{k\ell}$ is its (k, ℓ)-element, A' denotes its transpose, and $\det(A)$ its determinant. If, in addition, A is non-singular, that is, if $\det(A) \neq 0$, then A^{-1} is its inverse. If B is another matrix, then the Kronecker product between A and B is denoted $A \otimes B$. The n-dimensional identity matrix is denoted by I_n.

3.1 Granger-Causality

As mentioned above, Granger (1969) was the first to propose a notion of causality within the realm of time series econometrics. Nowadays, it can be considered the most prominent and, equivalently, the most controversial method of causal inference, serving as a reference point that enabled researchers to come up with refinements through criticizing Granger's initial method (for examples see Kilian and Lütkepohl 2018, Ch. 7; Pfister et al. 2019, p. 1264). The basic idea of Granger-causality is this: "[A] cause cannot come after the effect" (Lütkepohl 2005, p. 41). In other words, if a variable $y_{2,t}$ exhibits an impact on another variable $y_{1,t}$, the former should help forecasting the latter. As a consequence, a variable $y_{2,t}$ is said to *Granger-cause* another variable $y_{1,t}$, if the information contained in past values of the former help reducing the prediction error made for the latter (ibid.). Clearly, this means that Granger-causality centers around the predictive relationship between the variables of interest, or, in Granger's (1969, p. 430) own words, the "definition of causality [...] is based entirely on the predictability of some series". Yet, as the saying goes, correlation does not imply causation and hence, mere predictability need not necessarily be related to causal effects between the variables of interest. In fact, this is the aspect that is used most frequently in the literature to question the helpfulness of Granger-causality for causal inference (Lütkepohl 2005, p. 48; Kilian and Lütkepohl 2018, p. 49).

Let us now proceed by formalizing the notion of Granger-causality in the context of a VAR model and, subsequently, the way in which it can be assessed empirically. In order to do so, fix the number of variables $K = 2$ without loss of generality and assume that the two variables are generated by a bivariate VAR(p) process,

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} + \sum_{i=1}^{p} \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} \begin{pmatrix} y_{1,t-i} \\ y_{2,t-i} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \tag{1}$$

$$y_t = \nu + \sum_{i=1}^{p} A_i y_{t-i} + u_t \tag{2}$$

where ν_1, ν_2 are intercept terms, $a_{k\ell,i}$ are reduced-form coefficients and $u_{1,t}, u_{2,t}$ are white noise reduced-form errors, such that $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$ and $E(u_t u_s') = 0$ for $s \neq t$ (Lütkepohl 2005, p. 13). As is the case for all VAR(p) processes, one can also express the process in (1) in the so-called *companion form*, that is, as a VAR(1)

$$Y_t = \boldsymbol{\nu} + \boldsymbol{A} Y_{t-1} + U_t \tag{3}$$

$$\underset{(2p\times 1)}{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}} = \underset{(2p\times 1)}{\begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}} + \underset{(2p\times 2p)}{\begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_2 & 0 & \cdots & 0 & 0 \\ 0 & I_2 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_2 & 0 \end{bmatrix}} \underset{(2p\times 1)}{\begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix}} + \underset{(2p\times 1)}{\begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \tag{4}$$

where $\boldsymbol{A}$ is the so-called *companion matrix* of the VAR(p) process (Kilian and Lütkepohl 2018, p. 25). Beyond the obvious reason of conciseness, representing a VAR(p) process as a VAR(1) highly facilitates to assess the processes' *stability*. This is because the VAR(p) in (1) is stable if

$$\det(I_{2p} - \boldsymbol{A} z) \neq 0 \quad \forall z \in \mathbb{C},\ |z| \leq 1, \tag{5}$$

which amounts to the condition that all eigenvalues of the companion matrix $\boldsymbol{A}$ must have modulus less than 1 (ibid.). The stability condition mainly gains it relevance from the result that "stability implies stationarity" (Lütkepohl 2005, p. 25, in particular Prop. 2.1), which guarantees that a stable VAR(p) process is also *stationary*, that is, its first and second moments are time invariant. As a consequence, it is straightforward to assess the stability and stationarity of a VAR model by analyzing the eigenvalues of the companion matrix. For the remainder of this section, I will assume stability and stationarity of processes as it facilitates the following exposition of Granger-causality tests. In the empirical part below, I will test whether the assumption is fulfilled making use of condition (5) in order to check whether the methodology can be applied properly.

As mentioned above, Granger-causality centers around the predictability of one series through information contained in another one. Thus, consider a setting as the one presented in Kilian and Lütkepohl (2018, pp. 48), where forecasts of $y_{1,t}$ are computed from the VAR(p) process in (3). In this setting, let Ω_t be the information set consisting of information relevant to forecasting $y_{1,t}$, that is, information on $y_{1,t}$ and $y_{2,t}$ up to and including period t. Furthermore, denote the optimal h-step forecast of $y_{1,t}$, based on the information set Ω_t, by $y_{1,t+h|\Omega_t}$. In this context, 'optimal' is interpreted such that the forecast, $y_{1,t+h|\Omega_t}$, minimizes the *Mean Squared Prediction Error* (MSPE) in the sense that

$$MSPE(y_{1,t+h|\Omega_t}) \equiv E(y_{t+h} - y_{1,t+h|\Omega_t})^2 \leq MSPE(\bar{y}_{1,t+h|\Omega_t}) \qquad (6)$$

for any other forecast $\bar{y}_{1,t+h|\Omega_t} \neq y_{1,t+h|\Omega_t}$. Now that we have defined the overall VAR framework as well as the prediction framework that I just outlined, we are almost in the position to formally state the notion of Granger-causality. In order to make the last step, define $MSPE(y_{1,t+h|\Omega_t}) := \sigma^2_{y_1}(h|\Omega_t)$ and let $\Omega_t \setminus \{y_{2,s}|s \leq t\}$ denote the set of all relevant information except past and present information about the $y_{2,t}$ process. Then, according to the definition of Granger-causality, $y_{2,t}$ Granger-causes $y_{1,t}$ if

$$\exists\, h \in \{1, 2, \dots\} \colon \sigma^2_{y_1}(h|\Omega_t) < \sigma^2_{y_1}(h|\Omega_t \setminus \{y_{2,s}|s \leq t\}). \qquad (7)$$

Expression (7) clearly shows why the concept of Granger-causality relies on predictive relationships between different variables, for the definition states that $y_{2,t}$ Granger-causes $y_{1,t}$ if a lower MSPE in forecasting the latter variable can be achieved once the information set $\Omega_t \setminus \{y_{2,s}|s \leq t\}$ is augmented by past and present information on the former variable.

From an empirical point of view, the next question that needs to be addressed concerns the way in which the definition of Granger-causality given in (7) can be tested. To answer the question, recall the bivariate VAR(p) model in (1). Intuitively, to assess whether $y_{2,t}$ Granger-causes $y_{1,t}$ and thus, whether information about $y_{2,t}$ helps forecasting $y_{1,t}$, the corresponding coefficients in the first line of the VAR are crucial: If all coefficients for the lags of $y_{2,t}$ equal zero, they do not seem to possess predictive power for realizations of $y_{1,t}$ and $y_{2,t}$ is not Granger-causal for $y_{1,t}$. If, however, lags of $y_{2,t}$ appear with nonzero coefficients, the situation is reversed and $y_{2,t}$ is indeed Granger-causal for $y_{1,t}$ (Kilian and Lütkepohl 2018, p. 49). As mentioned above, I assume y_t to be a stable VAR(p) process with nonsingular white noise reduced-form covariance matrix Σ_u. This assumption entails that the conditions of Corollary 2.2.1 in Lütkepohl (2005, p. 45) are fulfilled and that, as a consequence, the crucial condition for empirically testing Granger-causality can be formalized as follows: $y_{2,t}$ is *not* Granger-causal for $y_{1,t}$, or, using the definition in (7), $\sigma^2_{y_1}(h|\Omega_t) = \sigma^2_{y_1}(h|\Omega_t \setminus \{y_{2,s}|s \leq t\})$, $h = 1, 2, \ldots,$ if and only if $a_{12,i} = 0$, $i = 1, \ldots, p$. Two aspects are worth mentioning here: First, note that the condition is formulated in terms of Granger-*non*causality rather than Granger-causality. This is due to the methodological insight that empirical evidence can in no way be strong enough to verify or at least confirm a given hypothesis, but is at best able to reject and thus *falsify* the hypothesis.[2] Since the ultimate goal is to identify Granger-causal relations between different variables, this can only be achieved by formulating the hypothesis conversely and rejecting it afterwards. As for the second aspect, it becomes obvious, that the condition for Granger-(non)causality embraces the intuition from above that predictive relevance and hence Granger-causality must be related to the VAR coefficients that indicate the relation between both variables. Furthermore, it thereby provides a natural starting point to test zero restrictions on the coefficients of estimated VAR models.

Lütkepohl (2005, pp. 102) proposes to test multiple restrictions on the LS-estimates of VAR coefficients by means of standard Wald tests. To be precise, the test consists of the hypotheses

$$H_0\colon C\boldsymbol{\beta} = c \quad \text{against} \quad H_1\colon C\boldsymbol{\beta} \neq c, \tag{8}$$

where, following the notation in Lütkepohl (2005, pp. 70 and 102),

[2] This line of argumentation that stems from the philosophy of science can at least be traced back to Karl Popper who coined the idea of *falsificationism*, but is already foreshadowed in Hume's famous problem of induction.

$$\boldsymbol{\beta} := \text{vec}\left(\underset{(2\times(2p+1))}{\begin{bmatrix} \nu_1 & a_{11,1} & a_{12,1} & \cdots & a_{11,p} & a_{12,p} \\ \nu_2 & a_{21,1} & a_{22,1} & \cdots & a_{21,p} & a_{22,p} \end{bmatrix}} \right) \tag{9}$$

$$= \underset{((4p+2)\times 1)}{\begin{bmatrix} \nu_1 & \nu_2 & a_{11,1} & a_{21,1} & a_{12,1} & a_{22,1} & \cdots & a_{11,p} & a_{21,p} & a_{12,p} & a_{22,p} \end{bmatrix}'}, \tag{10}$$

C is a $(N \times (4p+2))$ matrix of rank N, c is a $(N \times 1)$ vector and N denotes the number of restrictions to be tested. For instance, suppose we wanted to test Granger-noncausality from $y_{2,t}$ to $y_{1,t}$ in the context of the VAR(p) model in (1). In that case, we would have to test whether $a_{12,i} = 0$, $i = 1, \ldots, p$ and thus $N = p$ joint restrictions. This implies

$$C = \underset{(p\times(4p+2))}{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}}, \quad c = \underset{(p\times 1)}{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}, \tag{11}$$

which could be used directly in the hypotheses in (8). As a test statistic for the Granger-causality test, Lütkepohl (2005, p. 103) proposes the following quantity,

$$\lambda_F = \hat{\boldsymbol{\beta}}' C' \left[C((ZZ')^{-1} \otimes \widehat{\Sigma}_u) C' \right]^{-1} C\hat{\boldsymbol{\beta}}/N \tag{12}$$

with

$$Z := \underset{((2p+1)\times T)}{\begin{bmatrix} 1 & \cdots & 1 \\ y_{1,0} & \cdots & y_{1,T-1} \\ y_{2,0} & \cdots & y_{2,T-1} \\ y_{1,-1} & \cdots & y_{1,T-2} \\ y_{2,-1} & \cdots & y_{2,T-2} \\ \vdots & \ddots & \vdots \\ y_{1,-p+1} & \cdots & y_{1,T-p} \\ y_{2,-p+1} & \cdots & y_{2,T-p} \end{bmatrix}}, \tag{13}$$

where $\hat{\boldsymbol{\beta}}$ denotes the LS-estimator of $\boldsymbol{\beta}$ and $\widehat{\Sigma}_u$ denotes the LS-estimator of Σ_u. Under the null hypothesis H_0, the test statistic approximately follows an F-distribution with N—the number of restrictions—as the numerator degrees of freedom and $T - 2p - 1$ as the denominator degrees of freedom, that is, $\lambda_F \approx F(N, T - 2p - 1)$ with T denoting the number of observations (ibid.). As for all Wald-type tests, the null hypothesis H_0 can be rejected at some predefined level of significance, if the test statistic λ_F exceeds the critical value obtained from the $F(N, T - 2p - 1)$-distribution. Conversely, if the value of the test statistic remains below the critical value, H_0 is accepted (Lütkepohl 2005, p. 103).

As we have seen, the concept of Granger-causality might be considered somewhat misleading, for the preceding discussion revealed that it relies on the notion of prediction rather than causation. Consequently, it is only able to grasp mere correlation patterns between variables instead of actual causality, while neglecting instantaneous effects altogether (Kilian and Lütkepohl 2018, p. 50). Another shortcoming of Granger-causality is the fact that the relations between variables, once identified, are highly sensitive to changes in the information set. For instance, they might break down if another variable is included or removed (Kilian and Lütkepohl 2018, p. 201). These—severe—issue notwithstanding, Granger-causality remains a valuable—and, in fact, widely used—econometric tool that allows to assess dynamic relationships in economic time series. I regard it as a benchmark against which the performance of other methods can be evaluated.

3.2 Structural Vector Autoregressive Models

The preceding discussion of Granger-causality evolved around VAR models in their so-called *reduced form*, for instance in expressions (1) through (4). A reduced-form VAR model might be interpreted as an approximation of some underlying VAR DGP and can prove useful for forecasting (Kilian and Lütkepohl 2018, p. 2). It does not, however, allow for an investigation of instantaneous relationships between different variables and is in particular unable to model structural properties of the economy. This is the purpose of *structural vector autoregressive* (SVAR) models, that offer an alternative to Granger-causality analysis when it comes to conducting causal inference in macroeconometrics.

