Essays in Macroeconometrics

Lukas Hoesch

TESI DOCTORAL UPF / Year 2021

THESIS SUPERVISOR

Professor Barbara Rossi

Department of Economics and Business

To my family.

Acknowledgements

Many people have been instrumental to my journey to this Ph.D. thesis and I would like to take the opportunity to thank them here.

First and foremost, I would like to extend my heartfelt gratitude to my advisor Barbara Rossi for her invaluable guidance and advice, her generous support, her constant enthusiasm for my ideas and her encouragement throughout the years of my Ph.D. studies.

I also owe special thanks to Geert Mesters and Tatevik Sekhposyan. Thank you for your great advice, your mentoring and the many valuable comments on my work.

Many other professors have given me their advice and time whom I would like to thank here: Majid Al Sadoon, Christian Brownlees, Kirill Evdokimov, Libertad González, Tassos Magdalinos and Katerina Petrova. I would also like to thank Marta Araque, Laura Agustí, Mariona Novoa and Anna Cano for their outstanding and kind administrative support.

I received many useful comments on the papers included in this thesis from seminar participants at the 27th International Conference on Computing in Economics and Finance, the 2020 European Winter Meetings of the Econometric Society, the 2020 Spanish Economic Association SAEe Meeting, the 40th International Symposium on Forecasting, the 7th Barcelona GSE Jamboree, the 8th SIdE Workshop for Students in Econometrics and Empirical Economics and the 6th Barcelona GSE Summer Forum.

In addition, I would like to thank my colleagues and friends Bjarni Einarsson, Christian Höynck, Adam Lee, André Souza, Philipp Tiozzo and Yiru Wang for the wonderful time throughout our shared years in Barcelona, for countless coffee breaks and for many fruitful discussions over lunch.

I would like to thank Tina and Detlef for supporting me throughout my journey to this thesis. Your encouragement and advice, especially throughout the academic job market, have been invaluable.

I would like to thank my wonderful family, my parents Tine and Paul and my brother Jannis, for their love and support throughout my entire life. Thank you for encouraging me to pursue my dreams and for always being just a phone call away.

Finally, I would like to thank my partner Katharina for embarking on this journey together and for always supporting me and my research. Thank you for making me smile every day and for always being there for me. Your love, patience and unwavering support have made all this possible.

Abstract

This thesis consists of three chapters on topics in Macroeconometrics. Chapter 1 develops a hypothesis test to evaluate economic models and their forecasts robust to instabilities. The test is particularly powerful in the presence of multiple breaks and can be applied to in-sample and out-of-sample moment conditions. An application to predictability of the U.S. equity premium provides evidence in favour of "predictability pockets". Chapter 2 investigates the evolution of the Federal Reserve information advantage and the information channel of U.S. monetary policy. It provides evidence that the information channel is historically relevant, but finds substantially weaker evidence of its presence in recent years, once instabilities are accounted for. Chapter 3 develops a semi-parametric approach to conduct inference in non-Gaussian SVAR models robust to "weak" non-Gaussianity. The method exploits non-Gaussianity when it is present, while yielding correct coverage regardless of the distribution of the structural errors. An application revisits U.S. labor supply and demand elasticities and highlights the limitations of using non-Gaussianity for identification.

Resum

Aquesta tesi consta de tres capítols sobre temes de macroeconometria. El capítol 1 desenvolupa una prova d'hipòtesi per avaluar els models econòmics i les seves previsions robustes a les inestabilitats. La prova és particularment potent en presència de trencaments múltiples i es pot aplicar a condicions de moment dins i fora de la mostra. Una aplicació a la predictabilitat de la prima de renda variable dels Estats Units proporciona evidències a favor de "bosses de predictibilitat". El capítol 2 investiga l'evolució de l'avantatge informativa de la Reserva Federal i el canal d'informació de la política monetària dels Estats Units. Proporciona evidències que el canal d'informació és històricament rellevant, però troba proves substancialment més febles de la seva presència en els darrers anys, un cop es comptabilitzen les inestabilitats. El capítol 3 desenvolupa un enfocament semiparamètric per conduir la inferència en models SVAR no gaussians robustos a no gaussianitat "feble". El mètode explota la no-gaussianitat quan hi és present mentre proporciona una cobertura correcta independentment de la distribució dels errors estructurals. Una aplicació revisa les elasticitats de la oferta i la demanda de mà d'obra dels Estats Units i posa de manifest les limitacions de l'ús del no gaussianisme per a la identificació.

Preface

This thesis consists of three independent chapters on topics in Macroeconometrics.

In the first chapter, titled "Specification Tests Robust to Multiple Instabilities", I develop a hypothesis test for model evaluation which is robust to time-variation in parameters. The proposed method is general and can be applied to any economic model which is characterized by moment conditions. The test can be conducted in-sample to select between two nested specifications of an economic model in the presence of parameter instabilities or out-of-sample to evaluate the performance of model or judgmental forecasts robust to time-variation. The key feature of the proposed test is that it is particularly powerful in the presence of multiple shifts in parameters without imposing a specific form of time-variation. Further, the test statistic provides narrative evidence on which parts of the sample drive the rejection of the null hypothesis. Monte-Carlo simulations of the finite-sample performance of the proposed test show that the test is accurately sized and has high power even when model parameters only undergo one shift or are constant. Hence, researchers can use the test even when there is uncertainty about whether and how parameters change over time. In the empirical part of the paper, I use the test to document the presence of short-horizon predictability in the U.S. equity premium during the postwar period. I find evidence of predictability for a large set of variables once time-variation is taken into account. Further, the test provides evidence of heterogeneity in the location of predictability episodes across variables. The findings explain why traditional tests often fail to uncover predictability in the full sample and why studies that split the sample at different dates often arrive at conflicting results regarding the predictive ability of a wide class of financial variables.

