WORKING PAPER 01/2021

Machine Learning and Central Banks:

Ready for Prime Time?

Hans Genberg

and

Özer Karagedikli

The South East Asian Central Banks (SEACEN) Research and Training Centre (80416-M)

Kuala Lumpur, Malaysia

WORKING PAPER 01/2021

Machine Learning and Central Banks: Ready for Prime Time?

Hans Genberg

and

Özer Karagedikli

March 2021

Table of Contents

Machine Learning and Central Banks: Ready for Prime Time?*

Hans Genberg[a] and Özer Karagedikli[b,c]

[a] Asia School of Business
[b] South East Asian Central Banks (SEACEN) Research and Training Centre
[c] Centre for Applied Macroeconomic Analysis (CAMA)
[d] University of Marburg

March 25, 2021

Abstract

In this article we review what machine learning might have to offer central banks as an analytical approach to support monetary policy decisions. After describing the central bank's "problem" and providing a brief introduction to machine learning, we propose to use the gradual adoption of Vector Auto Regression (VAR) methods in central banks to speculate how machine learning models must (will?) evolve to become influential analytical tools supporting central banks' monetary policy decisions. We argue that VAR methods achieved that status only after they incorporated elements that allowed users to interpret them in terms of structural economic theories. We believe that the same has to be the case for machine learning model.

JEL Codes: J31, J64

*The views expressed here are the views of the authors and do not necessarily represent the views of the South East Asian Central Banks Research and Training (SEACEN) Centre, and its member central banks or monetary authorities. We thank Jose Eduardo P. Sto. Domingo, Shien Jin Ong, Ole Rummel, Benjamin Wong and participants at the SEACEN Centre seminar for comments. Correspondence: Hans Genberg – hans.genberg@asb.edu.my, Özer Karagedikli – ozer@seacen.org.

1 Introduction

Machine learning is the new 'car' on the street, and as with most things that are new, it is flashy, popular and full of excitement. Central banks, or at least some central bank economists, appear to share this excitement, and as a result, a number of central banks have invested resources in machine learning.[1] Is the excitement warranted?

Advances in artificial intelligence research has transformed many tasks previously carried out 'manually'. Areas include interpreting x-ray images, translation, and facial recognition to mention processes that already use machine-learning methods successfully. Artificial intelligence has also outperformed humans in playing highly complex games such as chess and Go, and holds the promise of being able to give us driver-less cars. Similar successes are also seen in banking applications, for example in credit scoring (Gambacorta et al. (2019)). In addition, machine learning is increasingly used to inform public policy decisions (Athey (2015)).

Little wonder that central banks ask themselves whether these underlying methods can be used to cater to the needs of central banks. After all, if computers can be trusted to drive cars, trucks and buses safely on busy streets filled with cyclists and pedestrians, can they not also be trusted with some aspects of the conduct of monetary policy? In this article, we will attempt to answer this important question.

Our analysis starts by providing answers to the following two questions: First, what does a monetary policy making authority 'want' from analytical models, and second, how have existing models, particularly VARs, evolved to serve these needs? In other words, we will use the evolution of VARs as a lens through which we will assess whether machine learning methods can be used to answer the questions that the central banks have.

What does a monetary policy making authority or a central bank require from an empirical tool? We argue that a monetary policy making institution would like to i) describe and summarise macroeconomic data, ii) make forecasts, iii) conduct

[1] The recent Federal Reserve Board, the Bank of England and the King's College Conference on "Modelling with Big Data Machine Learning: Measuring Economic Instability" is one example of this.

structural analysis, including risk and scenario analysis, and iv) communicate its decisions/analyses.

We believe that the VAR literature made a critical contribution to the way the central banks think about data, forecasting, and structural analysis. We will therefore use the developments and achievements of VAR methodology, and the challenges it has had to overcome, as a lens through which we can assess the potential usefulness of machine learning for the central banks.

To avoid misunderstanding, we would like to make it clear that the authors of this article are not machine learning experts. However, we do have experience in the use of VARs in central banking, and we will use this experience, and our understanding of the current state of machine learning, to speculate about how these tools need to evolve to challenge and perhaps overtake VARs as a dominant analytical tool for central banks.

We also want to emphasize that, in this article, we are focusing only on the potential use of machine learning in the monetary policy functions of the central bank. Machine learning and artificial intelligence have been used more actively in other aspects of central banking, for example: financial stability, crises forecasting and banking supervision.[2] We will not comment on these important applications for machine learning in this paper.

Almost two decades ago, Adrian Pagan argued that macroeconomic models used in central banks face a trade-off between their theoretical consistency or data-consistency (figure 1). Given that central banks have a fixed amount of resources (staff) available this trade-off could be interpreted as a production possibility frontier. One could also think about this trade-off chart as the preferences of policymakers or researchers between theory-driven and data-driven approaches. The models that we consider to be on the frontier are listed in the figure.[3]

[2] Fouliard et al. (2020), for example, predicts financial crises with machine learning algorithms. Babii et al. (2020) is a recent time-series application to nowcasting the US GDP growth. For a detailed treatment of time-series forecasting with machine learning see Masini et al. (2021).

[3] One might argue that old fashioned econometric models of the 1960s and the 1970s are perhaps inside the frontier as opposed to being on the frontier.