The starting point of structural vector autoregressive analysis is the premise that any reduced-form VAR(p) model—for instance the one presented in equation (1) above—might be conceived of as representing data that was generated by a structural VAR(p) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t \tag{14}$$

$$= \sum_{i=1}^{p} B_i y_{t-i} + w_t, \tag{15}$$

where y_t, $t = 1, \ldots, T$, is a possibly K-dimensional vector of observed time series data, the intercept terms are omitted for notational convenience, B_i, $i = 1, \ldots, p$, are $(K \times K)$ matrices of autoregressive slope coefficients and w_t is a $(K \times 1)$ vector of mean zero and serially uncorrelated structural shocks with variance-covariance matrix Σ_w (Kilian and Lütkepohl 2018, p. 2). Usually, Σ_w is normalized such that

$$E(w_t w_t') \equiv \Sigma_w = I_K, \tag{16}$$

which implies that Σ_w is diagonal and of full rank and the number of shocks is equivalent to the number K of variables in the model (Kilian and Lütkepohl 2018, pp. 213). Several aspects are interesting about this setup: First, note that by means of the non-singular matrix B_0, it is possible to model and analyze instantaneous effects between the variables under investigation. Second, the representation is truly structural in the sense that the elements of w_t are mutually uncorrelated and hence, each element has the interpretation of a shock that causes movements in the data. Put differently, this means that each element in w_t has—or, at least, ought to have—a distinct economic interpretation (Kilian and Lütkepohl 2018, pp. 2). Clearly, in empirical work, one is interested in the structural shocks and their impact on the variables in the model. Yet, in general, they are not directly observable and need to be recovered from the reduced-form representation of the structural VAR, for it is not possible to estimate the latter due to the matrix B_0 of instantaneous effects on the left-hand side.

To move from the structural to the reduced-form VAR representation that is needed for estimation and to recover the structural shocks, the vector y_t needs to be expressed as a function of lags of y_t only. To achieve this aim, both sides of (14) are premultiplied by B_0^{-1} which yields

$$B_0^{-1} B_0 y_t = B_0^{-1} B_1 y_{t-1} + \cdots + B_0^{-1} B_p y_{t-p} + B_0^{-1} w_t \tag{17}$$

$$\Leftrightarrow \quad y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{18}$$

such that $A_i = B_0^{-1}B_i$, $i = 1, \ldots, p$, and the reduced-form errors are a weighted average of the structural shocks, $u_t = B_0^{-1}w_t$ (Kilian and Lütkepohl 2018, p. 214). The discussion of Granger-causality above revealed that a reduced-form VAR may be estimated consistently by standard methods such as least squares. The crucial question in structural vector autoregressive analysis is, how to recover the matrix B_0 and thus, the structural shocks from the reduced-form estimates. In fact, the relevance of the issue cannot be overstated and a myriad of different ways to approach it have been proposed in the literature.[3] The key insight and point of departure for most approaches is this: By construction of the setup above, we have $u_t = B_0^{-1}w_t$. For the variance of u_t, this implies that

$$
\begin{aligned}
\Sigma_u = E(u_t u_t') &= E\left(B_0^{-1}w_t \left(B_0^{-1}w_t\right)'\right) && (19)\\
&= E\left(B_0^{-1}w_t w_t' B_0'^{-1}\right) && (20)\\
&= B_0^{-1}E\left(w_t w_t'\right)B_0'^{-1} && (21)\\
&= B_0^{-1}\underbrace{\Sigma_w}_{:=I_K}B_0'^{-1} && (22)\\
&= B_0^{-1}B_0'^{-1}. && (23)
\end{aligned}
$$

As a consequence, the variance-covariance matrix of the reduced-form errors, Σ_u, can be conceived of as a system of nonlinear equations in the unknown parameters of B_0^{-1} (ibid.). Due to the general symmetry of variance-covariance matrices, this system consists of $K(K+1)/2$ independent equations and the elements of B_0^{-1} might be recovered from it by imposing constraints on selected elements of B_0^{-1} (Kilian and Lütkepohl 2018, p. 215). It is in the motivation of these additional constraints, where the different approaches that exist to recover the structural shocks from reduced-form estimates diverge. Since the structural shocks of a given model come with an economic interpretation, many approaches proceed by imposing constraints that are motivated by insights from economic theory about the behavior of certain quantities in the short or long run. However, the above review of the literature on uncertainty and the business cycle revealed that there is no consensus as to the short and long run dynamics of both variables, which is why I will leave aside approaches that try to ground identifying restrictions in clear-cut theoretical assumptions.

[3] The excellent monograph by Kilian and Lütkepohl (2018) provides a thorough treatment of the most common methodologies.

The approach that I present in the following and employ in the empirical part of the text below relies on what is known as *recursive identification* instead (Kilian and Lütkepohl 2018, Ch. 8.2). The key idea is this: Since the structural shocks are mutually uncorrelated by definition, one can obtain them from the reduced-form errors by *orthogonalizing* the latter. This is accomplished as follows: One defines $\Sigma_u := PP'$, where P is a $(K \times K)$ lower-triangular matrix with a positive diagonal that is known as the lower-triangular Cholesky decomposition of Σ_u (Kilian and Lütkepohl 2018, p. 216). Thus, taking together this definition of Σ_u and expression (23), $B_0^{-1} = P$ is a possible solution to recover the structural shocks w_t from the reduced-form VAR, in particular, because the lower-triangular structure of P implies a number of $K(K-1)/2$ zero parameters which guarantees that B_0^{-1} is indeed exactly identified (ibid.). As easy and straightforward this way of achieving identification may seem, there is an important issue that must not go umentioned. Imposing a lower-triangular structure on B_0^{-1} via P entails a lower-triangular structure for B_0. This means that any structural model whose shocks are recovered through a Cholesky decomposition is recursive in the sense that a particular causal ordering of the variables under investigation is *defined* right from the beginning rather than *learned* from the data. Consequently, a recursive identification is only reasonable if the recursive structure of the resulting model can be justified on economic grounds (Kilian and Lütkepohl 2018, pp. 261). In the empirical part of the text, I will discuss how this justification might be achieved in the context at hand and I will argue why this identification scheme is better suited than others that are based on either short run, long run, or sign restrictions.

For the time being, note that in order to perform any kind of causal inference, we require some way of representing the elements of the structural model, for instance, the structural shocks w_t, once they are recovered from the reduced-form VAR. Arguably the most popular and illustrative tool to do so is the so-called *impulse response function* (IRF). Their aim is to capture the responses of each element of $y_t = (y_{1t}, \ldots, y_{Kt})'$ to a one-time shock in $w_t = (w_{1t}, \ldots, w_{Kt})'$ over time such that

$$\frac{\partial y_{t+i}}{\partial w_t'} = \Theta_i, \quad i = 1, \ldots, H, \tag{24}$$

where Θ_i is a $(K \times K)$ matrix and H is the maximum propagation horizon of the shocks (Kilian and Lütkepohl 2018, p. 108). Consequently, for some fixed index i, the elements of matrix Θ_i are given by

$$\theta_{jk,i} = \frac{\partial y_{j,t+i}}{\partial w'_{kt}}, \tag{25}$$

and a straightforward point of departure in order to determine them are the responses of y_{t+1} to the reduced-form errors u_t which are most easily obtained by considering the reduced form of a VAR(p) once more in its companion form, that is, in its VAR(1) representation as already shown in equation (3) above. Thus, transforming the VAR(p) in (18) to its VAR(1) representation yields

$$Y_t = \boldsymbol{A}Y_{t-1} + U_t, \tag{26}$$

where all objects are as defined in (4) except that possibly $K \neq 2$ (ibid.). Next, by successive substitution for Y_{t-i}, this representation can be transformed to

$$Y_{t+i} = \boldsymbol{A}^{t+i}Y_{t-1} + \sum_{j=0}^{i} \boldsymbol{A}^j U_{t+i-j}, \tag{27}$$

and left-multiplication of this expression with $J \equiv [I_K, \mathbf{0}_{K\times K(p-1)}]$ yields

$$y_{t+i} = J\boldsymbol{A}^{t+i}Y_{t-1} + \sum_{j=0}^{i} J\boldsymbol{A}^j U_{t+i-j} \tag{28}$$

$$= J\boldsymbol{A}^{t+i}Y_{t-1} + \sum_{j=0}^{i} J\boldsymbol{A}^j J'JU_{t+i-j} \tag{29}$$

$$= J\boldsymbol{A}^{t+i}Y_{t-1} + \sum_{j=0}^{i} J\boldsymbol{A}^j J'u_{t+i-j}. \tag{30}$$

The last line reveals the convenience of employing the VAR(1) representation for the impulse-response analysis, because the response of some variable $j = 1, \ldots, K$ in the original VAR(p) to a unit shock in the reduced-form error u_{kt}, $k = 1, \ldots, K$, i periods ago is obtained in a straightforward manner by

$$\underset{(K\times K)}{\Phi_i} = [\phi_{jk,i}] \equiv J\boldsymbol{A}^i J', \tag{31}$$

where matrix J and its transpose J' can be thought of "picking out" the first $(K \times K)$ elements of companion matrix $\boldsymbol{A}^i$. Since we just derived the responses of variables in the VAR to reduced-form shocks, one might equivalently state that we derived the *reduced-form* impulse responses (Kilian and Lütkepohl 2018, p. 109). A small step has to be taken to finally arrive at the *structural* impulse responses that we were initially interested in. In order to achieve this step, recall that we assume covariance stationarity of all processes throughout this text. Given this assumption, the process $y_t = (y_{1t}, \ldots, y_{Kt})'$ can be expressed in its moving average (MA) representation (ibid.):

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \sum_{i=0}^{\infty} \Phi_i B_0^{-1} B_0 u_{t-i} = \sum_{i=0}^{\infty} \Theta_i w_{t-i}, \tag{32}$$

where the last equality follows from the fact introduced above that $w_t = B_0 u_t$ and by defining $\Theta_i := \Phi_i B_0^{-1}$. Obviously, the MA representation allows to derive the structural impulse responses as given in (24), since

$$\frac{\partial y_t}{\partial w'_{t-i}} = \frac{\partial y_{t+i}}{\partial w'_t} = \Theta_i. \tag{33}$$

Thus, the last question regarding the determination of structural impulse responses must concern the actual computation of Θ_i for any horizon $i = 1, \ldots, H$. In fact, the computation is brought about by a simple procedure that exploits the definition $\Theta_i := \Phi_i B_0^{-1}$ (ibid.):

$$\Theta_0 = \Phi_0 B_0^{-1} = I_K B_0^{-1} = B_0^{-1} \tag{34}$$
$$\Theta_1 = \Phi_1 B_0^{-1} \tag{35}$$
$$\Theta_2 = \Phi_2 B_0^{-1} \tag{36}$$
$$\vdots$$

So, in summary, the computation of structural impulse responses relies on the companion-form representation of a reduced-form VAR as well as its transformation to the MA representation. Clearly, to obtain structural impulse responses in practice, all unknown objects introduced above have to replaced by consistent reduced-form estimates such that in the end, $\widehat{\Phi}_i$ and $\widehat{\Theta}_i$ can be computed (Kilian and Lütkepohl 2018, p. 109).

This, however, raises another final issue in the context of SVAR models and structural IRFs: Since the latter are obtained from reduced-form estimates, they are estimates and associated with some degree of uncertainty as well, which makes it necessary to perform some kind of inference. The standard procedure consists in computing confidence intervals for the structural IRFs, in most cases using some bootstrapping procedure that can be applied flexibly, requires a minimum of distributional assumptions, and yields accurate results even in small samples (Kilian and Lütkepohl 2018, p. 335). The most common bootstrapping procedure is the residual-based wild bootstrap proposed by Gonçalves and Kilian (2004). It proceeds as follows: First, the reduced-form VAR(p), for instance in (18), is estimated via multivariate LS to obtain estimates $\hat{A}_i$, $i = 1, \ldots, p$ and the reduced-form residuals $\hat{u}_t$. Second—and this is the central idea of this particular procedure—bootstrap residuals $u_t^{*r} = (u_{1t}^{*r}, \ldots, u_{Kt}^{*r})'$ are generated according to $u_t^{*r} = \hat{u}_t \eta_t$ with $\eta_t \overset{iid}{\sim} (0, 1)$ (Gonçalves and Kilian 2004, p. 96). The crucial point is that the bootstrap residuals are not sampled from some predefined distribution or from the reduced-form residuals as in other (non-)parametric bootstrapping procedures, but are instead obtained by multiplying each element of the reduced-form residual vector $\hat{u}_t$ by a scalar draw, η_t, from an auxiliary distribution with mean zero and variance one, for instance the standard normal distribution, $\mathcal{N}(0, 1)$ (Kilian and Lütkepohl 2018, p. 340).[4] This comes with the advantage that the reduced-form errors need not be i.i.d. as is usually assumed by other bootstrapping approaches, which is in fact an overly strong assumption, since "many monthly macroeconomic variables [...] exhibit evidence of conditional heteroskedasticity" (Gonçalves and Kilian 2004, p. 93). Now, as a third step, and after obtaining the bootstrap residuals, a pseudo time series—the bootstrap sample—$\{y_t^{*r}\}_{t=-p+1}^{T}$ is generated by adding the bootstrap residuals to the fitted model. Then, the LS estimates are updated, new reduced-form residuals are computed and the process starts from the beginning. Repeating this procedure for $r = 1, \ldots, R$ allows to construct the empirical distribution for the structural IRF $\hat{\theta}_{ik,h}^{*r}$ and, as a consequence, the computation of confidence intervals for the IRF (Kilian and Lütkepohl 2018, p. 339).