The second chapter, titled "Has the Information Channel of Monetary Policy Disappeared? Revisiting the Empirical Evidence", which is joint work with Barbara Rossi and Tatevik Sekhposyan, explores the empirical importance of the information channel of U.S. monetary policy. The paper studies two questions related to the information channel, making an important departure from the literature by explicitly taking instabilities into account. First, we investigate a sufficient condition for the information channel and evaluate whether the Federal Reserve has an information advantage relative to market participants when

forecasting the current and future state of the economy. We show that instabilities are an important empirical feature of the data and find that the Federal Reserve lost its short-horizon information advantage regarding key macroeconomic variables in recent years. Second, we study the evolution of the information channel. Specifically, we allow the nature of monetary policy shocks to vary over time, depending on whether the information advantage is present in the data. Similar to the information advantage, we find that the information channel appears to be historically important, but that there is substantially weaker empirical evidence of its presence in recent years, once instabilities are accounted for. Our analyses show that in the most recent sample period, (i) market surprises are no longer predictable by the Federal Reserve's forecasts, (ii) private forecasters' responses are less sensitive to monetary policy shocks, and (iii) impulse responses to monetary policy shocks are no longer confounded while an information-robust instrument is required to recover responses with the expected signs in the previous sample period. Our results are consistent with the hypothesis that the decline in the relevance of the information channel is linked to improvements in the communication strategy of the Federal Reserve in recent years.

The third chapter, titled "Robust Inference in Structural VAR Models Identified by Non-Gaussianity", which is joint work with Adam Lee and Geert Mesters, develops robust inference methods for structural vector autoregressive (SVAR) models that are identified using non-Gaussian error distributions. A growing literature exploits non-Gaussianity identification to conduct inference in SVAR models. These papers build on a mathematical result that in a stationary SVAR model with independent shocks, if at least all but one components of the error term have a non-Gaussian density, all model parameters are identified (up to signs and a potential re-ordering of the structural shocks). However, as we show in this paper, existing inference methods in non-Gaussian SVARs are not robust to situations in which the densities of the structural errors that generated the data are "close" to a Gaussian distribution. In such cases, local identification deteriorates and coverage distortions occur. We provide a solution to this problem by treating the SVAR model as a semi-parametric model and using a semi-parametric equivalent of the Neyman-Rao score statistic in order to conduct inference on the possibly weakly identified (or not identified) parameters of the SVAR. We conduct a large simulation study to assess the finite-sample performance of our method and find that the empirical rejection frequencies of the proposed test are close to the nominal size, regardless of the true distribution of the structural errors. Further, we find that the power of the semi-parametric score test comes close to that of a parametric test which uses the true structural error densities. Finally, we employ the proposed approach in an empirical study that revisits identification of supply and demand elasticities in the U.S. labor market. We construct identification-robust confidence regions for the labor supply and labor demand elasticities and construct robust confidence bands for the impulse responses. We find that a non-Gaussianity identification strategy is not sufficient to identify the dynamic response of the economy.

Contents

List of Figures

List of Tables

Chapter 1

Specification Tests Robust to Multiple Instabilities

1.1 Introduction

Instabilities in models of economic and financial time series are widespread. For example, when forecasting U.S. stock returns, financial ratios might contain useful information during some periods, but have no predictive ability during other periods (Farmer et al., 2019; Chinco et al., 2019). Similarly, when estimating a structural model of real economic activity, the absence of financial frictions in economic models might be unproblematic if the estimation sample covers "normal" periods, but might be a crucial omission during financial crises (Christiano et al., 2018). Instabilities are often implicitly acknowledged by conducting sensitivity checks over different subsamples and are sometimes explicitly addressed by testing for structural breaks in model parameters. However, they are commonly ignored when evaluating the specification of a model by means of hypothesis tests. Recently, a growing literature has raised concerns about this practice, arguing that in unstable environments, traditional specification tests have low power and may give conflicting results depending on the subsample considered (Rossi, 2013, 2020). This issue is particularly relevant when estimation samples span long time periods which cover different policy regimes, making it likely that model parameters undergo more than one shift. In such an environment, researchers face an econometric problem: "How can we take multiple instabilities into account when evaluating economic models or their forecasts?"

In this paper, I provide a general approach to test whether a parameter should be included in a model robust to instabilities. The proposed hypothesis test can be applied in-sample to select between two nested specifications of an economic model in the presence of parameter instabilities[1] or out-of-sample to evaluate the forecasting performance of model

[1]Such tests are widely used in the macroeconomic and financial literature and many studies document evidence

or judgmental forecasts robust to time-variation.[2] The main advantage of the proposed test is that it is particularly powerful in the presence of multiple shifts in parameters without imposing a specific form of time variation. At the same time, the test is accurately sized in finite-samples and has high power even when model parameters only undergo one shift or are constant. This makes the test particularly useful when the researcher faces uncertainty about whether and how parameters change over time. The test is simple to compute, can be efficiently implemented by a dynamic programming algorithm provided in the paper and the test statistic path can be plotted to provide narrative evidence on which parts of the sample drive the rejection of the null hypothesis.