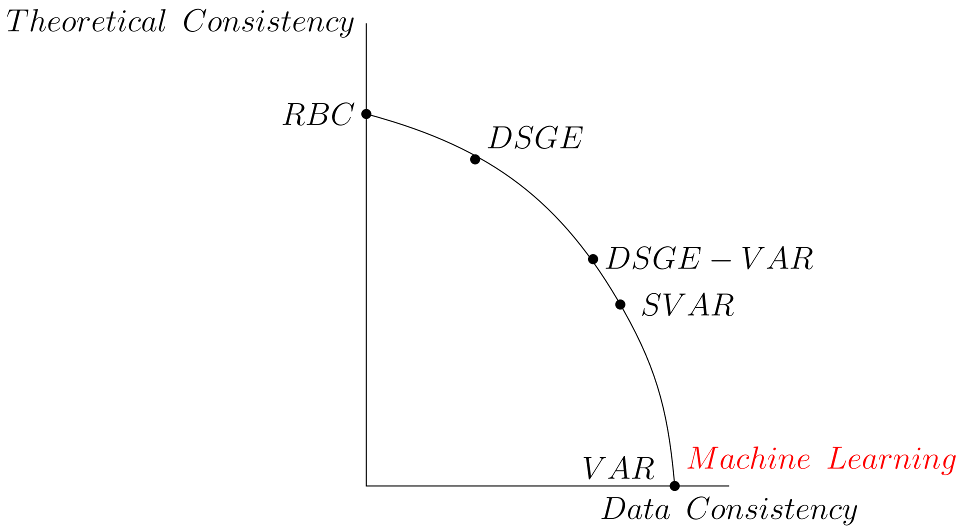

Figure 1
The Pagan Frontier

Models like the unrestricted VARs are located close the lower right hand part of the frontier reflecting their data-heavy feature. Close to the upper left part we find Real Business Cycle (RBC) and Dynamic Stochastic General Equilibrium (DSGE) models that are typically not estimated in the conventional sense. Rather, they are calibrated using estimates of structural parameters obtained from other studies.

Over the past four decades VARs have evolved substantially, so that they now cover a wider spectrum in the Pagan frontier. The development of the range of VARs that incorporate restrictions to give them structural interpretations (SVARs) have perhaps pushed them towards the middle of the Pagan frontier. For example, it has been argued that DSGE-VARs provide a framework with which one could evaluate the DSGE models (Del Negro and Schorfheide (2006)). In other words, the development of the VAR has been rather dynamic not static.[4]

The big question, we believe, is whether ML models that currently are located on the data-heavy corner of the frontier, can move towards the middle, and what modifications need to be introduced to enable them to do so. In other words, can they combine

[4]See (Kilian and Lutkepohl (2018) and Ramey (2016) for a reviews of identified VARs for example.

good forecasting performance with an ability to provide insights on more structural aspects of the macroeconomy that central banks need?

The remainder of this article is structured as follows: Section 2 sets out what we think central banks need from a model. Section 3 summarises our understanding of current generation machine learning models and how they might be used by central banks. Section 4 describes briefly the development of the VAR literature, and Section 5 discusses how well the VARs have served in solving or answering central banks' needs. Section 6 discusses whether the Machine Learning can develop into something similar or better than VARs. Section concludes.

2 Central Bank's problem

We propose a list of what central banks need from models by extending the Stock and Watson (2001)'s description of what macroeconometricians at policy institutions do. Stock and Watson (2001) argued that they do four things: 1) describe and summarize macroeconomic data, 2) make macroeconomic forecasts, 3) quantify what we do or do not know about the true structure of the macroeconomy, and 4) advise (and sometimes become) macroeconomic policymakers.

We adapt this classification to central bank policy making. Further, we argue that, to a first approximation, policy making involves the following steps, and models play a role in every one of them:[5]

1. Central banks summarise and analyse data,
2. They forecast the key macroeconomic variables,
3. They conduct risk analysis and balance of uncertainties.
4. They do structural/causal analysis, as well as scenario analysis.

[5]These five points are also in the spirit of Sims (1980), where he argued that the usefulness of VARs could come from three fronts: (1) forecasting macroeconomic variables; (2) designing and evaluating economic models; (3) evaluating the consequences of alternative policy actions.

5. They take decisions and communicate and justify these decisions vis-a-vis the public.

In the 1970s, these four tasks — data description, forecasting, structural inference and policy analysis — were performed using a variety of techniques. These ranged from large models with hundreds of equations to single-equation models that focused on interactions of a few variables to simple univariate time series models involving only a single variable. As we note in Section 4, dissatisfaction with the performance of these models let Christopher Sims proposing a new class of models, VARs, in his seminal Econometrica article published in 1980 (Sims (1980)).

In the remainder of the paper, we will be arguing that the VARs have served well in each of these five steps listed above. We will also argue that to be helpful to policymakers in central banks, Machine Learning methods must be able to perform well in these areas.

3 What is Machine Learning?

3.1 Machine Learning is so far principally a prediction tool, but change may be coming

Machine Learning (ML) has its origins in computational statistics. Its primary concern has been the use of algorithms to identify patterns or interrelationships that exist in data, and use these patterns in prediction. While the algorithms used in ML can be common techniques such as ordinary least squares regressions, they can also be more complex methodologies such as decision trees, clustering algorithms, and deep learning multi-layer neural networks.

Currently, the principal use of machine learning in economics has been for prediction This is particularly the case in macroeconomic applications (Kuhn and Johnson (2013)). The growing popularity of ML comes from its ability to uncover complex patterns in the data that have not been pre-specified a priori. This flexibility is in sharp contrast with approaches to forecasting traditionally used in central banks. In forecasting inflation, for example, researchers usually start with a structure that comes from theoretical considerations linking inflation to a set of causal determinants. For analytical and empirical tractability, the structure is often represented by a model that is linear in the underlying variables. But if forecast accuracy is the prime goal, this pre-specified linear structure can be a weakness, and the advantage of machine learning is that it does not have to impose it on the estimation of the model and hence on the forecast. For example, deep learning neural network algorithms do not impose any particular functional form of relationship between the explanatory variables and the forecast target. Instead this functional form is the outcome of the network algorithm.[6] The variables that are included in the forecasting model, be it a simple linear regression or a complex deep learning neural network, are still determined by the researcher, however.

While the main use of machine learning in economics is prediction, recent developments have shown that, in certain contexts, it is also possible to design algorithms that will allow causal inference. As Athey (2015) states "prediction and causal inference are distinct (though closely related) problems. Outside of randomized experiments, causal inference is only possible when the analyst makes assumptions beyond those required for prediction methods, assumptions that typically are not directly testable and thus require domain expertise to verify."[7]As we will argue below, for central bankers to make full use of machine leaning techniques, it is important that causal analysis also be possible in non-experimental contexts.

[6]Note that the algorithms themselves may favour some functional forms over others by construction, a feature referred to as inductive bias (Kelleher (2019)).

[7]See Pearl (2018) and Pearl and Mackenzie (2018) for detailed treatments of causal inference in machine learning.