3.3 Invariant Causal Prediction

Unlike the methodologies outlined above, the formal framework of ICP was developed rather recently, in Peters et al. (2016). Subsequently, Pfister

[4] More general yet concise treatments of bootstrapping procedures in multivariate time series analysis can be found in Kilian and Lütkepohl (2018, ch. 12.2) as well as in Lütkepohl (2005, Appendix D.3).

et al. (2019) extended it to the case of sequential data, since the original method is tailored to cross-sectional data. However, the central idea of ICP has already been discussed for a considerable period of time, for instance by Haavelmo (1944, pp. 28) and Aldrich (1989). Their key insight is that, given a model that consists of a response variable and several covariates, what discerns a causal model from a non-causal one is the property of *invariance*, which means that the relations between the variables within the model are not altered once there are interventions on the covariates. Peters et al. (2016, p. 948) state this idea in a concise way by observing that "the conditional distribution of the target variable of interest [...] given the complete set of corresponding direct causal predictors, must remain identical under interventions on variables other than the target variable." Clearly, this conceptual idea establishes a link between the invariance of a model under interventions and causality, thereby opening the door for a method of causal inference that exploits this link and statistically tests a given model's invariance.

The formalization of ICP proceeds as follows: Just as Pfister et al. (2019, p. 1265), I assume to be given data from a sequence $(Y_t, X_t)_{t\in\{1,\dots,n\}}$ where the covariates are defined as a $(1 \times d)$ vector X_t and a scalar response Y_t. Response variables and covariates for all periods $t = 1, \dots, n$ can be summarized in the $(n \times 1)$ vector $\boldsymbol{Y} := (Y_1, \dots, Y_n)'$ and the $(n \times d)$ matrix $\boldsymbol{X} := (X_1', \dots, X_n')'$, which leads to the overall $(n \times (d+1))$ data matrix $(\boldsymbol{Y}, \boldsymbol{X}) = (Y_t, X_t)_{t\in\{1,\dots,n\}}$. Finally, Pfister et al. (2019, p. 1266) assume that for any set $S \subseteq \{1, \dots, d\}$, the $(1 \times |S|)$ vector X^S contains only the variables $\{X^k; k \in S\}$. A first important step now involves the formalization of *invariance* which the authors achieve by means of the following

Definition 1 (invariant set S (Pfister et. al. 2019, Definition 1))**.** *A set $S \subseteq \{1, \dots, d\}$ is called invariant with respect to $(\boldsymbol{Y}, \boldsymbol{X})$, if there exist parameters $\mu \in \mathbb{R}, \beta \in (\mathbb{R} \setminus \{0\})^{|S|\times 1}$ and $\sigma \in \mathbb{R}_{>0}$ such that*

(a) $\forall t \in \{1, \dots, n\}\colon Y_t = \mu + X_t^S \beta + \varepsilon_t$ *and* $\varepsilon_t \perp\!\!\!\perp X_t^S$

(b) $\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.

Here, the symbol $\perp\!\!\!\perp$ denotes stochastic independence. From an intuitive point of view, the definition embodies the idea that the conditional distribution of response Y_t given the covariates from S remains unchanged for all periods and for all kinds of interventions that might occur to the system, as long as they do not affect the response itself. An important assumption is that a set $S^* \subseteq \{1, \dots, d\}$ which fulfills Definition 1 indeed exists (Pfister et al. 2019, p. 1266, Assumption 1). The goal of ICP is to infer this so-called invariant set S^* that consists of the covariates that exhibit a causal influence on the response. Thus, as insinuated above, the formal framework that I just outlined gives rise to a setting in which testing for invariance equates to testing for causality (Pfister et al. 2019, pp. 1266). The crucial question then is how invariance in the sense of Definition 1 might be assessed with a statistical test.

Clearly, the point of departure needs to be a specific null hypothesis that states properties to be tested. In the case of ICP, Pfister et al. (2019, p. 1266) define the null hypothesis in its simplest form as

$$H_{0,S}\colon S \text{ is an invariant set with respect to } (\boldsymbol{Y}, \boldsymbol{X}). \tag{37}$$

This means that the test should verify whether some S fulfills Definition 1 or not. Note, that at first blush, this approach of formulating the null hypothesis seems rather counterintuitive: Above, we have seen that testing for Granger-causality in fact means testing the null hypothesis of Granger-*non*causality while hoping to reject it, for this is as far as we can get using statistical evidence. In the case of ICP, however, the null hypothesis is formulated such that invariance and thus the existence of a causal effect is tested while its rejection leads to the conclusion that there is no causal effect. In an instant, I will point out why the authors nevertheless formulated $H_{0,S}$ as shown in (37), but first, observe that there is no straightforward way to test this general hypothesis. Consequently, Pfister et al. (2019, p. 1273) make it more precise by stating that for a fixed set $S \subseteq \{1, \dots, d\}$,

$$\widetilde{H}_{0,S,p}\colon \begin{cases} \exists \eta \in \mathbb{R}^{(|S|+(d+1)p)},\ \sigma \in (0, \infty) \text{ such that } \forall t \in \{p+1, \dots, n\}\colon \\ Y_t = Z_t^{S,p}\eta + \varepsilon_t \text{ with } \varepsilon_t \perp\!\!\!\perp Z_t^{S,p} \text{ and } \varepsilon_{p+1}, \dots, \varepsilon_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \end{cases} \tag{38}$$

As before, the lag order is given by $p \in \{0, \dots, n-2\}$ and the data is given in terms of a sequence $(Y_t, X_t)_{t \in \{1,\dots,n\}}$. According to Pfister et al. (2019, p. 1273), the main idea is to set up a regression using up to p lags of all variables and to regress the response on $Z_t^{S,p}$ instead of X_t^S. Hence, the linear model in Definition 1(a) can be transformed by defining an $(|S| + (d+1)p \times 1)$ vector $\eta := (\beta, B_1, \dots, B_p)$ where β is defined as in Definition 1 and B_k is a $((d+1) \times 1)$ vector for $k = 1, \dots, p$, such that

$$Y_t = Z_t^{S,p}\eta + \varepsilon_t \tag{39}$$

$$= \underbrace{\begin{pmatrix} X_t^S & Y_{t-1} & X_{t-1} & \cdots & Y_{t-p} & X_{t-p} \end{pmatrix}}_{(1\times|S|+(d+1)p)} \underbrace{\begin{pmatrix} \beta & B_1 & \dots & B_p \end{pmatrix}'}_{(|S|+(d+1)p\times 1)} + \varepsilon_t \tag{40}$$

$$= X_t^S\beta + \sum_{k=1}^{p} \begin{pmatrix} Y_{t-k} & X_{t-k} \end{pmatrix} B_k + \varepsilon_t. \tag{41}$$

In summary, the intuition of the hypothesis in (38) is that for all cases in which $\widetilde{H}_{0,S,p}$ is rejected and, consequently, the set S is not invariant, the dependence of Y_t on $Z_t^{S,p}$ is not captured by the same linear function across all periods (Pfister et al. 2019, p. 1267). Hence, $\widetilde{H}_{0,S,p}$ can be tested by testing the goodness-of-fit of the model in (41) (ibid.). To do so, let $\boldsymbol{Z}^{S,p} = (Z_{p+1}^{S,p}, \dots, Z_n^{S,p})$ and $\boldsymbol{Y} = (Y_{p+1}, \dots, Y_n)$. Since the model is estimated using OLS, the projection matrix as presented in standard textbooks (see, for instance, Hayashi 2001, p. 18) is given by $\boldsymbol{P}_{\boldsymbol{Z}}^{S,p} := \boldsymbol{Z}^{S,p}((\boldsymbol{Z}^{S,p})'\boldsymbol{Z}^{S,p})^{-1}(\boldsymbol{Z}^{S,p})'$ and the OLS residuals are $\boldsymbol{R}^{S,p} := (I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\boldsymbol{Y}$. Furthermore, under the null hypothesis, that is, assuming that $\widetilde{H}_{0,S,p}$ is true, Pfister et al. (2019, p. 1273) define the *scaled residuals* by

$$\widetilde{\boldsymbol{R}}^{S,p} := \frac{(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\boldsymbol{Y}}{||(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\boldsymbol{Y}||_2} = \frac{(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\boldsymbol{\varepsilon}}{||(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\boldsymbol{\varepsilon}||_2} = \frac{(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\tilde{\boldsymbol{\varepsilon}}}{||(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\tilde{\boldsymbol{\varepsilon}}||_2}, \tag{42}$$

where $\tilde{\boldsymbol{\varepsilon}} := \boldsymbol{\varepsilon}/||\boldsymbol{\varepsilon}||_2$ is the scaled noise and $||\cdot||_2$ is the Euclidean norm by which the scaling is performed.

The scaled residuals are of particular relevance, since Pfister et al. (2019, pp. 1269) make use of them to construct several test statistics, two of which will be discussed in the following. The motivation of the test statistics is twofold: First, they are meant to capture violations of the model in (39) that might occur at unknown periods, so-called change points, and, second, that they take on small absolute values whenever the model is true and large ones in the opposite case, since the model might be considered the central ingredient of the null hypothesis $\widetilde{H}_{0,S,p}$ of invariance. Clearly, in practice, the true change points are unknown, which is why a grid of potential change points that defines different *environments* across all periods is considered. In this setting, "violations in the invariance occur due to differences in the structural form of model [(39)] between two different environments" Pfister et al. (2019, pp. 1269). This gives rise to a collection of environments $\mathcal{E} \subseteq \mathcal{P}(\{1, \dots, n\})$ and all of the test statistics incorporate the pairwise comparisons between environments $e, f \in \mathcal{E}$ by some means or other in order to detect the most important violations of the model's invariance[5]: Either different regression coefficients or different residual variances obtained from regressing $\boldsymbol{Y}$ on $\boldsymbol{Z}^{S,p}$ in both environments e and f separately (ibid.). According to the authors, both types of violations can be detected by separate regressions of $\widetilde{\boldsymbol{R}}^{S,p}$, the scaled residuals in (42), on $\boldsymbol{Z}^{S,p}$ for two environments e and f. The regression coefficient obtained from such regression across all possible environments $h \subset \{1, \ldots, n\}$ is given by

$$\hat{\gamma}_{h,S,p} := ((\boldsymbol{Z}_h^{S,p})' \boldsymbol{Z}_h^{S,p})^{-1} (\boldsymbol{Z}_h^{S,p})' \widetilde{\boldsymbol{R}}_h^{S,p} \tag{43}$$

the corresponding sample variance by

$$\hat{s}^2_{h,S,p} := \frac{(\widetilde{\boldsymbol{R}}_h^{S,p} - \boldsymbol{Z}_h^{S,p} \hat{\gamma}_{h,S,p})'(\widetilde{\boldsymbol{R}}_h^{S,p} - \boldsymbol{Z}_h^{S,p} \hat{\gamma}_{h,S,p})}{|h|}, \tag{44}$$

respectively (ibid.). Having defined the central objects, one can follow Pfister et al. (2019, p. 1270) and separately test for a violation of invariance due to differences in the regression coefficients obtained from environments $e, f \in \mathcal{E}$ with the test statistic

$$T^1_{e,f}(\widetilde{\boldsymbol{R}}^{S,p}) := ||\hat{\gamma}_{e,S,p} - \hat{\gamma}_{f,S,p}||_2 \tag{45}$$

and for differences in the residual variance with the test statistic

$$T^2_{e,f}(\widetilde{\boldsymbol{R}}^{S,p}) := \frac{\hat{s}^2_{e,S,p}}{\hat{s}^2_{f,S,p}} - 1. \tag{46}$$

Clearly, one would like to obtain one single p-value for the test that Pfister et al. (2019, p. 1270) refer to as *decoupled test*, so it is necessary to combine both test statistics that I just mentioned. Since this combination of $T^1_{e,f}$ and $T^2_{e,f}$ is a problem of multiple testing, that is, of testing several hypotheses simultaneously, the authors propose to employ a Bonferroni correction to preserve the test's level α which means that the probability of one or more

[5] Here, $\mathcal{P}$ denotes the *power set*, that is the set of all subsets of $\{1, \ldots, n\}$.

false rejections does not exceed this predefined level.[6]

A second option proposed by the authors is the test statistic

$$T^3_{e,f}(\widetilde{\boldsymbol{R}}^{S,p}) := \frac{(\widetilde{\boldsymbol{R}}^{S,p}_e - \boldsymbol{Z}^{S,p}_e \hat{\gamma}_{f,S,p})'(\widetilde{\boldsymbol{R}}^{S,p}_e - \boldsymbol{Z}^{S,p}_e \hat{\gamma}_{f,S,p})}{\hat{s}^2_{f,S,p}|e|} - 1 \tag{47}$$

which tests for both types of violations, that is, for differences in the regression coefficients as well as for differences in the residual variances simultaneously, which is why Pfister et al. (2019, p. 1270) refer to it as *combined test*. As a last step, note that the test statistics (45)–(47) rely on pairwise comparisons of environments $e, f \in \mathcal{E}$. In order to obtain a final test statistic for $\widetilde{H}_{0,S,p}$ that incorporates all pairwise comparisons, Pfister et al. (2019, p. 1269) propose to take either the maximum

$$T^{\max,\mathcal{F}}_i(\widetilde{\boldsymbol{R}}^{S,p}) := \max_{(e,f)\in\mathcal{F}} \{T^i_{e,f}(\widetilde{\boldsymbol{R}}^{S,p})\} \tag{48}$$

or the sum

$$T^{\text{sum},\mathcal{F}}_i(\widetilde{\boldsymbol{R}}^{S,p}) := \sum_{(e,f)\in\mathcal{F}} T^i_{e,f}(\widetilde{\boldsymbol{R}}^{S,p}), \tag{49}$$

where $i \in \{1, 2, 3\}$ and $\mathcal{F} \subseteq \mathcal{E} \times \mathcal{E}$.