The proposed test is a joint hypothesis test for *both* parameter instability *and* a constant non-zero value of the parameter. In particular, the null hypothesis of the test specifies that the parameter, which can potentially be time-varying, has a zero value *at every point in time* throughout the sample. The test rejects against alternatives in which the parameter has a non-zero value *at some point in time* over the sample. Therefore, the test detects departures from the null hypothesis even when they only occur over short periods of the sample. This makes the test more powerful than traditional hypothesis tests (such as t-tests, Wald or LM tests) which are based on the full sample and fail to reject the null hypothesis if instabilities "average out" over the sample. The joint null hypothesis also distinguishes the test from tests of multiple structural breaks which are designed to detect parameter instability *only* and do not reject against constant alternatives.

The novel test statistic is intuitive and flexible. The test statistic jointly considers all possible splits of the sample at K splitting points into a sequence of consecutive blocks of variable lengths. For each block, a statistic is computed which evaluates whether the data inside each block supports a rejection of the null hypothesis. The test rejects if the combined information from all possible splits supports the alternative hypothesis. This allows the test to achieve high power in the presence of multiple shifts. The test can be constructed based on a set of moment conditions involving the parameters of interest using both a Lagrange-Multiplier (LM) form and a Wald form. The LM form *imposes* the null hypothesis that the parameter is zero at every point in time. In contrast, the Wald form *estimates* the entire parameter vector for each considered block by a partial-sample Generalized Method of Moments (GMM) estimator. The test statistic can be computed for a fixed number of splits or by specifying an upper bound of splits to take into account. Regardless of the number of splits taken into account, the test statistic can be implemented efficiently by a dynamic programming algorithm provided in the paper.

of instabilities. For example, Rossi (2006b) evaluates whether exchange rates are random walks and finds instabilities in the parameters of interest. Similarly, Rossi (2013) finds instabilities when evaluating predictive models of inflation. Welch and Goyal (2007) and Timmermann (2008) evaluate a wide set of predictive models for US stock returns and document that predictive ability is time-varying.

[2]There is a large literature evaluating out-of-sample forecast performance by means of specification tests; Clark and McCracken (2013) provide an overview. Rossi and Sekhposyan (2016) and Rossi (2020) document that out-of-sample specification tests are affected by instabilities and discuss how to take instabilities into account when evaluating forecasts.

Finally, the test is widely applicable. In particular, the test can be applied to any economic model which is described by a set of moment conditions. Moment conditions can be derived from many reduced form models and structural models such as linear regressions, vector autoregressions, structural equations identified using instrumental variables or even dynamic stochastic general equilibrium models. Alternatively, the test can be used to evaluate out-of-sample forecasting performance by applying it to a moment condition describing a sequence of out-of-sample forecast errors. These forecast errors can be obtained either from a forecasting model whose parameters are estimated using a recursive scheme or from model-free forecasts such as survey or judgmental forecasts. Applications include testing in the presence of instabilities for forecast unbiasedness, rationality, efficiency or forecast encompassing.

This paper makes three contributions to the literature.

First, I provide an instability-robust hypothesis test for a general class of models, explicitly taking multiple discrete shifts in parameters into account. In contrast to structural break tests which test parameter stability only, the procedure *jointly* tests parameter stability and a linear hypothesis on the parameter vector specified by the researcher. Contrary to tests which assume a single break in parameters and only indicate the location of the largest shift, the path of the proposed test statistic can be plotted to provide narrative evidence on which periods of the sample are driving the rejection of the null hypothesis. I derive the limiting distribution of the test statistic which is a function of independent Brownian Motions and tabulate its critical values.

Second, I investigate the finite sample performance of the proposed test across a series of data-generating processes and compare its performance to that of other tests from the literature. The simulations illustrate that asymptotically, traditional hypothesis tests using the full sample and tests of structural change, such as the $UD\max$ test of Bai and Perron (1998), have no power against some of the relevant alternatives. In contrast, the proposed test exhibits significant and monotonic power for these alternatives. The simulations further show that the proposed test is accurately sized across a variety of sample sizes and various forms of serial correlation. Finally, I compare the finite sample power of the proposed procedure to that of traditional specification tests based on the full sample and the QLR_T^* specification test imposing one break by Rossi (2005). I find that the proposed test yields substantially larger finite-sample power if the data-generating process exhibits multiple shifts in parameters. Further, if the data-generating process exhibits one shift or constant coefficients, the power loss compared to existing tests is small. Thus, researchers can use the proposed test without prior knowledge of whether and how parameters vary over time.

Third, I use the proposed test to document the presence of local short-horizon predictability in the U.S. equity premium during the 1946-2019 period using a set of financial variables considered by Welch and Goyal (2007). Recently, various studies have provided theoretical and empirical evidence that predictability is concentrated in short-lived periods, so-called

"pockets of predictability" (Timmermann, 2008).[3] This form of predictability is particularly difficult to detect using traditional specification tests (Rossi, 2020) and previous efforts are based on repeated tests in overlapping samples of the data, leading to issues associated with multiple testing (Farmer et al., 2019). In contrast, the test proposed in this paper explicitly takes the search across multiple subsamples into account, thereby avoiding the multiple testing problem. Hence, the test can be used to detect predictability even in the presence of "predictability pockets". I find that one-month-ahead excess market returns are predictable from a larger set of variables than typically found in the literature once multiple shifts in predictability are taken into account. In contrast to traditional predictability tests, the conclusions from the proposed test are invariant to starting the sample after the 1951 Treasury Accord Act. Furthermore, the paths of the test statistics provide evidence of heterogeneity in the location of predictability episodes across predictors. The findings explain why traditional tests often fail to uncover predictability in the full sample and why studies that split the sample at different dates often arrive at conflicting results regarding the predictive ability of a wide class of variables.