3.2 ‘Training’ vs. ‘Validation’ and the need for large amounts of data

Since machine learning algorithms impose relatively little a priori structure, there is a potential for overfitting whereby the algorithm will continue to add complex relationships until it perfectly explains the data it is given. Some restraints must therefore be imposed. This can be done both at the stage of designing the algorithm that is used to explain the data and at the stage of validating the output of the algorithm.

For example, the number of nodes of a decision tree determines its complexity and ability to explain the data it is confronted with. The risk of overfitting - explaining not only the systematic relationships imbedded in the data but also random noise resulting from errors in measurements, variations in the data due to idiosyncratic shocks that are not forecastable but nevertheless influence the outcome – can be reduced by limiting ex ante the complexity of the algorithm used in the analysis.

In addition to limiting the complexity ex ante, the analyst will typically divide the data sample into two parts, a training sample and a validation sample. The former is used to generate a model that best explains the data in the sample, while the latter is used to check how well the model explains data that are assumed to be driven by the same data-generating process but not used in the initial training phase. If the model’s ability to explain the validation data set is significantly worse than its ability to explain the training data set, it is likely that over-fitting the data is a problem. The fundamental parameters of the algorithm are then adjusted and the training and validation process is repeated. The iteration process will end when the model’s explanation of the training data set is not significantly better than its ability to explain the validation data set.[8]

The practice of using part of the available data as a training ground for the ML algorithm and part of it as a validation implies that machine learning modelling is best suited to situations where we have a large number of observations on the events

[8]This modelling process is familiar to the traditional econometricians as in-sample estimation and out-of-sample forecast evaluation.

that we are trying to forecast. This can potentially be a problem for forecasting of inflation, output fluctuations, and other macroeconomic variables at business cycle frequencies that central banks are particularly concerned with. In fact, Athey and Imbens (2019) argue that the methods developed in the ML literature have been particularly successful in “big data” settings.

Although there is no universally agreed definition, ‘big data’ generally refers to very large structured and/or unstructured data sets containing tens of thousands of observations on bank customers, holders of insurance policies, users of online payment platforms, etc. Textual data can also be digitized and made available for computer-aided analysis of content. Examples include the digitization of documents containing the latest financial regulations so that these can be incorporated in compliance routines; newspaper articles to aid in the search for indicators of economic uncertainty; or useful information for regulators.

These sources of information have, of course, existed for a long time, but it is only with the advent of inexpensive data storage facilities, huge increases in computing power, and new analytical techniques capable of dealing with very large data sets that the full benefits of big data have been realized.

3.3 Cars with ML drivers vs. Economies with ML policy decisions: Similarities and differences.

If we can design cars that navigate safely using machine learning algorithms, why should we not be able to design monetary policy rules based on the same principles? After all, a self-driving car must be able to observe the environment in real time (including possibly idiosyncratic behaviour of pedestrians, motorcyclists, and other vehicles), interpret what it observes in terms of objectives that it wants to achieve (arriving at a pre-determined destination, avoiding accidents, adhering to driving rules and conventions, etc.), and decide on adjustments to be made to driving direction, speed, avoidance manoeuvres, etc.

Similarly, in order to conduct monetary policy, the central bank management needs to obtain observations on the current state of the economy (the current and probable future rate of inflation and growth, for example), interpret this information in relation to the objectives it is charged with achieving (macroeconomic and financial stability in particular), and take the appropriate policy action (changing policy interest rates, intervene in the foreign exchange market, etc.).

At this high level of generality, the tasks of creating a driver-less car seems to be similar to designing a machine learning model to conduct monetary policy. However, as we dig deeper, significant differences are revealed that need to be recognized and addressed before monetary policy can be driven, at least partially, by digital means.

Consider first observing the economy. There have been considerable advances in now-casting methods to provide a detailed picture of the current state of inflation, growth and other variables relevant for monetary policy decisions. These methods typically make use of a wide variety of indicators and sophisticated analytical approaches. Machine learning tools are also being used to provide forecasts of these same variables. However, as we will discuss in some detail later in the paper, determining policy actions typically requires more than knowledge of the current state of the economy and having a forecast of its future path. For example, it is often necessary to know how the economy has arrived at its current state, and what the underlying determinants of the forecasted future path are: are they driven by supply or demand shocks, by implicit forecasts of changes in economic activity and inflation abroad, or by projected changes in consumer behaviour? The appropriate policy response will depend on the context. In contrast, the driverless car does not really need to know whether the pedestrian who is crossing the street is coming from a movie theatre or the grocery store, or, at least to a first approximation, whether the bend in the road ahead will be followed by a straightaway or another bend. In other words, narratives that provide a context are important for monetary policy making, and here the 'black-box' aspects of current generation machine learning models constitute a hurdle.

Another difference between the driver-less car example and the challenges facing the central bank, is that the time pattern of the responses to the actions taken by the driving algorithm and the actions taken by the central bank are quite different, in the parlance of monetary policy, the transmission mechanism is different. If a pedestrian crosses the street, the algorithm in the driver-less car will initialize actions to slow down or come to a complete stop. The car will react virtually immediately to the actions. For the central bank, on the other hand, if the ML algorithm initiates actions to slow the economy by raising the policy interest rate, the impact on the economy will be felt only several quarters later. It would be more like trying to stop a super tanker than stopping a car. To be sure, the algorithm can build this reaction lag into the decision-making structure, but this adds another layer of complexity in the central banking example. The algorithm must also learn how the economy reacts to a policy rate increase, whereas the algorithm for the driverless car can incorporate the mechanical relationships directly. But understanding the reaction of the economy to a policy change requires either having a reliable model of the transmission mechanism or training the algorithm on past experiences with policy rate changes. In the latter case, the required data may be insufficient to allow for a relatively unstructured machine learning approach.

Finally, there is the communication aspects of monetary policy decisions. The effects of a particular monetary policy are likely to depend on how well it is explained and justified. The policy maker needs to provide a plausible causal story for both the assessment of the current state of the economy and the policy decisions taken. It is not likely to be sufficient to just refer to a computer algorithm. In contrast, the driver-less car does not have to provide a story about why it stopped for the pedestrian, as long as it did so in a timely manner. [9]

In the following sections we will first elaborate on how central banks have tried to deal with the challenges of formulating policy based on a mixture of theoretical and empirical modelling, before discussing how machine learning needs to (will?) evolve to provide more support for monetary policy decisions of central banks.