So far, we have seen how to formalize the notion of invariance, how to formulate an appropriate null hypothesis to test for its existence, and how to define test statistics that allow to capture a model's invariance or violations thereof. Yet, to turn ICP into an applicable testing procedure, the test statistics' distribution under the null hypothesis $\widetilde{H}_{0,S,p}$ needs to be carved out, as it forms the basis to obtain critical values. Not surprisingly, these distributions are not easily obtained analytically, which is why Pfister et al. (2019, p. 1268) propose a resampling procedure that is based on the idea of bootstrapping to approximate the distribution empirically. The procedure works as follows: In a first step, OLS estimation is applied to model (39) to obtain the projection matrix $\boldsymbol{P}^{S,p}_{\boldsymbol{Z}}$. Then, scaled bootstrap residuals are

[6] For an extensive discussion of the problem of multiple testing, the reader is referred to Lehmann and Romano (2008, chapter 9) and in particular to Lehmann and Romano (2008, p. 350, Theorem 9.1.1) for an explanation of the Bonferroni correction. In the present case of combining two test statistics, each separate p-value must be below $\alpha/2$ for a rejection of the null hypothesis.

computed by $\widehat{\boldsymbol{R}}^{S,p} := (I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\tilde{\boldsymbol{\varepsilon}}/||(I_n - \boldsymbol{P}_{\boldsymbol{Z}}^{S,p})\tilde{\boldsymbol{\varepsilon}}||_2$ exploiting that under $\widetilde{H}_{0,S,p}$, it holds that $\tilde{\boldsymbol{\varepsilon}} \sim \mathcal{N}(0, I_n)$. Next, the scaled bootstrap residuals are used to calculate test statistics (45)–(47). Finally, empirical distributions of the test statistics and critical values to which they can be compared are obtained by simply replicating the procedure I just outlined $B \in \mathbb{N}$ times. The absolute value of the test statistics should be *small* under $\widetilde{H}_{0,S,p}$, so the null hypothesis is rejected for values of the test statistics that are *larger* than the critical values and accepted otherwise (Pfister et al. 2019, p. 1268).

This last observation leads back to the remark from above regarding the rather counterintuitive formulation of the null hypothesis and, in fact, provides a justification for it: The discussion of the present section reveals that ICP seeks to infer causal relations by testing for a model's invariance across different environments within the data. This idea is captured by the test statistics above, since all of them measure differences between environments. In this setting, invariance equates to differences of zero which means that the test statistics are bounded from below and take on a fixed value under the null hypothesis. Additionally, the test statistics are not bounded from above and hence, there is no constraint on evidence against a model's invariance that would give rise to high differences between environments. From a statistical point of view, this is a convenient setup and it is not clear how to reach an equally convenient situation along with a more intuitive null hypothesis that is formulated in terms of non-causality and, hence, non-invariance.

To conclude the outline of ICP, one remark regarding the central properties of any statistical test, its size and its power, seems in place. While a test's size refers to the maximum risk to commit a *type I error*, which amounts to falsely reject a true null hypothesis, and should correspond to the predefined significance level α, the latter corresponds to one minus the probability to commit a *type II error* and hence gives the probability of rejecting the null hypothesis when it is false, which should ideally equal one (Bierens 2007, pp. 125). Pfister et al. (2019, p. 1273, Proposition 3) show that the tests based on (45)–(47) indeed achieve the correct level as the number of bootstrap replications $B \in \mathbb{N}$ tends to infinity. Furthermore, the authors show that the power of the decoupled test and the combined test converges to one as the sample size and the number of bootstrap replications goes to infinity, which entails that both tests are in fact consistent (Pfister et al. 2019, pp. 1271, Theorems 1 and 2). Yet this convergence happens at different rates for both tests, being faster for the decoupled test, and, in general, rather slow. In fact, the simulation results provided by Pfister et al. (2019, Supplementary Material (SM), pp. 2) reveal that for small violations of invariance and for samples up to 300 observations, the power of decoupled and combined test is rather low. This casts doubt on the applicability of ICP, in particular in macroeconometrics, where small samples are still commonplace, for instance due to low-frequency data. The empirical analyses in the next chapter will shed light on the method's performance in comparison to Granger-causality testing and SVAR models.

4 Application

After providing the econometric basis for the present investigation of the relationship between uncertainty and the business cycle, this chapter is concerned with the results of the empirical analyses. Given the diverging theoretical as well as empirical results in the literature, my goal is to provide an analysis inspired by the *ceteris paribus* interpretation that is well known to the econometrician: I employ three different econometric methods for causal inference—namely those that were outlined above—to a dataset that is left fixed across all analyses. This allows for a comparison of results obtained from different methodologies, which is an advantage over existing contributions to the literature that often rely both on different methodological approaches and different data. Thus, in the course of this chapter, I will first outline the data used for the subsequent econometric analyses with a particular focus on the question as to how uncertainty should be measured properly. Second, I will present the empirical results obtained from testing for Granger-causality, from SVAR analysis and, finally, from ICP.

4.1 The Data

As pointed out above, the key challenge for every empirical analysis of uncertainty is to come up with an appropriate way of measuring the object of interest. In fact, several ways have been proposed in the literature, since "no objective measure of uncertainty exists" (Jurado et al. 2015, p. 1178). In the following, I will employ two different measures, the macro uncertainty index (MUI) proposed by Jurado et al. (2015) and the EPU index proposed by Baker et al. (2016). There are several reasons for this choice: First, employing multiple measures of uncertainty instead of only one provides an immediate robustness check for the results. Second, it is common practice in the recent literature to consider both the MUI and the EPU index (for instance in Ludvigson et al. 2020). Third, using these measures allows to exploit their distinct benefits while avoiding the shortcomings of other measures. For instance, the EPU index is based on keywords in newspapers which makes it inherently "forward looking in that [it] reflect[s] the real-time uncertainty perceived and expressed by journalists" (Baker et al. 2020, p. 4). The MUI, on the other hand, is specifically constructed to capture uncertainty about economic fundamentals and starts from the—very plausible—premise that uncertainty is tightly linked to predictability. One might argue that this idea is also embodied in approaches that try to measure uncertainty by forecaster disagreement, that is, by considering the dispersion of point forecasts about some economic outcome of interest.

However, as pointed out by Baker et al. (2020, p. 5), while "there is a long history of using such disagreement measures to proxy for uncertainty" there is "also a long history of disagreement about their suitability for that purpose." So let us take a closer look at how exactly the MUI and the EPU index are constructed.

As already mentioned, the idea of the MUI is to remove the forecastable component of a given series and to compute its conditional volatility afterwards (Jurado et al. 2015, p. 1179). Thus, the h-period ahead uncertainty in some variable $y_{jt} \in Y_t = (y_{1t}, \ldots, y_{Nt})'$ can be defined as

$$\mathcal{U}^y_{jt}(h) \equiv \sqrt{E\left[\left(y_{j,t+h} - E\left[y_{j,t+h}|I_t\right]\right)^2 |I_t\right]}, \tag{50}$$

where the information available at time t is denoted by I_t (Jurado et al. 2015, p. 1178). To obtain an overall index, individual uncertainties are aggregated using aggregation weights w_j, such that

$$\mathcal{U}^y_t(h) \equiv \text{plim}_{N\to\infty} \sum_{i=1}^{N} w_j \mathcal{U}^y_{jt}(h) \equiv E_w\left[\mathcal{U}^y_{jt}(h)\right]. \tag{51}$$

The main contribution of Jurado et al. (2015) is to obtain estimates for (50) and (51) that in fact allow for a construction of an index that measures macroeconomic uncertainty. Without going too far into the details that are beyond the scope of the present text, it can be said that their strategy is to approximate $E\left[y_{j,t+h}|I_t\right]$ with h-step-ahead forecasts, to obtain the corresponding forecast error $V^y_{j,t+h} \equiv y_{j,t+h} - E\left[y_{j,t+h}|I_t\right]$ as well as the conditional variance, $E\left[\left(V^y_{t+h}\right)^2 |I_t\right]$, and to build the final index $\mathcal{U}^y_t(h)$ from an equally-weighted average of individual uncertainties $\mathcal{U}^y_{jt}(h)$ (Jurado et al. 2015, pp. 1179). Furthermore, the construction of the MUI relies on data from $N = 132$ different macroeconomic time series that contain information on

> "real output and income, employment and hours, real retail, manufacturing and trade sales, consumer spending, housing starts, inventories and inventory sales ratios, orders and unfilled orders, compensation and labor costs, capacity utilization measures, price indexes, bond and stock market indexes, and foreign exchange measures" (Jurado et al. 2015, p. 1189).

The EPU index proposed by Baker et al. (2016), on the other hand, does not rely on macroeconomic data or econometric estimates, as it is solely based on newspaper articles. The construction of the index is described in detail in Baker et al. (2016, p. 1599) and proceeds as follows: Ten leading U.S. newspapers, for instance the New York Times and the Wall Street Journal, are tracked to create a monthly count of articles that contain a combination of the terms ‘uncertainty’ or ‘uncertain’; ‘economic’ or ‘economy’; and one policy term such as ‘Congress’, ‘Federal Reserve’ or ‘White House’. In order to control for the overall number of articles published in a given newspaper and month, the raw counts obtained from newspaper $i = 1, \dots, 10$ are scaled by the overall number of articles in that corresponding newspaper and month, resulting in the series X_{it}. Subsequently, the scaled counts are standardized to a unit standard deviation, which yields a new series Y_{it}. The final EPU index at time t is obtained by computing the mean of the latter series across all newspapers by month and by normalizing it to 100 for the period from 1985 to 2009.

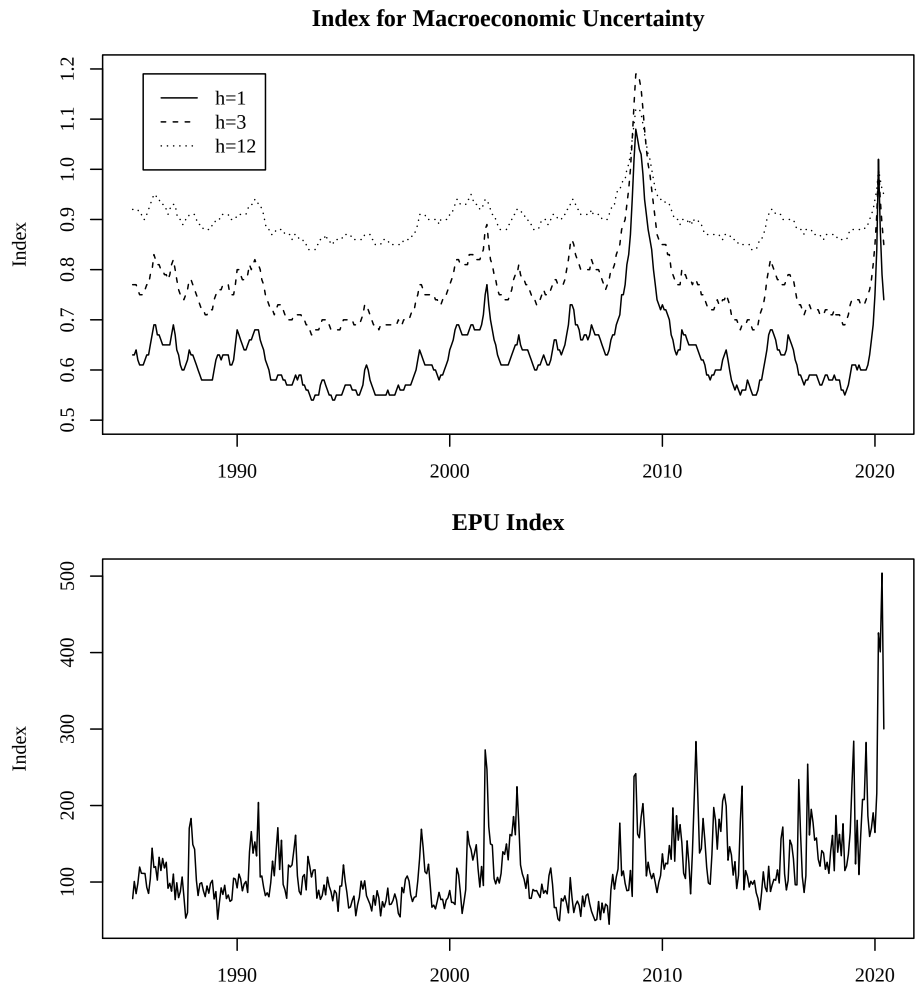


Figure 1: Time series plots for the MUI (top) and the EPU indices (bottom) ranging from 1985:01 through 2020:06.