LITERATURE. Several papers have proposed specification tests robust to instabilities. However, these focus either on the case of a single break or on a different class of alternatives than the one considered in this paper.

A related method to the one proposed in this paper which also builds on moment conditions and tests a joint hypothesis is developed in Rossi (2005) which considers optimal tests for the case of a single break in parameters. In contrast, the test statistic considered in this paper allows the researcher to consider an unknown number of shifts in parameters up to a specifiable upper bound and nests the case of a single break. This makes the test more powerful in the presence of multiple shifts in parameters while retaining comparable power if the parameter shifts only once. In addition, the path of the proposed test statistic provides narrative evidence on which periods of the sample are driving the rejection of the hypothesis whereas a test imposing one break indicates the location of the largest shift in the parameter vector.

A different strand of the literature designs hypothesis tests which are robust against particular alternatives. These include tests for predictability in threshold models (Gonzalo and Pitarakis, 2012, 2017), tests of relative forecasting performance under Markov-switching alternatives (Odendahl et al., 2020), automated model-selection in the presence of instabilities (Castle et al., 2012) and real-time detection of predictability regimes (Harvey et al., 2020). The advantage of the test which is proposed in this paper is that it remains agnostic about the specific process driving the changes in parameters and thus offers a general approach that can be used if the researcher has no prior knowledge on whether

[3] Timmermann (2008) notes that "[...] there appear to be pockets in time where there is modest evidence of local predictability; [...] the identity of the best forecasting method can be expected to vary over time, and there are likely to be periods of model breakdown where no approach seems to work".

and how parameters vary over time.

Finally, there is a large literature on testing for multiple structural changes. A seminal contribution is Bai and Perron (1998) who proposed $\sup F$ tests in a class of linear regression models. Structural break tests of multiple changes have since been extended to more general classes of models (Sowell, 1996; Perron and Qu, 2006; Qu and Perron, 2007; Elliott and Müller, 2006). A related literature provides structural change tests for predictive regression models; Pitarakis (2017) and Georgiev et al. (2018) are two examples of recent contributions. In contrast to the approach presented in this paper which tests a joint hypothesis, structural change tests focus on the null hypothesis of parameter stability only and therefore do not have power against some of the alternatives considered in this paper.

OUTLINE. Section 1.2 discusses the hypotheses of interest and proposes the test statistic. Sections 1.3 and 1.4 discuss the relevant asymptotic theory to conduct in-sample and out-of-sample inference, respectively. Section 1.5 provides a guide for implementing the test. Section 1.6 explores the finite sample performance of the proposed test by means of extensive Monte Carlo simulations. Section 1.7 applies the test to study the predictability of the U.S. equity premium robust to instabilities. Section 1.8 concludes.

1.2 Specification tests robust to multiple instabilities

This section formalizes the testing problem of parameter inclusion under instabilities and introduces the test statistic.

1.2.1 Model and hypotheses

Consider a model indexed by a v-dimensional parameter vector θ_t for $t = 1, \dots, T$. Assume the parameter vector partitions $\theta_t = (\beta_t', \delta)$ where β_t is $(p \times 1)$ and δ is $(q \times 1)$. Further, assume the model satisfies the following m-dimensional moment condition

$$\mathbb{E}\left[f(z_t, \beta_t, \delta) \right] = 0 \tag{1.1}$$

where z_t is an r-dimensional random vector of data and $f : \mathbb{R}^r \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^m$.

Based on the moment condition in (1.1), the researcher wants to test whether the sequence of *possibly* time-varying parameters β_t should be included when modeling the data. The relevant hypotheses are

$$H_0 : \ \beta_t = 0 \quad \forall\, t \qquad \text{vs.} \qquad H_A : \ \beta_t \neq 0 \quad \text{for some} \ \ t \geq 1 \tag{1.2}$$

Under the null hypothesis β_t is zero *at any point in the sample* and thus can be excluded from the model. Under the alternative, β_t is different from zero *at some point in the sample* and should be included in the model.[4] The test of parameter inclusion is thus required to be sensitive to situations in which β_t departs from zero only during short periods over the sample.

It will be useful to express the null hypothesis as an intersection. Let β_0 denote the true parameter value under the null hypothesis that β_t is constant in which case $\theta_0 = (\beta_0, \delta_0)$. Then, we can express the null hypothesis in (1.2) as

$$\begin{aligned} H_0 : \beta_t \in B^1 \cap B^2 \qquad \text{with} \quad & B^1 \equiv \left\{\beta_t \in \mathbb{R}^p : \beta_t = \beta_0 \; \forall \; t\right\} \\ & B^2 \equiv \left\{\beta_0 \in \mathbb{R}^p : \beta_0 = 0\right\} \end{aligned} \tag{1.3}$$

Denote the first part of this null hypothesis imposing a constant parameter vector by $H_0^{(1)} : \beta_t \in B^1$ and the second part imposing a zero value by $H_0^{(2)} : \beta_t \in B^2$.