[9] There may be cases where (the owner of) the driver-less car may have to provide an explanation of the car's behaviour. For example, if stopping for the pedestrian in time is impossible, and the car swerves to avoid her and in the process hits an oncoming motorcycle, then the owner may be required to provide an explanation to a court why it took that decision. {See Boudette (2017), "Tesla's self-driving system cleared in deadly crash." New York Times.

4 The Development of the VARs at a Glance

More than four decades ago Sims (1980) provided a new macroeconometric framework in a paper with the provocative title 'Macroeconomics and Reality'. The motivation was a view that the econometric models used predominantly at the time were unreliable because they relied on "incredible identification" (Sims (1980), p. 1) based on theories which in Sims' judgment were not justifiable.[10] Sims introduced an alternative methodology, estimation and analysis based on vector autoregressions (VARs), designed to 'let the data speak' more freely about the relationship between major macroeconomic variables without being constrained by a priori restrictions.

Specifically, in general terms, the VAR system proposed is an n-equation, n-variable linear model in which each variable is in turn explained by its own lagged values, plus current and past values of the remaining n - 1 variables.

This simple framework provided a systematic way to capture rich dynamics in multiple time series, and the statistical toolkit that came with VARs was easy to use. As Sims (1980) and others argued in a series of influential early papers, VARs held out the promise of providing a coherent and credible approach to data description, forecasting, structural inference and policy analysis.

Will models based on ML and AI be able to extend Sims' 'let-the-data-speak' approach even further by doing away with the linear structure? Before we attempt to answer this question, we briefly review the evolution of the use of VAR models in empirical macroeconomics in general and in central banks in particular.

[10]"...claims for a connection between these models and reality – the style in which "identification" is achieved for these models – is inappropriate, to the point at which claims for identification in these models cannot be taken seriously." (Sims (1980) p. 1).

4.1 From Linear Dynamic Stochastic General Equilibrium Models to VARs and back

4.1.1 From structural models to VARs

Every linear DSGE model can be transformed into a VAR representation. Let Y_t be a vector of endogenous variables represented by the model. The linear DSGE model can then be written as

$$AY_t = BY_{t-l} + U_t \tag{1}$$

where for simplicity the dynamics is restricted to one lag. The elements of U_t are shocks to each of the structural equations, and they are assumed to be independent of each other.[11] A and B are structural parameters derived from the underlying theory. To convert this structural model into a VAR, simply pre-multiply both sides of the equation 1 by the inverse of the matrix A.

$$Y_t = CY_{t-1} + V_t \tag{2}$$

where $C = A^{-1}B$ and $V_t = A^{-1}U_t$

Four observations follow from the equation 2: First, the reduced form, equation 2, can be estimated efficiently equation by equation by ordinary least squares. Second, the parameters in C are linear combinations of the structural coefficients in A and B. Third, even if the elements in U are independent of each other, the elements in V will not be. Fourth, and most crucially for our purposes, it is in general not possible to infer the values of the structural coefficients from the reduced form parameters. Similarly, it is not possible to infer the nature of the structural shocks (the Us) from the properties of the reduced form errors (the Vs).

[11] In general, we use the term shock to refer to innovations (i.e., the residuals from a reduced-form vector autoregressive (VAR) model or to the structural disturbances in a simultaneous equation system). Bernanke (1986) argues that shocks should be“primitive" exogenous forces, not directly observed by the econometrician which buffet the system and cause oscillations. Because these shocks are primitive, i.e., they do not have common causes, it is natural to treat them as approximately uncorrelated”. See also Christiano et al. (1999) for a discussion on the definition of shocks in macro.

When the use of the VAR is for pure forecasting, these observations are not problematic. Indeed, in its simple form VARs are frequently used for this purpose. As we will be discussing later in the paper, more elaborate variants with time-varying parameters, stochastic volatility and a number of other extensions have also been developed and are frequently used by central banks.

But if we want to 'tell stories' of the type 'This is what our estimates tell us about the consequences of a unexpected increase in the policy interest rate' (assuming this variable is an element of Y), then the VAR estimates are not sufficient. We would need to reverse engineer the VAR to identify the underlying structural equations. This 'identification problem' has spurned a large literature.

4.1.2 Giving VARs a structural interpretation

To give a VAR a structural interpretation, we need to impose restrictions on the model so that we can recuperate the elements of A and B from the estimates of C, and to ascertain the nature of the Vs from the estimated Us. A very large literature has evolved to do just that.

The method originally proposed by Sims was to assume a temporal ordering through which the shocks had an impact on the endogenous variables. For example, an unanticipated change in the short term monetary policy interest rate would affect a longer term interest rate before it would have an impact on aggregate demand or any other variable in the model. Similarly, an unexplained shock to aggregate demand would affect output before it would change inflation. If such a temporal ordering was imposed on variables in the economy, it would be possible to recuperate the original structural parameters in the model as well as the time pattern of the structural shocks. That identification would then make it possible to trace out the effects of any shock in the system on all of the endogenous variables,[12] i.e. to 'tell stories' of the type mentioned in the last paragraph of the previous subsection.[13]

[12] In the jargon of the literature this entailed calculating 'impulse response functions'.

[13] Given the centrality of understanding shock propagation in the economy, impulse response functions are key to understanding propagation of structural shocks. In the more recent decade or two, since Jordà

Other identification schemes followed; some based on restrictions on the contemporaneous interactions between the endogenous variables in the structural model (the A matrix);[14] others on the sign pattern of the impulse response function.[15] Yet others achieved identification based on the theoretical presumption that a particular shock would have no long run effect on a specific variable,[16] based on theoretical 'signs', the information coming from high frequency data or heteroscedasticity patterns in the data.

The goal of these identification strategies was always the same; to be able to conduct causal analysis, to understand and importantly, be able to communicate to stakeholders how monetary policy affects the economy.

4.1.3 Incorporating large data sets in VARs

The number of variables that can be directly included in a VAR model is frequently constrained by the available number of observations. A n-variable model that includes four lags of each variable will have 4n+1 free parameters to estimate per equation. If the frequency of observation is quarterly, then a four-variable system will require forty years of observations if the rule-of-sum requirement of ten observations per parameter is respected. For many economies this is very demanding.