Both the MUI—for forecast horizons $h = 1, 3, 12$—and the EPU index are depicted in figure 1. As becomes evident, they exhibit a similar behavior over time while stressing different events to a different extent. For instance, on the one hand, the financial crisis of 2008 seems to have created a sharp increase in the MUI, regardless of the forecast horizon, while the increase in the EPU index is more moderate. The ongoing COVID-19 pandemic, on the other hand, appears as a sharp increase in both the MUI and the EPU index, but while the latter attains an all-time high, the level of the former remains below the one in the midst of the financial crisis. The bottom line of these observations is that, albeit evolving rather similarly, different measures of uncertainty are appropriate to capture different aspects of the overall concept of uncertainty. While the financial crisis mainly induced uncertainty about economic fundamentals as incorporated by the MUI, the ongoing pandemic seems to generate forward-looking uncertainty about economic policy as captured by the EPU index. Note, that in order to make the following analyses involving the MUI and the EPU index comparable, I restrict the attention to the MUI series with $h = 1$. This is because the EPU index and hence the newspaper articles it incorporates are inherently forward looking, but presumably for a rather restricted period of time. In fact, it seems reasonable to assume that the number of articles addressing an issue that affects uncertainty about economic policy increases, the closer this certain issue—consider, for instance, a presidential election—lies ahead. Consequently, one ought to consider the MUI series with $h = 1$ for comparison, since its focus is on one-month-ahead (un)predictability and the corresponding volatility.

Apart from the measures of uncertainty, the subsequent econometric analyses require macroeconomic data, both to capture the business cycle and to control for other economically relevant factors. To achieve the former, I follow the existing literature and employ U.S. industrial production—an index normalized to the value of 100 in 2012—as a measure of business cycle activity.[7] Other factors that I include as control variables following the literature are the Federal Funds Rate in percent, employment measured in thousands of persons, and the S&P 500 stock index. All variables are sampled at monthly frequency and cover the period from January 1985 to June 2020, which yields a sample of size $T = 426$ (for more details regarding the sources see appendix B). The depiction of the raw macroeconomic time series in figure 2 reveals two issues that need to be taken into account in any econometric analysis: First, the ongoing pandemic exhibits severe consequences by inducing a sharp drop in all of the series. Clearly, this

[7] Examples for this strategy can be found in Baker et al. (2016), Berger et al. (2020), Jurado et al. (2015), and Ludvigson et al. (2020).

needs to be conceived of as an outlier that one should exclude or for which at least robustness check ought to be performed. Second, all series exhibit a clear-cut trending behavior, either upward sloping in the case of industrial production, employment, and the S&P 500, or downward sloping in the case of the Federal Funds Rate.

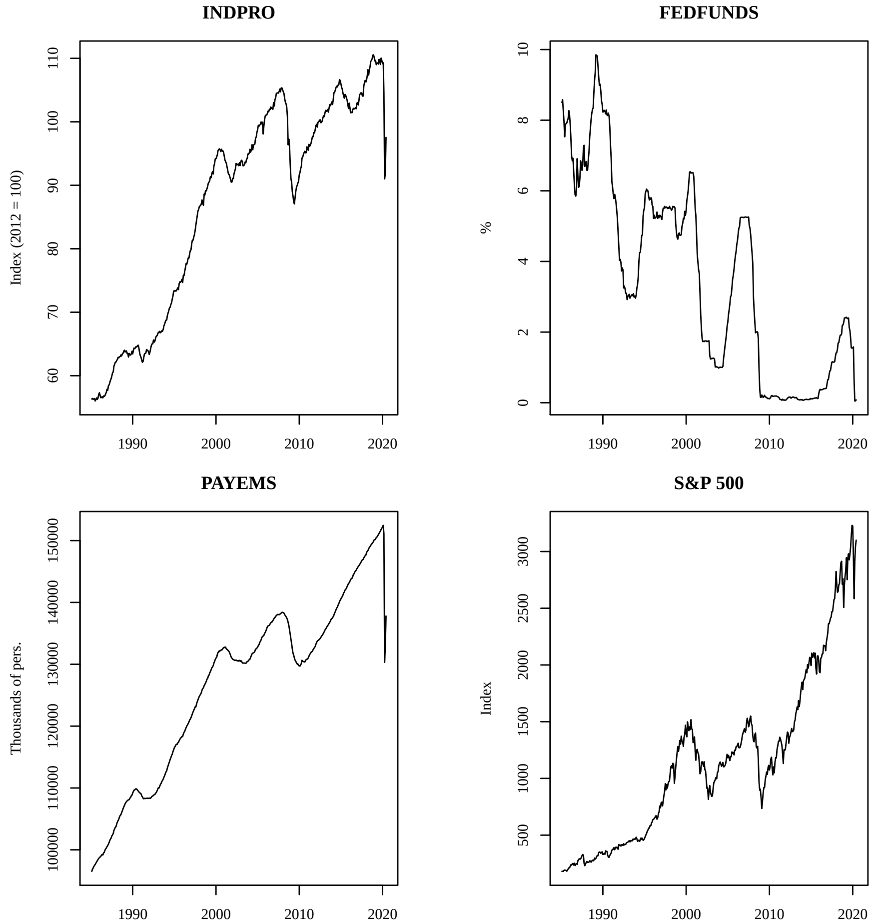


Figure 2: Time series plots for monthly U.S. industrial production (INDPRO), Federal Funds Rate (FEDFUNDS), employment (PAYEMS) and the S&P 500 ranging from 1985:01 through 2020:06.

This is strong evidence against the assumption that the series arise from a stationary data generating process, which is why I transform them to (log-)differences. The transformed series are depicted in figure 3. Here, all series seem to evolve around some time-invariant mean and with some time-invariant variance. Furthermore, note, that the graphs in figure 3 only range until January 2020 in order to exclude the outlier caused by the COVID-19 pandemic.

4.2 Empirical Results

After the mere description of the data in the preceding section, the present section is concerned with the econometric analyses that I employed in order

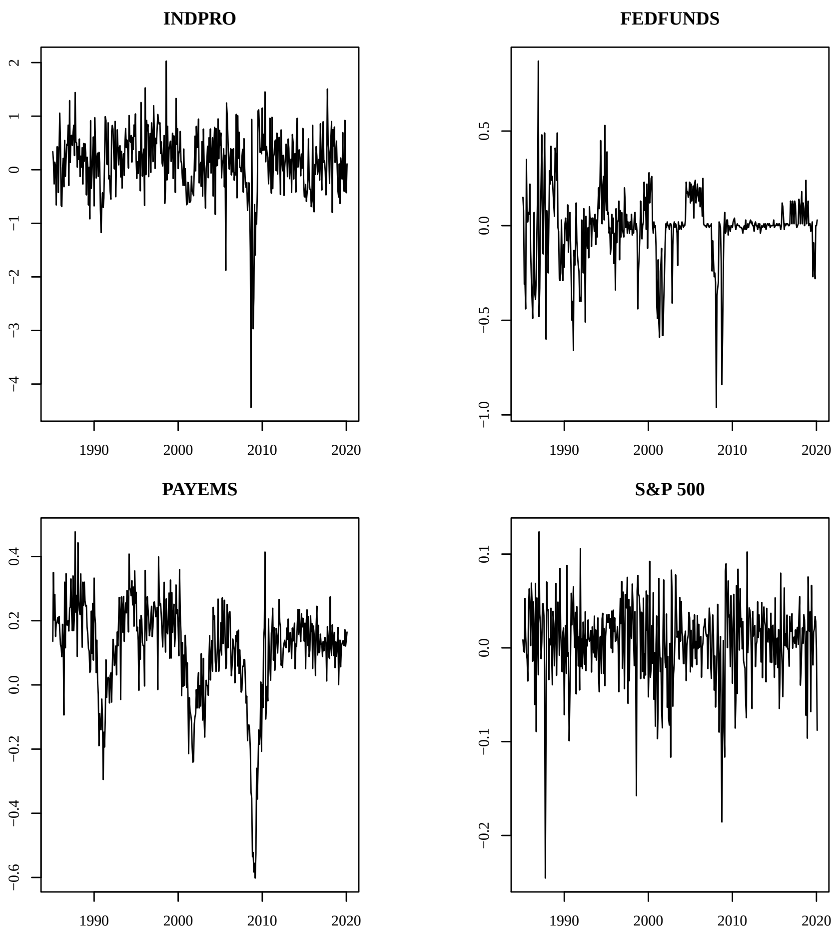


Figure 3: Time series plots for log-differences of monthly U.S. industrial production (INDPRO), employment (PAYEMS) and the S&P 500 as well as first differences for the Federal Funds Rate (FEDFUNDS) ranging from 1985:01 through 2020:01.

to illuminate the relationship between uncertainty and the business cycle. Recall, that the idea is to assess different methods of causal inference and the corresponding results while leaving everything else equal to enable a proper comparison. Thus, the following analyses were performed using the same dataset that I introduced above.

4.2.1 Granger-Causality

The procedure of testing for Granger-causality starts from a generalization of the reduced-form VAR in (1) to $K = 5$ variables. These are, in the order of their occurrence in the model, an uncertainty measure, industrial production in log-differences, the Federal Funds Rate in first differences, and the employment as well as the S&P 500 index in log-differences. Before the test for Granger-causality itself can be performed, the final structure of the model on which the test is based, and in particular its lag order p, has to be determined. To do so, the model is estimated using OLS for several lag orders up to $p = 16$ and subsequently, three different information criteria are computed following Lütkepohl (2005, pp. 147), namely the *Akaike Information Criterion* (AIC), given by

$$AIC(m) = \log(\det(\widehat{\Sigma}_u(m))) + \frac{2mK^2}{T}, \tag{52}$$

the *Schwarz Information Criterion* (SIC)

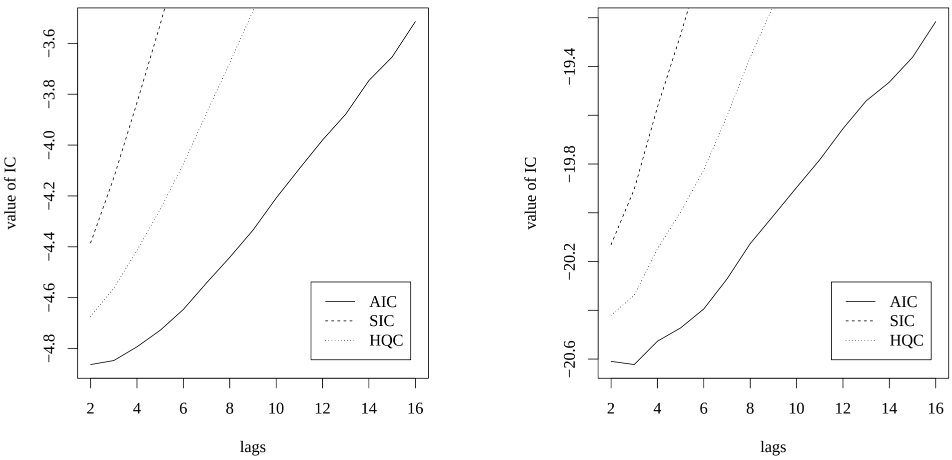
$$SIC(m) = \log(\det(\widehat{\Sigma}_u(m))) + \frac{\log(T)}{T} mK^2, \tag{53}$$

and the *Hannan-Quinn Criterion* (HQC)

$$HQC(m) = \log(\det(\widehat{\Sigma}_u(m))) + \frac{2\log\log(T)}{T} mK^2. \tag{54}$$

In all cases, m denotes the lag order at which each criterion is evaluated, T is the number of observations used in the estimation and $\widehat{\Sigma}_u$ is the (5×5) estimate for the variance-covariance matrix obtained from the OLS residuals in the way presented in Kilian and Lütkepohl (2018, p. 31, (2.3.4)). All information criteria rely on the idea of a trade-off between an improved fit of a model for a higher lag order, as reflected by a lower $\log(\det(\tilde{\Sigma}_u(m)))$, and the parsimony of the model, as reflected by the rightmost terms in (52)–(54) that increase with m. As a consequence, the lag order should be chosen such that the information criteria are minimized. Figure 4 reveals that apart from one case, all information criteria are minimized for $m = 2$, regardless whether the EPU index (panel (a)) or the MUI (panel (b)) is used as the measure of uncertainty. Consequently, a VAR(2) is used to test for Granger-causality.

OLS estimation of the five-variable VAR(2) leads to results that justify the assumption of stable and stationary processes: Regardless whether the MUI or the EPU index is used as measure of uncertainty, the eigenvalues of the model's (10×10) companion matrix have modulus less than one. This in turn implies that the use of the standard Wald testing procedure outlined above is asymptotically valid (see Kilian and Lütkepohl 2018, p. 42 and pp. 49 for procedures that deal with violations of stability). Thus, it is straightforward to test both directions of causality separately, that is, from business cycle activity to uncertainty and vice versa. The null hypothesis of no Granger-causality from business cycle activity to uncertainty can be expressed in terms of the VAR(2) coefficients as



(a) (b)

Figure 4: Values of AIC, SIC and HQC for lag orders $m \in \{1, \ldots, 16\}$, calculated from a five-variable VAR including (a) the EPU index or (b) the MUI as well as log-differences of industrial production, first differences of the Federal Funds Rate, log-differences of both employment and the S&P 500.

$$H_0\colon a_{12,1} = a_{12,2} = 0, \tag{55}$$

or as in (8) by defining a (2×1) vector c and the (2×55) matrix $C_{BC,UC}$ with elements $c_{1,7} = c_{2,32} = 1$ and all other elements equal to zero. Equivalently, the opposite direction, that is the null hypothesis of no Granger-causality from uncertainty to business cycle activity can be expressed in terms of the VAR(2) coefficients as

$$H_0\colon a_{21,1} = a_{21,2} = 0, \tag{56}$$

or as in (8) using the (2×55) matrix $C_{UC,BC}$ with elements $c_{1,6} = c_{2,31} = 1$ and all other elements equal to zero.