It is important to note that the hypotheses in (1.2) are different from those of a test assuming constant parameters. In particular, if the researcher misspecifies the moment condition in (1.1) by assuming that θ is constant and tests the hypothesis $H_0 : \beta = 0$ against the alternative $H_A : \beta \neq 0$, the test will reject if the constant parameter β is different from zero rather than β_t being zero. Depending on the form of time-variation in β_t, it can be the case that $\beta = 0$ but that $\beta_t \neq 0$ for some t. In that case instabilities "average out" over the sample and a test of the hypotheses in (1.2) will reject while a test assuming constant parameters will not.[5]

One might be tempted to think that a test of the hypotheses in (1.2) is simply a test of parameter stability. However, as (1.3) shows, the relevant null hypothesis is a *joint hypothesis* which simultaneously imposes (i) a constant coefficient vector and (ii) a zero value on the constant coefficient vector. In contrast, the null hypothesis of a test for parameter stability only focuses on the first part, that is $H_0^{(1)}$. This implies that there are data-generating processes for which a test of parameter stability does not reject while the test considered in this paper does. In particular, this would be the case when the true parameter vector is constant, but different from zero.

The test statistic proposed in this paper detects violations of the null hypothesis in equation (1.2) by combining information obtained from partitioning the sample into a series of discrete consecutive blocks. In particular, the test searches over all possible combinations of $K+1$ discrete segments and evaluates whether there is evidence to reject the null hypothesis. While the test does not impose an alternative of a particular form, it has particularly high power when the parameter undergoes a series of discrete changes. Specifically, assume that β_t has a "baseline value" of β_A but that it undergoes a series of K discrete changes where

[4] Note that by redefining β_t one can similarly test any linear hypothesis on the parameter.

[5] For a detailed explanation of this argument see the discussion in Rossi (2005) and Rossi (2013).

the value of β_t departs from β_A. Collect these change-points expressed as a fraction of the sample in a K-dimensional vector $\lambda^K := (\lambda_1, \ldots, \lambda_K)$ where $\lambda_j \in (0,1)$ and $\lambda_j > \lambda_{j-1}$. Further, collect the magnitudes of these changes in a K-dimensional vector β_Δ. Under this assumption, the alternative hypothesis in (1.2) can be expressed as

$$H_A : \beta_t = \beta_A + \sum_{j=1}^{K} \mathbb{1}\left([\lambda_j T] < t \leq [\lambda_{j+1} T]\right) \cdot \beta_{\Delta,j} \qquad t = 1, \ldots, T \tag{1.4}$$

where $\beta_{\Delta,j}$ denotes the j-th element of β_Δ, $\lambda_0 \equiv 0$, $\lambda_{K+1} \equiv 1$ and $[\cdot]$ is the integer part operator.[6] The proposed test statistic searches across all possible values of λ^K to best approximate the path of β_t and evaluates for each candidate λ^K whether the null hypothesis $\beta_t = 0 \; \forall t$ can be rejected. If the path of β_t takes the form described in (1.4), there is a value of K and λ^K for which the approximation will be exact. This makes the test particularly powerful in the presence of discrete changes.

Tests of the hypothesis considered above have many applications in empirical work. At the end of this section, I provide various examples of problems which have been studied in the macroeconomic and financial literature and illustrate how they fit into the testing framework.

1.2.2 Test statistics

This section introduces a class of test statistics which can be used to test the null hypothesis defined in (1.2). As previewed before, the main idea of the tests is simple: To detect departures from the null hypothesis, the test statistic jointly considers all possible splits of the sample at K splitting points into a sequence of $K+1$ consecutive blocks of variable length. For each block, a statistic is computed which evaluates whether the data inside each block supports a rejection of the null hypothesis. If there is a sample split for which the sum of statistics computed on each block supports the alternative, the test rejects. In what follows, I first describe how to construct the test statistic for a fixed number of sample splits K. Consecutively, I discuss how to robustify the test against the choice of K.

Test with a fixed number of splits K

Assume the researcher wants to test parameter inclusion robust to time-variation according to the hypotheses in (1.2) on a sample $t = T_0, \ldots, T$ where T_0 denotes the first observation and is typically set to 1. Let λ^K denote a K-dimensional vector of splitting points $\lambda^K := (\lambda_1, \ldots, \lambda_K)$ where $\lambda_j \in (0,1)$ and $\lambda_j > \lambda_{j-1}$. Each value of λ^K implies a different partition of the sample into a sequence of $K+1$ consecutive blocks where block j spans data from $t = [\lambda_{j-1} T] + 1, \ldots, [\lambda_j T]$ with $T_0 \equiv [\lambda_0 T] + 1$ and $\lambda_{K+1} \equiv 1$.

[6] For example, when $\beta_t = 0$ for $t = 1, \ldots, [T/2]$ and $\beta_t = \beta_\Delta$ for $t = [T/2]+1, \ldots, T$, then $\beta_A = 0$, $\lambda_1 = 1/2$, $K = 1$ and β_Δ is a scalar. In contrast, when β_t is constant at β_A for $t = 1, \ldots, T$, $K = 0$.