Two types of solutions to this problem have been proposed in the literature. One is to reduce the effective number of parameters by 'soft' constraints on the lag patterns associated with each variable in the system. This can be achieved by applying the so-called 'Minnesota priors' during the estimation process.[17]

(2005), local projections have become an increasingly widespread alternative econometric approach to understanding the shock propagation. Recently, Plagborg-Møller and Wolf (forthcoming) showed that the impulse responses from these two approaches are the same in large samples. However, we will not be delving into this literature.

[14]For example that inflation does not depend contemporaneously on the size of aggregated demand, or that export volumes do not respond contemporaneously to a change in the real exchange rate.

[15]That aggregate demand should respond negatively to an unexpected increase in the monetary policy interest rate, for example.

[16]That monetary policy would have no long-run effects on real output.

[17]The Minnesota prior was introduced in Doan et al. (1986) as a shrinkage prior for autoregressive parameters in vector autoregressive (VAR) models. More recently, Giannone et al. (2015) shows that

Another solution involves reducing the number of variables by combining them into 'factors' and using these factors in the model instead of the underlying variables themselves. The resulting VARs in these cases have been given the name factor-augmented VARs (FAVARs).[18] For example, to study the propagation mechanism of monetary policy,[19] it might be desirable to include a number of financial variables to capture different channels of transmission. But doing so would render the size of the VAR too big to be estimated. Instead, the analysts combines the financial variables into a 'financial factor' and uses this in the VAR. Usually the factor is a linear combination of the underlying variables, often based on principal component analysis or dynamic factor models.

Another example of reducing the number of variables comes up in the context of open economy interactions with the rest of the word where, instead of letting all variable for all countries enter every equation, foreign variables are combined into foreign 'factors' representing 'world' prices, 'world' interest rates, 'world' output for example. Mumtaz and Surico (2009) for example use a data from a panel of 17 industrialized countries to investigate the international transmission mechanism.[20]

5 How well the VARs have served in 'solving' the Central Bank's problem?

5.1 Forecasting and Risk Analysis

The standard VARs and their extended variants such as time-varying parameter VARs, time-varying stochastic volatility VARs, factor-augmented VARs and so on have been

the optimal choice of the informativeness of priors can be 'treated as additional parameters to reduce the number and importance of subjective prior choices in the setting of the prior.'

[18] See Stock and Watson (2016) for the implications of factors models for the SVARs.

[19] See Bernanke et al. (2005) for an example of monetary policy analysis with factor-augmented vectorautoregression (FAVAR).

[20] Another approach to examine the open economy interactions is the Global Vector AutoRegressive (GVAR) methodology of Dees et al. (2007) that provides a VAR based interactions between a large number of countries.

extensively used in forecasting. Their forecasting performances have been found to be good and most central banks use them. One advantage of the VARs in forecasting is their system nature and hence every endogenous variable can be forecast with the same model. This is in contrast with machine learning models which, to the best of our knowledge, typically predict one variable at a time. The VAR forecasts have been successful in conducting point forecasts as well as density forecasts (Carriero, Clark, and Marcellino (2020)). Therefore, we can say without any qualifications that the VARs have been an important part of the forecasting toolkit of central banks.

Recently, monetary policy makers have also become interested in the tail-risks and taking into account the potential tail outcomes in forecasting. Building on finance literature in assessing tail risks in asset prices and returns, this line of research has emphasized the importance of tail risks in macroeconomic policymaking. Although this literature has started with the risks of significant declines in GDP growth, and has relied on quantile regression methods to estimate tail risks (Adrian et al. (2019)) the applications of the framework have widened to other important variables such as house prices and capital flows.[21] Recently, Chavleishvili and Manganelli (2019) extended this framework in a VAR context.[22] Carriero, Clark, and Marcellino (2020) and Carriero, Clark, and Massimiliano (2020) show that Bayesian VARs with stochastic volatility are able to capture tail risks in macroeconomic forecast distributions and outcomes. Bayesian VARs, which have been commonly used for point and density forecasting, are able to capture more time variation in downside risk as compared to upside risk for output growth consistent with the results of quantile regression of the earlier research. Moreover, the Bayesian VARs come with additional gains in the form of standard point and density forecasts.

The VAR literature has also been responsive to the most recent challenges with the post COVID-19 pandemic sample. With new data that has not been seen for the past century, are there implications for VAR forecasts? Primiceri and Tambalotti (2020)

[21]For example, Gelos et al. (2019) use the similar quantile-regressions framework to understand the asymmetric tail-risks to capital flows in emerging markets.

[22]Other avenues of work in considering tail risks are the copula modeling (e.g., Smith and Vahey (2016) and Loaiza-Maya and Smith (2020), and the copula-based forecast combinations; Karagedikli et al. (2019).

have recently shown that the VAR forecasts can, indeed, handle this massive anomaly in the data.

5.2 Structural and Scenario Analysis

The traditional econometrics or estimation used in central banks are often times concerned with questions beyond simple out-of-sample forecasting. This is an important difference between traditional machine learning applications and econometrics . In many, arguably most cases, central banks are interested in "average treatment effects" or other causal or structural relationships.[23] As we have already noted, the challenge for machine-learning models is to be able to make causal inferences also in contexts important for monetary policy decisions.

The structural VARs have now been used in identifying a large number of structural shocks and their causal effects including monetary policy shocks, demand shocks, commodity price shocks, oil market shocks and so on. Moreover, the structural VARs are used in distinguishing between different theoretical structures.

Structural analysis also includes scenario analysis which central banks often do. For example, a central bank may be interested in the consequences for the economy of alternative paths for the policy interest rate; holding it constant for an extended period before increasing it to a higher level compared to increasing immediately but doing so gradually until it reaches the new higher level. VARs can be and have been used to carry out such scenario analysis or conditional forecasting. The earlier applications of conditional forecasting includes Doan et al. (1986) and Waggoner and Zha (1999). Leeper and Zha (2003) for example, came up with a framework for computing and also evaluating forecasts of endogenous variables within a SVAR, conditional on hypothetical paths of monetary policy. Leeper and Zha (2003) is a good example of the VARs moving along the Pagan frontier as we discussed in the introduction: from a purely data driven objective, they also become more consistent

[23] See Abadie and Cattaneo (2018) for a survey. For more recent macro surveys on the issue, see Fuchs-Schündeln and Hassan (2016), Ramey (2016) and Nakamura and Steinsson (2018).

with theory. Their framework is based on the theoretical model that reports “when linear projections are reliable even though policy switches from one regime to another.”