The last step that concludes the test for Granger-causality requires to combine the estimation results and the notation introduced above, to compute the test statistic λ_F as in (12) and to compare it to the critical values c_α from the $F(2, 423 - 5 \cdot 2 - 1) = F(2, 412)$-distribution that can be obtained from Hamilton (1994, pp. 756, Table B.4) for significance levels $\alpha \in \{.05; .01\}$. Table 2 summarizes the results. As becomes evident, the results mirror the state of the literature in the sense that there is overwhelming evidence for *all* directions of Granger-causality and for both measures of uncertainty, since the values of the test statistic are exceptionally high and the null hypothesis of no Granger-causality can be rejected at the 1% level of significance in all cases. Furthermore, the results do not change qualitatively depending on whether the period from February until June 2020, that might be considered as an outlier, is in- or excluded in the analysis. This overall outcome, however, should only be interpreted cautiously as a first and rather descriptive step towards answering the question regarding the causal relationship between uncertainty and business cycle activity. We have seen above that the concept of Granger-causality relies on the notion of prediction rather than causation and, as a consequence, the results in table 2 only license the conclusion that business cycle activity helps to predict uncertainty and vice versa. Although this might be considered an interesting insight *per se*, further methods need to be employed in order to discern causal effects from non-causal ones.

Table 2: Results of testing Granger-noncausality from economic activity to uncertainty and vice versa using U.S. industrial production in log-differences to measure economic activity, EPU and MUI as uncertainty measures, and 5% as well as 1% significance levels. The sample ranges from 1985:02 to 2020:06 ($T = 425$).

		Critical values	
Direction tested	λ_F	$c_{.05}$	$c_{.01}$
INDPRO → EPU	10.17***	3.04	4.71
EPU → INDPRO	152.09***		
INDPRO → MUI	60.00***		
MUI → INDPRO	2903.95***		

Note: (***) indicates that the null hypothesis of Granger-noncausality can be rejected at the 1%-level of significance. Critical values were obtained from Hamilton (1994, p. 760, Table B.4).

4.2.2 Structural Vector Autoregressive Analysis

The next step consists in transforming the reduced-form VAR that was used to test for Granger-causality to a SVAR and to employ impulse response functions to investigate if and how a shock to one variable of interest, say uncertainty, affects the other variable of interest, industrial production in the case at hand, over time. As we have seen above, the crucial step in any SVAR analysis is that of identification: Recall, that in order to recover the structural shocks from the reduced-form VAR requires constraints on the elements of matrix B_0^{-1} to avoid underidentification. The literature review and its summary in table 1 revealed that the recursive identification scheme by means of a Cholesky decomposition as outlined above is the most common approach to achieve identification in the context of uncertainty and the business cycle. On the one hand, this seems reasonable given the diverging results in the literature that impede the justification of specific short- or long-run restrictions that are other, widely-used identification strategies. On the other hand, a recursive identification scheme itself imposes a particular structure on the variables under consideration, for it entails that the variable that is ordered first in the VAR is not affected contemporaneously by any of the other variables, while the variable that is ordered last is affected contemporaneously by all of the other variables Kilian and Lütkepohl (2018, pp. 216).

In the present SVAR analysis, the $K = 5$ variables are ordered as in the test for Granger-causality above. That is, the measure of uncertainty—MUI or the EPU index—is ordered first, log-differences of industrial production are ordered second, followed by first differences of the Federal Funds Rate; log-differences of employment are ordered fourth and finally, as the fifth variable, log-differences of the S&P 500 are included in the model. The reasons for this particular ordering are twofold: First, the ordering closely mimics the one employed by Jurado et al. (2015, p. 1202) who justify it by the observation that it "affords the advantage of containing a set of variables whose dynamic relationships have been the focus of extensive macroeconomic research." Second, it differs from the ordering employed by Jurado et al. (2015) in the important aspect that the measure of uncertainty is ordered *first* instead of *last*. As mentioned above, ordering uncertainty first results in a setting in which it is not affected contemporaneously by any of the other variables. Consequently, this is the most conservative setting for assessing the endogeneity of uncertainty, since contemporaneous impacts of other variables are ruled out *ex ante*.

The results from conducting the SVAR analysis that I just described using the MUI as measure of uncertainty are depicted in figure 5. Since the focus of the investigation at hand is on the relation between uncertainty and business cycle activity, solely impulse responses for these two variables of interest are shown. Solid lines represent point estimates of the impulse response functions, while dashed and dash-dotted lines represent the corresponding one standard error and two standard error confidence intervals, respectively, that were computed using the bootstrapping procedure outlined above. The left column in figure 5 contains impulse responses of the MUI and industrial production to a one-standard-deviation shock to the MUI, that is, to uncertainty. For the MUI, clearly, the shock results in an immediate and enduring increase that is followed by a slow decline such that 25 months after the shock, the index is still above its initial level. Interestingly, however, convergence to the initial level happens a lot faster for industrial production: As one might have expected, an uncertainty shock leads to an immediate and pronounced drop at first, but only five months after the shock, industrial production is already close to its initial level. Furthermore, the confidence bands are rather tight in the case of industrial production, thereby providing evidence that the estimate for the IRF is reliable.

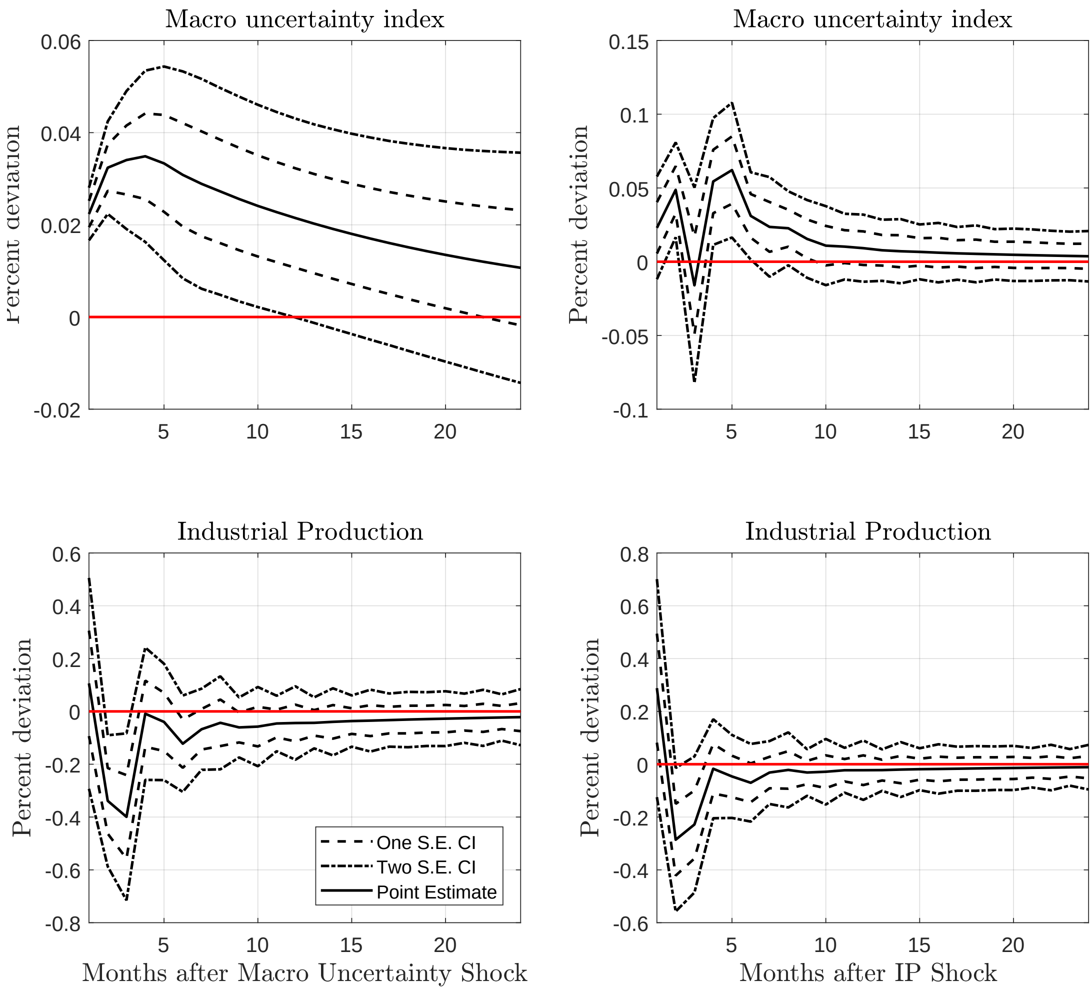


Figure 5: Point estimates and confidence intervals for impulse response functions for the Macro Uncertainty Index and U.S. industrial production after a shock to the Macro Uncertainty Index (left column) and to the industrial production (right column). The confidence bands were constructed using the bootstrapping procedure proposed by Gonçalves and Kilian (2004).

Now, what happens to both variables after a shock to industrial production (figure 5, right column)? In fact, the evolution of industrial production itself is similar to the one it exhibits after an uncertainty shock, with a sudden yet less pronounced drop and a quick recovery back to the initial level. The MUI, on the other hand, is increased by a shock to industrial production. The increase is even higher than the one after an uncertainty shock, but it only lasts for approximately five months and is replaced by a process of convergence to the initial level afterwards. Thus, as a first summary, one can say that uncertainty shocks affect the business cycle by lowering industrial production and that shocks to industrial production increase uncertainty. Both results seem to be relevant for rather short periods of time, since convergence back to the initial level happens within six months. However, quantitatively, the impact of uncertainty shocks to industrial production seems to be more severe than that of industrial production shocks to uncertainty: While industrial production deviates by -0.4% from its initial level after an uncertainty shock, the increase in uncertainty after a shock to industrial production only amounts to 0.1%. This is in line with the fact that the Granger-causality test statistics reported in table 2 are by far higher when testing for an impact of uncertainty on industrial production compared to the opposite direction.

In order to assess the robustness of the results that I just mentioned, the same analysis was repeated using the EPU index instead of the MUI as the measure of uncertainty. The results from this analysis are depicted in figure 6. First, mere eyeballing reveals that they are qualitatively similar to the ones obtained from using the MUI: An uncertainty shock leads to an immediate drop in industrial production that is followed by a convergence back to the initial level, a shock to industrial production generates an increase in uncertainty that disappears over time. In fact, the results are also qualitatively similar to the IRF reported by Baker et al. (2016, p. 1629) in the article in which they proposed the EPU index. The direct comparison to the MUI results, however, reveals two interesting aspects: First, the impulse responses generally seem to be associated with more uncertainty, since the confidence bands are considerably wider than above, at least for the first year after a shock. Second, the convergence back to the initial level takes more time for both uncertainty and industrial production. While the initial level was reached after approximately six months in figure 5 above, this process takes more or less one year once uncertainty is measured by the EPU index. The reason for this observation could be the inherently forward-looking nature of the EPU index that I mentioned above: While an increase in the MUI indicates that the economy has become less predictable, an increase in the EPU index might be due to very specific concerns about certain events in the future that are not ameliorated until the events finally materialize. Consequently, changes in or induced by the EPU index could exhibit an impact that lasts longer than the one of the MUI.

Overall, the SVAR analysis confirms the findings of the Granger-causality tests and extends them by introducing a dynamic perspective embodied by the impulse responses, that allow to track the different impacts over time. Until this point, there is evidence for both directions of causality, a result that is at least in line with the findings in the literature.

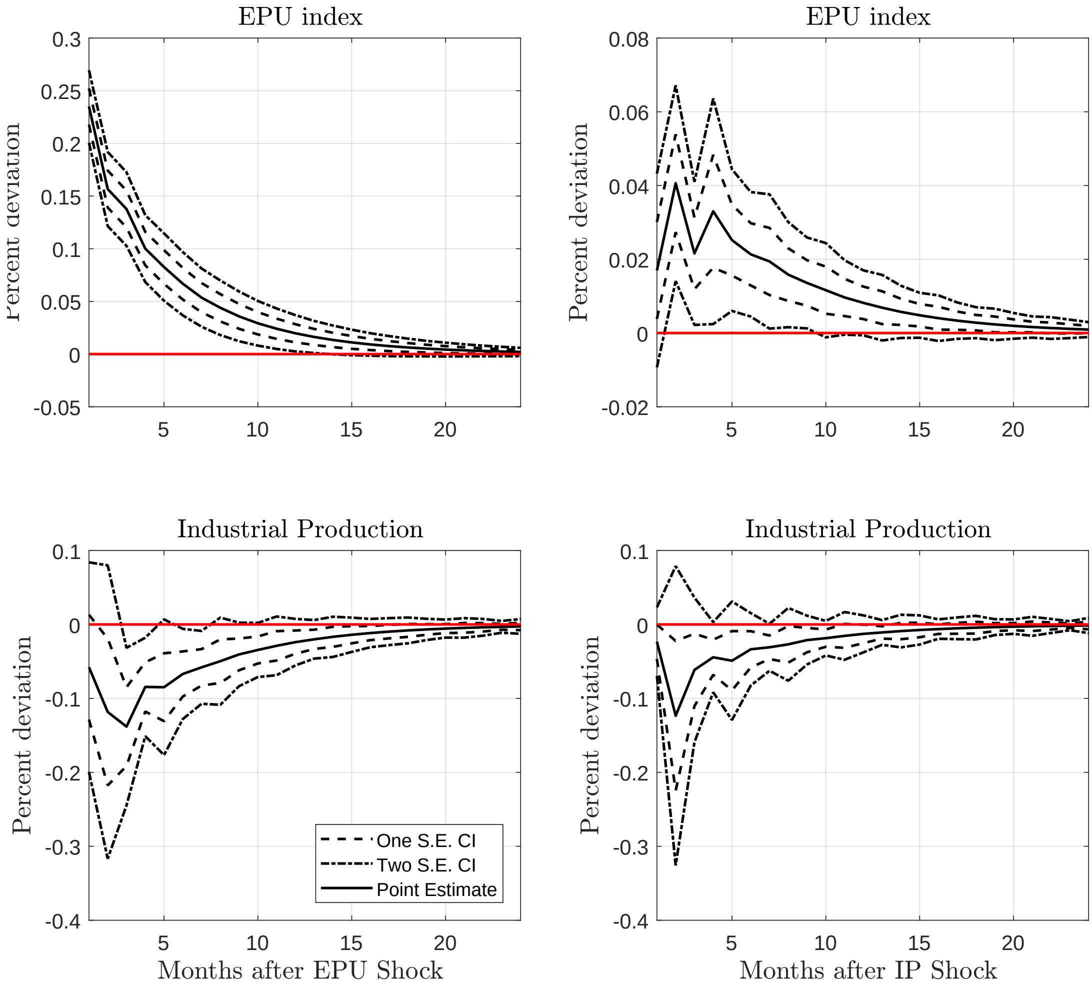


Figure 6: Point estimates and confidence intervals for impulse response functions for the EPU index and U.S. industrial production after a shock to the EPU index (left column) and to the industrial production (right column). The confidence bands were constructed using the bootstrapping procedure proposed by Gonçalves and Kilian (2004).