For simplicity of exposition, I first abstract from the choice of the number of splits and assume that K is a known value. The proposed test statistic for testing the null hypothesis in (1.2) using K sample splits takes the following form.

$$\begin{aligned} \sup \Phi_T(K) := \sup_{\lambda^K \in \Lambda_\epsilon} & \sum_{j=1}^{K+1} \Phi_{T,j}(\lambda_{j-1}, \lambda_j) \\ \Lambda_\epsilon \equiv \Big\{ \lambda_j : \;\; & \lambda_j \in (\lambda_0 + \epsilon, \lambda_{K+1} - \epsilon),\; \lambda_j > \lambda_{j-1} + \epsilon,\;\; j = 1, \ldots, K \Big\} \end{aligned} \tag{1.5}$$

For a given sample split λ^K, the test statistic is simply the sum of $K+1$ statistics $\Phi_{T,j}(\lambda_{j-1}, \lambda_j)$ computed on each block of the data. The $\sup_{\lambda^K}$ part of the test statistic searches over all possible combinations of K splitting points for the choice of λ^K which maximizes this sum. The value at the optimal choice of λ^K which yields the maximum value for the sum term is the final value of the test statistic.

Note that the search of λ^K is restricted to a set Λ_ϵ defined by a trimming parameter $\epsilon \in (0, 1)$ which imposes that each of the blocks contains at least $[\epsilon T]$ observations. This parameter is set by the researcher prior to conducting the test and its choice depends on the stochastic properties of the data.[7] The simulations presented in Section 1.6 provide guidance on the trade-offs of choosing a lower or higher value of ϵ. In most applications, a choice of $\epsilon = 0.05$ or $\epsilon = 0.1$ is sufficient.

Choice for $\Phi_{T,j}(\cdot, \cdot)$

The test statistic above crucially depends on the choice for $\Phi_{T,j}(\cdot, \cdot)$. In this paper, I consider two forms, a Lagrange-Multiplier (LM) statistic and a Wald statistic. The LM form imposes the value of β_t to be zero in each block while the Wald form estimates β_t in each block. In both cases, the test statistic builds on partial sums of the moment condition defined in equation (1.1), evaluated at estimates of δ.

The tests proposed in this paper can be conducted based on moment conditions formulated in-sample or out-of-sample. These two cases differ in the portion of the sample on which the test statistic is constructed as well as the estimation scheme which is used to estimate δ. In the in-sample case, the test is conducted on the full sample setting $T_0 = 1$ and δ is estimated based on the moment condition in (1.1). The resulting estimate is denoted $\hat{\delta}$. In contrast, in the out-of-sample case the test is conducted on an out-of-sample portion of the data setting $T_0 = R$ where $R \gg 1$. Here, δ is the parameter of a forecasting model identified by a separate moment condition which is estimated using a recursive scheme on an in-sample portion of the data, yielding a sequence of estimators of δ denoted $\{\hat{\delta}_t\}_{t=R}^T$. The in-sample case is discussed in more detail in Section 1.3 while the out-of-sample case is discussed in Section 1.4 of the paper.

[7] The use of trimming parameters is standard in tests for structural breaks, see e.g. Andrews (1993) or Bai and Perron (1998).

LAGRANGE-MULTIPLIER FORM. If the test is implemented using the Lagrange-Multiplier form, the test statistic is constructed using the following choice for $\Phi_{T,j}(\cdot,\cdot)$.

$$\begin{aligned}\Phi^{LM}_{T,j}(\lambda_{j-1},\lambda_j) &:= \hat{\mathcal{F}}'_{T,j} \times \hat{\Omega}_{T,j} \times \hat{\mathcal{F}}_{T,j} \\ \hat{\mathcal{F}}_{T,j} &:= (T-T_0+1)^{-1/2}\ \hat{\Sigma}_{ff}^{-1/2} \sum_{t=[\lambda_{j-1}T]+1}^{[\lambda_j T]} f(z_t,\tilde{\theta}_t)\end{aligned} \tag{1.6}$$

Note that $f(\cdot,\cdot)$ is the moment function which was defined in equation (1.1), $\hat{\Sigma}_{ff}$ is a consistent estimator of the long-run variance of the sample moments under the null hypothesis and $\hat{\Omega}_{T,j}$ is a consistent estimator of the long-run variance of $\hat{\mathcal{F}}_{T,j}$. Formulas to compute these estimators are given in Section 1.5 of this paper. $\tilde{\theta}_t$ is a restricted generalized method of moments (GMM) estimator of θ_t that imposes the joint null hypothesis defined in (1.2) which restricts $\beta_t = \beta_0 = 0\ \forall\ t$ while leaving δ unspecified.

The difference between testing in-sample and out-of-sample using the statistic above lies in how the estimate of δ, and consequently $\tilde{\theta}_t$ is obtained. In the out-of-sample case, the restricted estimator is formed as $\tilde{\theta}_t := (0_{p\times 1}, \hat{\delta}_t)$ where $\{\hat{\delta}_t\}_{t=T_0}^{T}$ is a sequence of estimates of δ which is obtained via a recursive estimation scheme. Section 1.4 discusses in detail how to obtain these estimates. In contrast, in the in-sample case $\tilde{\theta}_t = \tilde{\theta}\ \forall t$ where $\tilde{\theta}$ is a constant GMM estimator which is defined as follows.

$$\begin{aligned}\tilde{\theta} &:= \arg\max_{\theta\,\in\,\Theta}\ \hat{Q}_T(\theta) \qquad\qquad \hat{Q}_T(\theta) \equiv \hat{F}_T(\theta)'\ W_T\ \hat{F}_T(\theta) \\ \hat{F}_T(\theta) &\equiv (T-T_0+1)^{-1} \sum_{t=T_0}^{T} f(z_t,\theta) \\ &\text{subject to}\quad A\,\theta = 0_{p\times 1} \qquad\qquad A = \begin{bmatrix} I_{p\times p} & 0_{p\times q}\end{bmatrix}\end{aligned} \tag{1.7}$$

$\hat{F}_T(\theta)$ is the sample analogue of the moment condition defined in (1.1) and W_T is a positive definite weighting matrix.