Structural VARs are also used in understanding the history in terms of the drivers of the business cycles. With historical decompositions one can get a view on the historical drivers of the business cycle dynamics

5.3 Communication

Over the past few decades, transparency, accountability and the need for clear and timely communication have come to be widely recognized as essential components of successful central banking. In part, this development is the result of demands by the general public for greater transparency and accountability in government institutions.

In part, more active communication by central banks also stems from their attempts to manage and guide market expectations and thereby increase the effectiveness of monetary policy. For example, stating in a monetary policy briefing that ‘the short-term policy interest rate will be maintained at a low level for a considerable period’ can be part of a strategy to influence longer-term market interest rates that typically have a stronger effect on consumption and investment decisions.

A component of a central bank’s communication consists of explaining why and on the basis of what information a policy decision has been taken. This is typically followed by an explanation of how the policy change is expected to influence the economy and lead to desirable outcomes.

In order for the communication to be credible, the policy statement must therefore not only make it clear which variables the central bank is primarily focused on, but it must also explain the economic mechanisms that the central bank relies on to achieve its objectives. All this requires having a clear causal structure in mind. This structure can be a calibrated theoretical model that reflects the central bank’s understanding of how the economy works, or it can be a more data-driven model, such as a VAR, which has been given a structural interpretation using methods described above. But crucially, the economic reasoning behind the central bank's policy decision must be clear and transparent for it to be credible in the eyes of the public.[24]

6 Can Machine Learning do the same?

In section 2 we have characterized the central bank's problem as one of summarizing and analysing data; forecasting key macroeconomic variables; conducting risk analysis and balance of uncertainties; carrying out structural/causal as well as scenario analysis; and finally communicating and justifying policy decisions vis-a-vis the public. We then argued how VAR methodology has evolved and become a work-horse to assist central banks in responding to these tasks (Sections 4 and 5). In this section we ask how machine learning may need to evolve to become a complementary or alternative methodology guiding monetary policy decisions and communication.

6.1 Forecasting and Risk Analysis

As summarised in Mullainathan and Spiess (2017) , and as we have argued above, Machine Learning in its current form, is essentially about prediction. We argue that machine learning's advantage in short-term forecasting is significant, although it is subject to the similar issues as classical econometric techniques; for example, structural breaks in the data with an event like the COVID-19 pandemic, is likely to have implications for the training sample and forecasting.

Perhaps forecasts from more traditional models can be compared with those from ML algorithms, and, to the extent that the latter are superior, it may be possible to tweak the traditional models so 'having a story to tell'.

Risk analysis has become commonplace in traditional econometrics with the use of quantile regressions and density forecasts. We are not aware of similar applications in

[24] Blinder et al. (2008) argue that transparency is closely linked to central bank credibility: it simply means matching deeds to words.

the machine learning sphere, but it would seem that carrying out the equivalent of quantile regressions should be possible with extant machine learning tools. Estimating and forecasting entire densities appear to us to be more of a challenge as this requires some assumptions about the underlying data generating process.

6.2 Structural/Scenario Analysis and Communication

We have argued that clear and timely communication about its policy objectives, analytical framework, and potential risks is an essential component of a successful monetary policy framework. Having access to a well-specified economy-wide model would clearly be desirable in this context. Of course it would have to be estimated (or calibrated) on data for the economy in question in order to be useful. The large-scale econometric models that were commonly used in the 1960s and 70s were of this type, although critics would question whether they were 'well-specified'. As Sims, op.cit., said, they were built on 'incredible identifying restrictions'.

The popularity of the structural VARs that were developed following Sims' seminal paper, was due to their reliance on more palatable (in the views of those who developed these models) identifying restrictions that allowed the user to give causal interpretations of the estimated relationships.

To our knowledge, the current machine learning algorithms is at its infancy in terms of its penetration into the macro literature to provide the user with a similar structural interpretation between the input and output variables in the data, or indeed between the output variables themselves. These algorithms are often characterized as 'black boxes', although there is an emerging literature on interpretable machine learning algorithms (Molnar et al. (2020)). One recent macro example of this is Tiffin (2019). By using Causal Forest algorithm, Tiffin (2019) shows how one can estimate the average impact of a crisis. This avenue of identification has been gaining momentum in other fields of economics, and we see this as an exciting opportunity that will eventually arrive in macro issues. Moreover, we also see the potential of this line of identification marrying the structural VAR way of identification. So in this sense, we see the work of Pearl (2018) and Pearl and Mackenzie (2018) for example becoming useful for central banking in the future especially combined with the big data sets that the central banks are using more and more actively (Doerr et al. (2021)).[25]

We present a brief history of the field of interpretable machine learning (IML), give an overview of state-of-the-art interpretation methods, and discuss challenges. Research in IML has boomed in recent years. As young as the field is, it has over 200 years of old roots in regression modeling and rule-based machine learning, starting in the 1960s. Recently, many new IML methods have been proposed, many of them model-agnostic, but also interpretation techniques specific to deep learning and tree-based ensembles. IML methods either directly analyse model components, study sensitivity to input perturbations, or analyse local or global surrogate approximations of the ML model. The field is approaching a state of readiness and stability, with many methods not only proposed in research, but also implemented in open-source software. But many important challenges remain for IML, such as dealing with dependent features, causal interpretation, and uncertainty estimation, which need to be resolved for its successful application to scientific problems. A further challenge is a missing rigorous definition of interpretability which is accepted by the community. To address the challenges and to advance the field, we would need to not only recall our roots of interpretable, data-driven modeling in statistics and (rule-based) ML, but also to consider other areas such as sensitivity analysis, causal inference, and the social sciences.

Perhaps it may be possible algorithms would be constrained to produce forecasts that are considered to be consistent with generally accepted economic theory. But the VARs identified through sign restrictions on the impulse response functions may provide inspiration, although the 'generally accepted economic theory' would have to be specifically tailored to the machine learning environment. An example might be 'if the algorithm produces a forecast that variable y_1 will increase, then it must simultaneously forecast that variable y_2 will increase'.