4.2.3 Invariant Causal Prediction

The last step of the empirical investigation consists in applying ICP to the data that was already used in the Granger-causality test as well as the SVAR analysis above. Thus, after a particular focus on predictability (Granger-causality) and dynamic evolution (SVAR and IRF), the focus is now on causal effects that happen contemporaneously, that is, within some given period of interest. First, note that the sample size—as in all preceding analyses—is $T = 426$, which ameliorates concerns about ICP's low power in small samples. However, the sample at hand is certainly still small enough to be considered finite while at the same time, the results should be as robust as possible, which is why I test several specifications of both the combined and the decoupled test as introduced in (45)–(47). More precisely, I employ the following setups of the tests: The lag order of model (39) is *not* determined using one of the more sophisticated methods suggested in Pfister et al. (2019, p. 1274) but testing is performed for all $p = 1, \dots, 8$. To generate different environments within the data, the change points are set to $t \in \{0, 70, 140, T\}$ following the empirical example in Pfister et al. (2019, pp. 1274) in which the authors use a macroeconomic dataset with a comparable sample size. Following a suggestion in Pfister et al. (2019, p. 1272), each environment is tested against its complement and not against all other environments that could be constructed based on the change points, which greatly enhances computational efficiency. Finally, both of the methods given in (48) and (49) to combine the test statistics obtained from these pairwise comparisons are considered to construct overall test statistics, since the test's final outcome is rather sensitive to these combination methods. I set the level of significance to $\alpha = .05$ and the number of bootstrap replications to $B = 1,000$, the latter also for reasons of computational efficiency.

As above, I begin with an analysis in which uncertainty is measured by the MUI. The results from testing whether industrial production causally affects the MUI, namely the p-values obtained from the test—for reasons of presentation in their natural logarithms—, are depicted in figure 7. As becomes evident, for all types of the test (decoupled/combined) and for both combination methods (sum/max), there are situations in which there is a significant contemporaneous effect of industrial production on the MUI as captured by p-values that fulfill $\log(p) \leq \log(.05) \approx -3.00$. However, this evidence seems to be highly dependent on the chosen lag order, for the decoupled test only indicates an effect for $p = 5$ or $p = 6$, while the combined test indicates an effect for $p = 2$, $p = 3$ or $p = 4$. This casts doubt on the finding, especially because it should be robust rather than highly sensitive to the precise specification of the testing procedure. Nevertheless, the insight that industrial production affects the MUI is in line with the analyses above.

For the opposite direction of causality, that is, for the MUI causally and contemporaneously affecting industrial production, the results—depicted in figure 8—are even less instructive: On the one hand, the combined test does not lead to any significant result, for none of the lag orders and combination methods considered. On the other hand, the decoupled test yields evidence of instantaneous causality for lag orders $p = 2$ (combination method: sum, top left panel) and $p = 1$ (combination method: max, bottom left panel). Thus, the results seem to depend on the chosen lag order, but furthermore on the chosen test statistic and on the method that is used to combine the different test statistics across environments in the data. Without doubt this is a myriad of variables that needs to be fixed in a justified manner before conducting an empirical analysis. Before pursuing this methodological point a bit further, let me finalize the empirical investigation with a remark regarding the results obtained from measuring uncertainty by the EPU index.

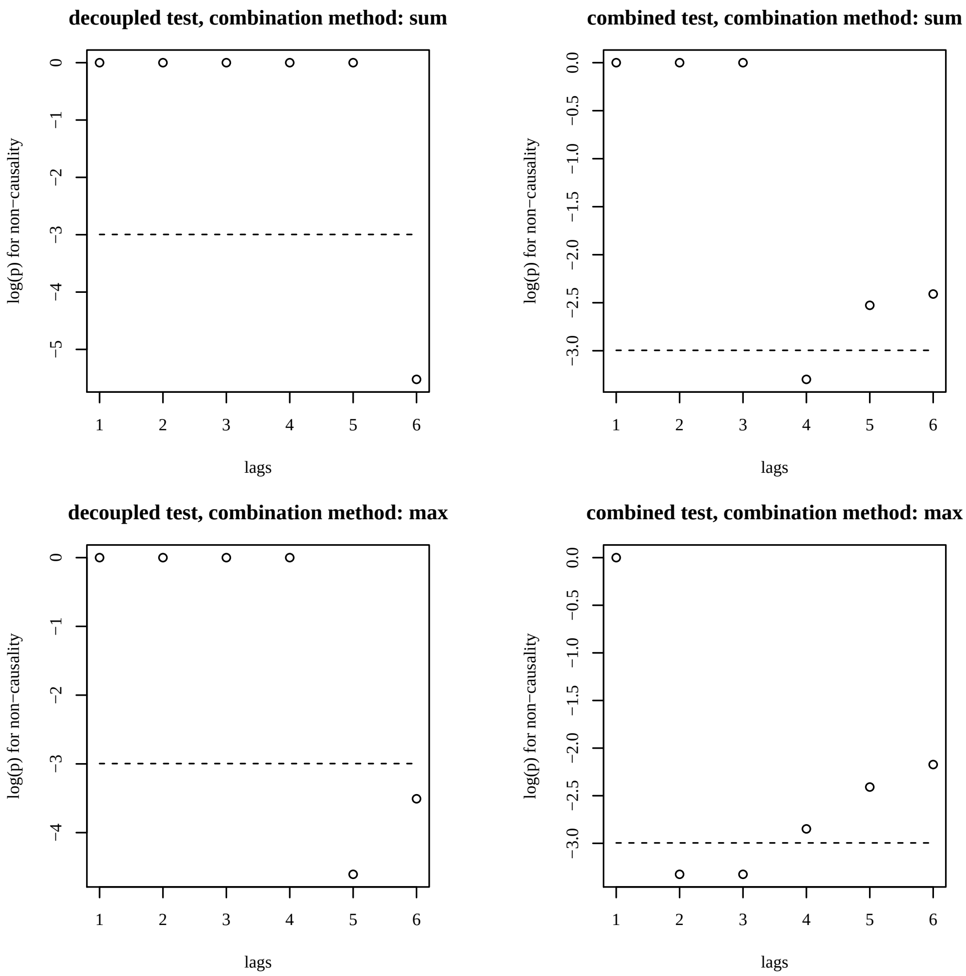


Figure 7: Results from testing causality from log-differences of U.S. industrial production to the MUI using the decoupled test (left column) and the combined test (right column) as well as the methods in (49) (top row) and (48) (bottom row) to combine individual test statistics for lag orders $p \in \{1, \ldots, 8\}$.

Surprisingly, in this setting, the only evidence for instantaneous causality could be obtained when using the decoupled test along with the maximum combination method to test whether the EPU index contemporaneously affects industrial production. In this case and for lag order $p = 4$, there is slight evidence for a causal relationship. However, as I just insinuated, for all other setups of the test and for the opposite direction of causality, no significant results could be identified, which is why I omit the corresponding plots at this point of the text. They can be found in appendix A.

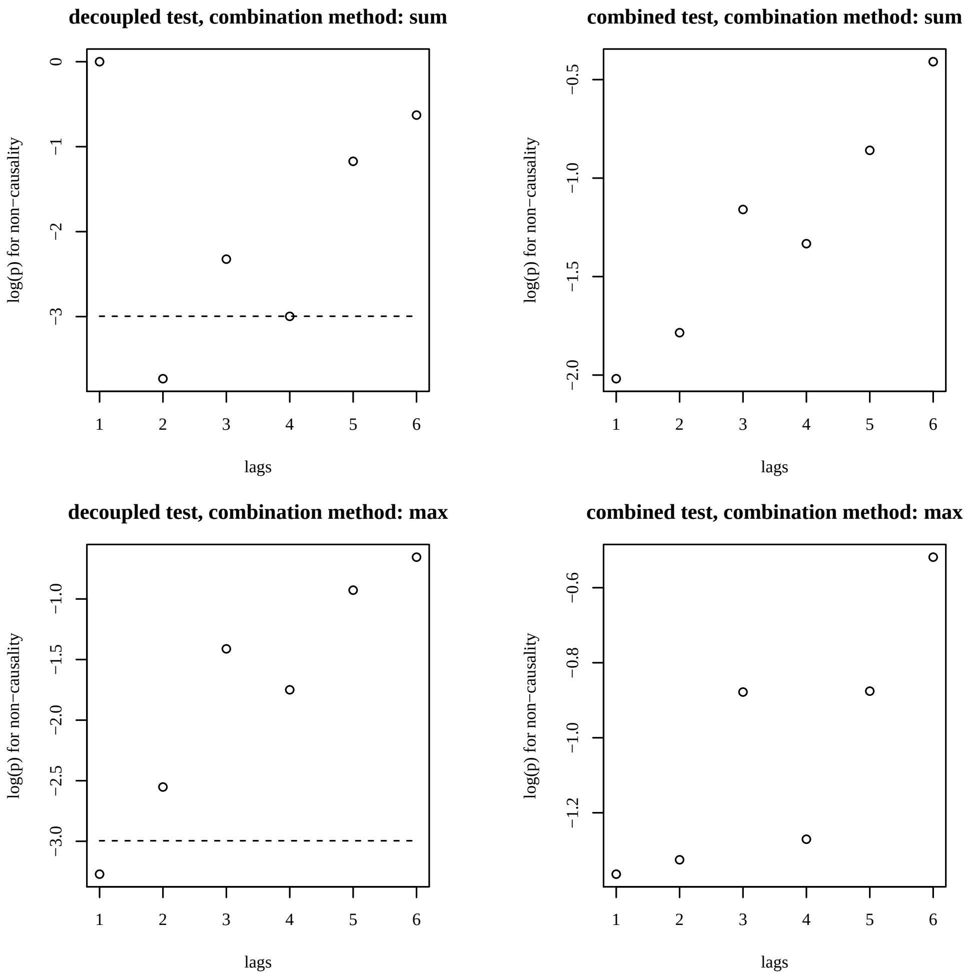


Figure 8: Results from testing causality from the MUI to log-differences of U.S. industrial production using the decoupled test (left column) and the combined test (right column) as well as the methods in (49) (top row) and (48) (bottom row) to combine individual test statistics for lag orders $p \in \{1, \ldots, 8\}$.

So, in summary, what is the upshot of employing ICP to investigate the relation of uncertainty and the business cycle? From a purely economic point of view, one might argue that if any, the most evident result is that there must be some causal impact of uncertainty on industrial production, for there was always at least one ICP setup that led to this conclusion. However, the fact that I am forced to refer to “at least one ICP setup” reveals the severe methodological shortcomings of the statistical testing procedure. As we have seen above, it is highly sensitive to the specific setup that is used and the final results can differ considerably, not only between the decoupled and the combined test, but also between the sum and the maximum test statistic. Unfortunately, the authors largely neglect these rather practical considerations and only make the attempt to provide a brief empirical explanation Pfister et al. (2019, pp. 1266).

5 Conclusion

In the course of this thesis, I tried to assess the causal relation between uncertainty and the business cycle. In order to turn this into a feasible undertaking and, in particular, to add some degree of unification and comparability to the diverging results in the literature, my strategy consisted in applying different econometric methods for causal inference to the same macroeconomic dataset.

From a theoretical point of view, I tried to employ a range of econometric methods, some of which are widely used in practice, to identify their specific benefits and disadvantages. From a more applied point of view, I used two different measures of uncertainty to capture different kinds of uncertainty—namely macroeconomic and forward-looking policy uncertainty—and to allow for an immediate robustness check.

Overall, it seems fair to say that the results of the present investigation confirm the state of the literature: There is no clear-cut evidence for one direction of causality, be it from the business cycle to uncertainty or vice versa. At best, the results obtained from testing for Granger-causality and from the SVAR analysis license the belief that the impact of business cycle activity as measured by industrial production on uncertainty is larger than the impact in the opposite direction. The ICP results partly confirm this finding and one might argue that they even refine it due to the fact that there is much evidence for an impact of industrial production on the MUI (figure 7) while there is none for an impact of industrial production on the EPU index (figure 9). Thus, as already stated in the introduction, uncertainty seems to be an ambiguous concept consisting of a variety of different characteristics which complicates its measurement.