WALD FORM. If the test is implemented using the Wald form, the test statistic is constructed using the following choice for $\Phi_{T,j}(\cdot,\cdot)$.

$$\Phi^{W}_{T,j}(\lambda_{j-1},\lambda_j) := (T-T_0+1)\ \left[\ \hat{\beta}_j(\lambda_{j-1},\lambda_j)\ \right]'\ \times\ \hat{\Omega}^{-1}_{T,j}\ \times\ \left[\ \hat{\beta}_j(\lambda_{j-1},\lambda_j)\ \right] \tag{1.8}$$

The difference between conducting the test in-sample and out-of-sample lies in how the estimate of $\hat{\beta}_j(\cdot,\cdot)$ is obtained. In the in-sample case, $\hat{\beta}_j(\lambda_{j-1},\lambda_j) := A\ \hat{\theta}_j(\lambda_{j-1},\lambda_j)$ where $A \equiv \begin{bmatrix} I_{p\times p} & 0_{p\times q}\end{bmatrix}$ and $\hat{\theta}_j(\lambda_{j-1},\lambda_j)$ is defined as the following GMM estimator

which assumes the parameter θ has a constant value in block j.

$$\begin{aligned} \hat{\theta}_j &:= \arg\max_{\theta \in \Theta} \hat{Q}_{T,j}(\theta) \qquad \hat{Q}_{T,j}(\theta) \equiv \hat{F}_{T,j}(\theta)' \, W_T \, \hat{F}_{T,j}(\theta) \\ \hat{F}_{T,j}(\theta) &\equiv \big([\lambda_j T] - [\lambda_{j-1} T]\big)^{-1} \sum_{t=[\lambda_{j-1}T]+1}^{[\lambda_j T]} f(z_t, \theta) \end{aligned} \tag{1.9}$$

where $\hat{F}$ is a partial-sample analogue of the moment condition defined in (1.1) and W_T is a positive definite weighting matrix.

In contrast, in the out-of-sample case $\hat{\beta}_j$ is obtained using a similar GMM estimator.

$$\begin{aligned} \hat{\beta}_j &:= \arg\max_{\beta \in B} \hat{Q}_{T,j}(\beta) \qquad \hat{Q}_{T,j}(\beta) \equiv \hat{F}_{T,j}(\beta)' \, W_T \, \hat{F}_{T,j}(\beta) \\ \hat{F}_{T,j}(\beta) &\equiv \big([\lambda_j T] - [\lambda_{j-1} T]\big)^{-1} \sum_{t=[\lambda_{j-1}T]+1}^{[\lambda_j T]} f(z_t, \beta, \hat{\delta}_t) \end{aligned} \tag{1.10}$$

In contrast to the in-sample case, the estimator does not define an estimate of δ, but evaluates the partial-sample moment at the given sequence of parameter estimates, $\{\hat{\delta}_t\}_{t=T_0}^{T}$.

In a given application, both the LM and the Wald form of $\Phi_{T,j}(\cdot,\cdot)$ can be used to construct the $\sup \Phi_T(K)$ test statistic defined above. However, their performance may differ in finite samples depending on the application considered. Generally, the LM form of the test statistic is computationally efficient as it only requires computing an estimate of θ under the null hypothesis. In contrast, the Wald form of the test statistic requires re-estimating θ for each block and every sample partition considered.[8] Because of its general computational benefits, the rest of the paper will mainly focus on the LM form of the test statistic.

To implement the LM and Wald statistics, one requires the variance estimators $\hat{\Sigma}_{ff}$ and $\hat{\Omega}_{T,j}$. Formulas to compute these estimators are given in Section 1.5. Computation of the $\sup_\lambda^K$ operator in (1.5) can be achieved efficiently by means of a dynamic programming algorithm which is also provided in Section 1.5. The algorithm computes the LM form in $\mathcal{O}(T^2)$ operations computing the sum term in equation (1.5).

Test for an unknown number of splits

The discussion in the previous section has made a simplifying assumption, namely that the researcher wants to conduct the test with a specific number of splits K in mind. However, in practice it is unlikely that the researcher has prior information about the appropriate choice of K. To circumvent this problem, this section presents a robustified version of the test statistic which abstracts from the choice of K by combining information from computing the test statistic for different values for K, starting at one and stopping at some

[8] For linear models, this estimator can be implemented efficiently by a recursion.

pre-defined ceiling, $\bar{K}$.[9] The proposed test statistic for testing the joint null hypothesis in (1.2) considering up to $\bar{K}$ splits of the sample has the following form.

$$D\sup\Phi_T(\bar{K}) := \max_{1\leq k\leq \bar{K}} \left\{ \sup\Phi_T(k)/k \right\} \tag{1.11}$$

where $\sup\Phi_T$ is the test statistic defined in equation (1.5).