Other constraints that might be introduced would be related to the long-run restrictions in VAR models. The ML algorithm would be instructed to ensure that a certain class

[25]For example, in a recent application on the effects of minimum wage, Cengiz et al. (2021) use machine learning tools to construct treatment groups that captures 75% of all minimum wage workers, which is significant improvement over the previous work which has largely focused on subgroups such as fast-food industry where the policy might have been in effect.

of variables would not have any effect on the long-run forecast of another class of variables. A possible combination of ML and FAVAR models would be to ask the ML algorithm to produce the factors that go into the FAVAR model, doing away with the linearity assumption of conventional principal component methods.

7 Conclusions

The thesis in this paper is that the evolution of VAR methodology has led to its gradual adoption in central banks as a useful analytical tool. VAR methodology started at the lower right hand (heavily data focused) corner of the Pagan frontier (Figure 1), but has gradually moved upwards to incorporate theoretical elements thus allowing central banks to use it to conduct structural/causal, including scenario analysis. Such analysis, we have argued, is necessary for the central bank's communication with the public about policy choices and decisions to be credible.

If our thesis is correct, and for machine learning models to be adopted widely by central banks in their monetary policy tool boxes, their causal structure and interpretation need to become more transparent. We hope that our analysis and modest suggestions in the previous section will lead to a conversation between traditional econometricians and machine learning experts on how to combine the best elements of each approach for more informed policy decisions and better understanding of these decisions by the general public.

References

Abadie, Alberto, and Matias D. Cattaneo. 2018. "Econometric Methods for Program Evaluation." *Annual Review of Economics* 10, no. 1 (August): 465–503. doi:10.1146/annurev-economics. https://ideas.repec.org/a/anr/reveco/v10y2018p465-503.html.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable Growth." *American Economic Review* 109, no. 4 (April): 1263–1289. https://ideas.repec.org/a/aea/aecrev/v109y2019i4p1263-89.html.

Athey, Susan. 2015. "Beyond prediction: Using big data for policy problems." *Science* 355, no. 6324 (May): 483–485. doi:DOI:10.1126/science.aal4321. https://science.sciencemag.org/content/355/6324/483/tab-article-info.

Athey, Susan, and Guido Imbens. 2019. *Machine Learning Methods Economists Should Know About.* Papers 1903.10075. arXiv.org, March. https://ideas.repec.org/p/arx/papers/1903.10075.html.

Babii, Andrii, Eric Ghysels, and Jonas Striaukas. 2020. *Machine Learning Time Series Regressions with an Application to Nowcasting.* arXiv: 2005.14057 [econ.EM].

Bernanke, Ben S. 1986. "Alternative explanations of the money-income correlation." *Carnegie-Rochester Conference Series on Public Policy* 25, no. 1 (January): 49–99. https://ideas.repec.org/a/eee/crcspp/v25y1986ip49-99.html.

Bernanke, Ben S, Jean Boivin, and Piotr Eliasz. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120 (1): 387–422.

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen. 2008. *Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence.* Technical report 4. December. https://ideas.repec.org/a/aea/jeclit/v46y2008i4p910-45.html.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2020. *Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions.* Working Papers 202002R. Federal Reserve Bank of Cleveland, January. doi:10.26509/frbc-wp-202002. https://ideas.repec.org/p/fip/fedcwq/87375.html.

Carriero, Andrea, Todd E. Clark, and Marcellino Massimiliano. 2020. *Nowcasting Tail Risks to Economic Activity with Many Indicators.* Working Papers 202013R2. Federal Reserve Bank of Cleveland, May. doi:10.26509/frbc-wp-202013. https://ideas.repec.org/p/fip/fedcwq/87955.html.

Cengiz, Doruk, Arindrajit Dube, Attila S Lindner, and David Zentler-Munro. 2021. *Seeing Beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes.* Working Paper, Working Paper Series 28399. National Bureau of Economic Research, January. doi:10.3386/w28399. http://www.nber.org/papers/w28399.

Chavleishvili, Sulkhan, and Simone Manganelli. 2019. *Forecasting and stress testing with quantile vector autoregression.* Working Paper Series 2330. European Central Bank, November. https://ideas.repec.org/p/ecb/ecbwps/20192330.html.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary policy shocks: What have we learned and to what end?" Chap. 2 in *Handbook of Macroeconomics,* edited by J. B. Taylor and M. Woodford, 1:65–148. Handbook of Macroeconomics. Elsevier. https://ideas.repec.org/h/eee/macchp/1-02.html.

Dees, Stephane, Filippo Di Mauro, M. Hashem Pesaran, and L. Vanessa Smith. 2007. "Exploring the International Linkages of the Euro Area: A Global Var Analysis." *Journal of Applied Econometrics* 22 (1): 1–38. ISSN: 08837252, 10991255. http://www.jstor.org/stable/25146503.

Del Negro, Marco, and Frank Schorfheide. 2006. "How good is what you've got? DSGE-VAR as a toolkit for evaluating DSGE models." *Economic Review* 91 (Q 2): 21–37. https://ideas.repec.org/a/fip/fedaer/y2006iq2p21-37nv.91no.2.html.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims. 1986. *Forecasting and conditional projection using realistic prior distribution.* Staff Report 93. Federal Reserve Bank of Minneapolis. https://ideas.repec.org/p/fip/fedmsr/93.html.

Doerr, Sebastian, Leonardo Gambacorta, and José María Serena Garralda. 2021. *Big data and machine learning in central banking.* BIS Working Papers 930. Bank for International Settlements, March. https://ideas.repec.org/p/bis/biswps/930.html.

Fouliard, Jeremy, Michael Howell, and Hélène Rey. 2020. *Answering the Queen: Machine Learning and Financial Crises.* Working Paper, Working Paper Series 28302. National Bureau of Economic Research, December. doi:10.3386/w28302. http://www.nber.org/papers/w28302.