Leaving the purely economic considerations behind, the central methodological result of this thesis seems to be that none of the econometric approaches considered above is entirely suited to identify true causal relations between macroeconomic quantities. We have seen that the notion of Granger-causality is about prediction rather than causation and the results indicate that uncertainty and the business cycle help predicting each other while casting doubt on whether this tells anything about their causal relation. The SVAR analysis brought about an insightful dynamic perspective by means of the IRFs, but identification is a major issue in this context as it introduces a great deal of flexibility for the researcher, which is why no strategy of identification will ever be accepted without controversy. Clearly, this also holds for the recursive identification scheme employed above. The issues with ICP go in a similar direction: Apart from the fact that its null hypothesis is formulated in a counterintuitive way, it is highly sensitive to the specific setup of the test.

Thus, to conclude, there are at least two important results that also offer avenues for future research: First, it is still necessary to investigate the relation between uncertainty and the business cycle to reach more robust conclusions. Second, causal inference on the macroeconomic level remains an important econometric challenge with plenty of room for methodological improvement.

References

Aldrich, J. (1989). Autonomy. *Oxford Economic Papers*, 41(1):15–34.

Angrist, J. D. and Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Bachmann, R., Elstner, S., and Sims, E. R. (2013). Uncertainty and Economic Activity: Evidence from Business Survey Data. *American Economic Journal: Macroeconomics*, 5(2):217–249.

Bachmann, R. and Moscarini, G. (2012). Business Cycles and Endogenous Uncertainty.

Baker, S. and Bloom, N. (2013). Does Uncertainty Reduce Growth? Using Disasters as Natural Experiments. *NBER Working Paper*, 19475.

Baker, S., Bloom, N., Davis, S., and Kost, K. (2019). Policy News and Stock Market Volatility. *NBER Working Paper*, 25720.

Baker, S., Bloom, N., Davis, S., and Terry, S. (2020). COVID-Induced Economic Uncertainty. *NBER Working Paper*, 26983.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Berger, D., Dew-Becker, I., and Giglio, S. (2020). Uncertainty Shocks as Second-Moment News Shocks. *The Review of Economic Studies*, 87(1):40–76.

Bernanke, B. S. (1983). Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics*, 98(1):85–106.

Bierens, H. J. (2007). *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press, Cambridge.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3):623–685.

Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives*, 28(2):153–176.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. J. (2018). Really Uncertain Business Cycles. *Econometrica*, 86(3):1031–1065.

Carriero, A., Clark, T. E., and Marcellino, M. (2018). Endogenous Uncertainty. *Federal Reserve Bank of Cleveland Working Paper*, 18-05.

Carriero, A., Clark, T. E., and Marcellino, M. (2020). Assessing International Commonality in Macroeconomic Uncertainty and its Effects. *Journal of Applied Econometrics*, 81:1–21.

Cesa-Bianchi, A., Pesaran, M. H., Rebucci, A., and van Nieuwerburgh, S. (2019). Uncertainty and Economic Activity: A Multicountry Perspective. *The Review of Financial Studies*, 115:1–53.

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

Fajgelbaum, P. D., Schaal, E., and Taschereau-Dumouchel, M. (2017). Uncertainty Traps. *The Quarterly Journal of Economics*, 132(4):1641–1692.

Fernández-Villaverde, J. and Guerrón-Quintana, P. (2020). Uncertainty Shocks and Business Cycle Research. *NBER Working Paper*, 26768.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form. *Journal of Econometrics*, 123(1):89–120.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438.

Haavelmo, T. (1944). The Probability Approach in Econometrics. *Econometrica*, 12(Supplement):iii–vi + 1–115.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Hassler, J. A. (1996). Variations in Risk and Fluctuations in Demand: A Theoretical Model. *Journal of Economic Dynamics and Control*, 20(6-7):1115–1143.

Hayashi, F. (2001). *Econometrics*. Princeton University Press, Princeton, NJ.

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

Kilian, L. and Lütkepohl, H. (2018). *Structural Vector Autoregressive Analysis*. Cambridge University Press, Cambridge, 1st edition.

Lehmann, E. L. and Romano, J. P. (2008). *Testing Statistical Hypotheses*. Springer, New York, 3rd edition.

Ludvigson, S., Ma, S., and Ng, S. (2020). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? *American Economic Journal: Macroeconomics*, Forthcoming.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press, Oxford.

Nakamura, E. and Steinsson, J. (2018). Identification in Macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86.

Peters, J., Bühlmann, P., and Meinshausen, N. (2016). Causal Inference By Using Invariant Prediction: Identification and Confidence Intervals. *Journal of the Royal Statistical Society: Series B*, 78(5):947–1012.

Pfister, N., Bühlmann, P., and Peters, J. (2019). Invariant Causal Prediction for Sequential Data. *Journal of the American Statistical Association*, 114(527):1264–1276.

R Core Team (2019). R: A Language and Environment for Statistical Computing.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, Egbert H, Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J. (2019). Inferring Causation From Time Series in Earth System Sciences. *Nature Communications*, 10(1):1–13.

Saijo, H. (2017). The Uncertainty Multiplier and Business Cycles. *Journal of Economic Dynamics and Control*, 78:1–25.

van Nieuwerburgh, S. and Veldkamp, L. (2006). Learning Asymmetries in Real Business Cycles. *Journal of Monetary Economics*, 53(4):753–772.

A Further Graphs

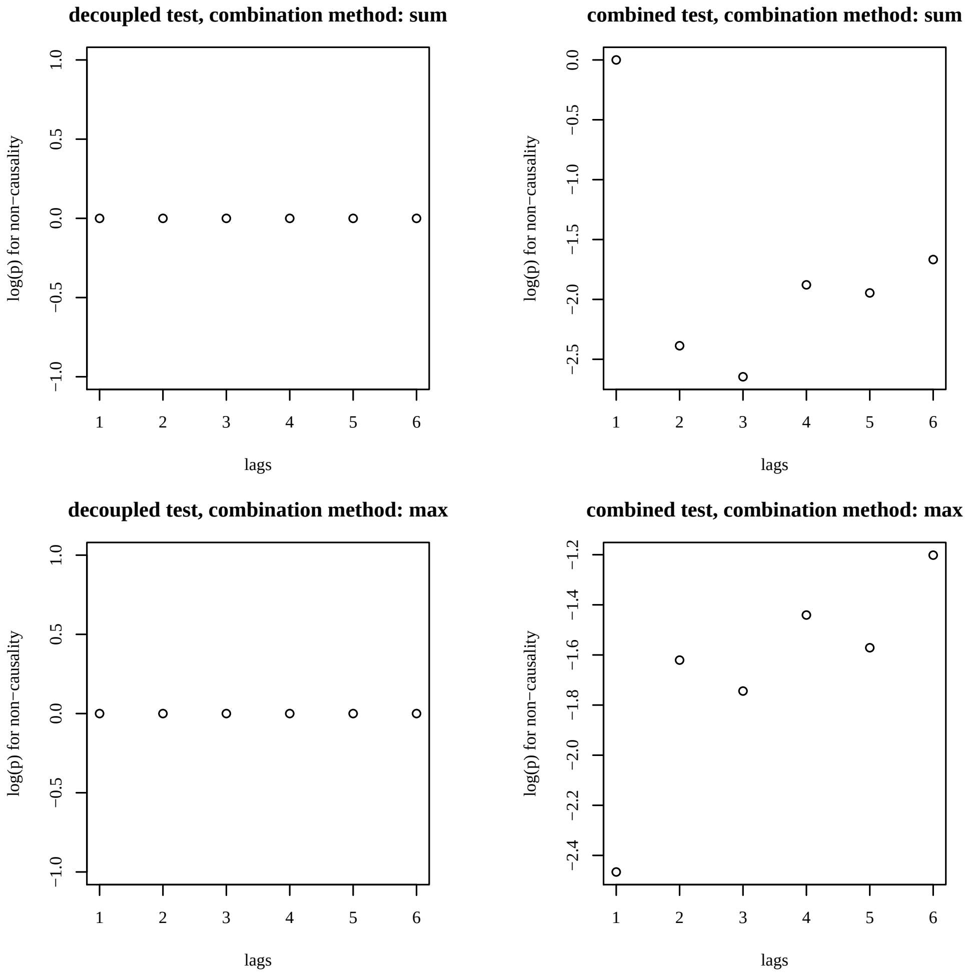


Figure 9: Results from testing causality from log-differences of U.S. industrial production to the EPU index using the decoupled test (left column) and the combined test (right column) as well as the methods in (49) (top row) and (48) (bottom row) to combine individual test statistics for lag orders $p \in \{1, \ldots, 8\}$.

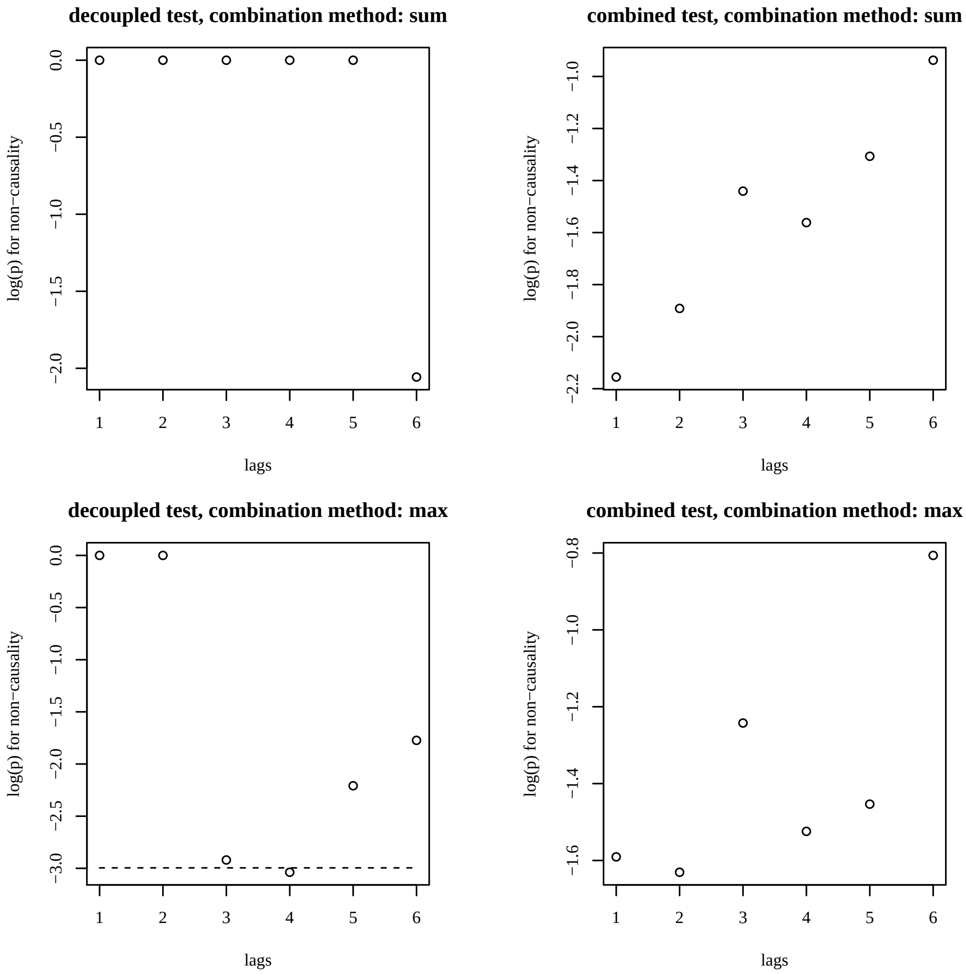


Figure 10: Results from testing causality from the EPU index to log-differences of U.S. industrial production using the decoupled test (left column) and the combined test (right column) as well as the methods in (49) (top row) and (48) (bottom row) to combine individual test statistics for lag orders $p \in \{1, \ldots, 8\}$.

B Data and Code

The data used in the empirical analyses were obtained from different sources, all of which are publicly accessible online. The details are summarized in the table below.

Description	Code	Source
Macro Uncertainty Index	none	Link
EPU index	none	https://www.policyuncertainty.com/us_monthly.html
U.S. industrial production	INDPRO	https://fred.stlouisfed.org/series/INDPRO
Federal Funds Rate	FEDFUNDS	https://fred.stlouisfed.org/series/FEDFUNDS
U.S. employment	PAYEMS	https://fred.stlouisfed.org/series/PAYEMS
S&P 500	none	https://finance.yahoo.com/quote/%5EGSPC/history?period1=473299200&period2=1598918400&interval=1mo&filter=history&frequency=1mo

Table B1: Sources for the data used in the empirical part of the paper.

The SVAR analyses were conducted using Matlab running under version 2017b. All other analyses were conducted using `R` running under version 3.5.3 (R Core Team 2019) and, in addition, using the following packages:

Name	Reference	URL
`parallelsugar`	VanHoudnos, N. (2019). *parallelsugar: mclapply() syntax for Windows machines*. R package version 0.0.0.2.	`https://github.com/nathanvan/parallelsugar`
`seqICP`	Pfister, N. and Peters, J. (2017). *seqICP: Sequential Invariant Causal Prediction*. R package version 1.1.	`https://CRAN.R-project.org/package=seqICP`.
`tidyverse`	Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1.	`https://CRAN.R-project.org/package=tidyverse`

Table B2: `R`-packages used in the empirical part of the paper.