Fuchs-Schündeln, N., and T.A. Hassan. 2016. "Natural Experiments in Macroeconomics." Chap. 0 in *Handbook of Macroeconomics,* edited by J. B. Taylor and Harald Uhlig, 2:923–1012. Handbook of Macroeconomics. Elsevier. doi:10.1016/bs.hesmac.2016.03. https://ideas.repec.org/h/eee/macchp/v2-923.html.

Gambacorta, Leonardo, Yiping Huang, Han Qiu, and Jingyi Wang. 2019. *How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm.* BIS Working Papers 834. Bank for International Settlements, December. https://ideas.repec.org/p/bis/biswps/834.html.

Gelos, R. G, Lucyna Gornicka, Robin Koepke, Ratna Sahay, and Silvia Sgherri. 2019. *Capital Flows at Risk: Taming the Ebbs and Flows.* IMF Working Papers 19/279. International Monetary Fund, December. https://ideas.repec.org/p/imf/imfwpa/19-279.html.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2015. "Prior Selection for Vector Autoregressions." *The Review of Economics and Statistics* 97, no. 2 (May): 436–451. https://ideas.repec.org/a/tpr/restat/v97y2015i2p436-451.html.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95, no. 1 (March): 161–182. doi:10.1257/0002828053828518. https://www.aeaweb.org/articles?id=10.1257/0002828053828518.

Karagedikli, Ozer, Shaun P. Vahey, and Elizabeth C. Wakerly. 2019. *Improved methods for combining point forecasts for an asymmetrically distributed variable.* CAMA Working Papers 2019-15. Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University, February. https://ideas.repec.org/p/een/camaaa/2019-15.html.

Kelleher, John D. 2019. *Deep Learning.* The MIT Press.

Kilian, Lutz, and Helmut Lutkepohl. 2018. *Structural Vector Autoregressive Analysis.* Cambridge Books. Cambridge University Press.

Kuhn, Max, and Kjell Johnson. 2013. *Applied predictive modeling.* http://www.amazon.com/Applied-Predictive-Modeling-Max-Kuhn/dp/1461468485/.

Leeper, Eric M., and Tao Zha. 2003. "Modest policy interventions." *Journal of Monetary Economics* 50, no. 8 (November): 1673–1700. https://ideas.repec.org/a/eee/moneco/v50y2003i8p1673-1700.html.

Loaiza-Maya, Rubén, and Michael Stanley Smith. 2020. "Real-Time Macroeconomic Forecasting With a Heteroscedastic Inversion Copula." *Journal of Business & Economic Statistics* 38, no. 2 (April): 470–486. doi:10.1080/07350015.2018.151. https://ideas.repec.org/a/taf/jnlbes/v38y2020i2p470-486.html.

Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. 2021. *Machine Learning Advances for Time Series Forecasting.* arXiv: 2012.12802 [econ.EM].

Molnar, Christoph, Giuseppe Casalicchio, and Bernd Bischl. 2020. *Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges.* Technical report. eprint: 2010.09337.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31, no. 2 (Spring): 87–106. https://ideas.repec.org/a/aea/jecper/v31y2017i2p87-106.html.

Mumtaz, Haroon, and Paolo Surico. 2009. "The Transmission of International Shocks: A Factor-Augmented VAR Approach." *Journal of Money, Credit and Banking* 41, no. s1 (February): 71–100. doi:10.1111/j.1538-4616.2008.. https://ideas.repec.org/a/wly/jmoncb/v41y2009is1p71-100.html.

Nakamura, Emi, and Jón Steinsson. 2018. *High Frequency Identification of Monetary Non-Neutrality: The Information Effect.* NBER Working Papers 19260. National Bureau of Economic Research, Inc, January. https://ideas.repec.org/p/nbr/nberwo/19260.html.

Pearl, Judea. 2018. "The Seven Tools of Causal Inference with Reflections on Machine Learning." *Communications of Association for Computing Machinery* 1, no. 1 (November): 1–6. https://ftp.cs.ucla.edu/pub/stat_ser/r481.pdf.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect.* New York: Basic Books. ISBN: 978-0-465-09760-9.

Plagborg-Møller, Mikkel, and Christian K. Wolf. Forthcoming. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica.* https://scholar.princeton.edu/sites/default/files/mikkelpm/files/lp_var.pdf.

Primiceri, G., and A. Tambalotti. 2020. *Macroeconomic Forecasting in the Time of COVID-19.* Working Paper. Federal Reserve Bank of New York.

Ramey, V.A. 2016. "Macroeconomic Shocks and Their Propagation." Chap. 0 in *Handbook of Macroeconomics,* edited by J. B. Taylor and Harald Uhlig, 2:71–162. Handbook of Macroeconomics. Elsevier. doi:10.1016/bs.hesmac.2016.03. https://ideas.repec.org/h/eee/macchp/v2-71.html.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48, no. 1 (January): 1–48. https://ideas.repec.org/a/ecm/emetrp/v48y1980i1p1-48.html.

Smith, Michael S., and Shaun P. Vahey. 2016. "Asymmetric Forecast Densities for U.S. Macroeconomic Variables from a Gaussian Copula Model of Cross-Sectional and Serial Dependence." *Journal of Business & Economic Statistics* 34, no. 3 (July): 416–434. doi:10.1080/07350015.2015.104. https://ideas.repec.org/a/taf/jnlbes/v34y2016i3p416-434.html.

Stock, James, and Mark Watson. 2001. "Vector Autoregressions." *Journal of Economic Perspectives* 15, no. 4 (December): 101–115. doi:10.1257/jep.15.4.101. https://www.aeaweb.org/articles?id=10.1257/jep.15.4.101.

———. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In *Handbook of Macroeconomics,* edited by John Taylor and Harald Uhlig, 2:415–525. Handbook of Macroeconomics. Elsevier. https://ideas.repec.org/h/eee/macchp/v2-415.html.

Tiffin, Andrew J. 2019. *Machine Learning and Causality: The Impact of Financial Crises on Growth.* IMF Working Papers 2019/228. International Monetary Fund, November. https://ideas.repec.org/p/imf/imfwpa/2019-228.html.

Waggoner, Daniel F., and Tao Zha. 1999. "Conditional Forecasts In Dynamic Multivariate Models." *The Review of Economics and Statistics* 81, no. 4 (November): 639–651. https://ideas.repec.org/a/tpr/restat/v81y1999i4p639-651.html.