To compute this test statistic, one computes the $\sup\Phi_T(k)$ statistic for different choices of $k = 1, \ldots, \bar{K}$. The resulting values are then weighted by the number of shifts used to compute them. The maximal value of this re-weighted series of test statistics is the final value of the test statistic.

In practice, a choice of $\bar{K}$ as low as five or ten is often sufficient in applications. In fact, simulations show that the choice of $\bar{K}$ has little impact on size and power of the test beyond these values.[10] In general, the admissible values of $\bar{K}$ are bounded above by the choice of the trimming parameter ϵ which is used to compute the $\sup\Phi_T$ statistic since it imposes a minimum number of observations for each segment and therefore implicitly defines an upper bound on K. For instance when $\epsilon = .1$, the maximum number of admissible splits which can be considered is $\bar{K} = 10$.

1.2.3 Examples

Tests which are robust to heterogeneity in parameters over time have many applications in empirical work. In the following, I provide two examples of testing problems which have been widely studied in the macroeconomic and financial literature and illustrate how they fit into the testing framework.

PREDICTIVE REGRESSIONS. Consider the following model

$$y_{t+h} = \delta + X_t'\,\beta + \eta_{t+h} \qquad\qquad t = 1, \ldots, T \tag{1.12}$$

where y_{t+h} is a scalar series to be predicted in-sample at horizon h, X_t is a $(p \times 1)$ vector of predictors which are suspected to have a time-varying relationship with y_{t+h}, δ is the constant of the regression and η_{t+h} is a sequence of unforecastable errors. A large literature in macroeconomics and finance studies tests of the hypothesis $H_0 : \beta = 0$ in the model above e.g. to determine the predictive ability of financial variables (Pitarakis and Gonzalo, 2019) or to evaluate model specifications for variables such as inflation (Rossi, 2005). However, in recent years many studies have documented that predictive ability in these models, which is captured by β, is time-varying and that tests based on the full sample may fail to reject in the presence of predictability (Welch and Goyal, 2007; Timmermann, 2008; Rossi, 2013).

[9] A similar method to robustify tests to the choice of K was first proposed in the context of structural break tests in Bai and Perron (1998).

[10] The simulation results are available on request.

To implement the test robust to instabilities in the parameters, replace β in equation (1.12) by β_t and note that the resulting model implies the following moment condition.

$$\mathbb{E}\left[f(z_t, \beta_t, \delta)\right] = 0 \qquad\qquad f(z_t, \beta_t, \delta) \equiv \begin{bmatrix} X_t \cdot (y_{t+h} - \delta - X_t'\beta_t) \\ 1 \cdot (y_{t+h} - \delta - X_t'\beta_t) \end{bmatrix}$$

where $z_t = (y_{t+h}, X_t')'$ and $\theta_t = (\beta_t', \delta)'$.

To conduct the $D\sup\Phi_T(K)$ test based on the LM statistic, compute the restricted GMM estimator $\tilde{\theta}$ via the estimator in (1.7) which imposes the null hypothesis $\beta_t = 0\ \forall t$. Set $T_0 = 1$ and use the variance estimator in (1.15) or (1.17) to compute $\hat{\Sigma}_{ff}$ and use (1.18) to compute $\hat{\Omega}_{T,j}$. Then, compute $D\sup\Phi_T(\bar{K})$ defined in (1.11) by means of the dynamic programming algorithm described in Section 1.5. Reject the null hypothesis of no predictive ability if the computed value of the test statistic is larger than the appropriate critical value reported in Section 1.3.

OUT-OF-SAMPLE PREDICTIVE ABILITY. West and McCracken (1998) proposed a framework to evaluate out-of-sample predictive ability by testing the null hypothesis H_0 : $\beta = 0$ vs. $H_A : \beta \neq 0$ in the following linear model

$$v_{t+h}(\hat{\delta}_t) = \hat{\xi}(z_t, \hat{\delta}_t)'\,\beta + \eta_{t+h}, \qquad t = R, \ldots, T \tag{1.13}$$

where $v_{t+h}(\hat{\delta}_t)$ is a sequence of forecast errors derived from a parametric forecasting model which depend on the sequence of estimated parameters of the forecasting model $\hat{\delta}_t$. By appropriately choosing the function $\hat{\xi}(z_t, \hat{\delta}_t)$, the framework includes popular test for forecast evaluation such as tests of forecast unbiasedness, forecast rationality or forecast encompassing. Many studies provide empirical evidence that out-of-sample predictive ability is time-varying e.g. for forecast rationality of private sector forecasts or for forecast encompassing tests evaluating the information-advantage of Federal Reserve forecasts (Rossi and Sekhposyan, 2016; Hoesch et al., 2020).

To implement the proposed test robust to heterogeneity in β, note that for time-varying β_t, the model above implies the following moment condition.

$$\mathbb{E}\left[f(z_t, \beta_t, \hat{\delta}_t)\right] = 0 \qquad\qquad f(z_t, \beta_t, \hat{\delta}_t) \equiv \hat{\xi}(z_t, \hat{\delta}_t) \cdot \left(v_{t+h}(\hat{\delta}_t) - \hat{\xi}(z_t, \hat{\delta}_t)'\beta_t\right) \tag{1.14}$$

where $z_t = v_{t+h}(\hat{\delta}_t)$.

For the purpose of this example, focus on a test of forecast unbiasedness which sets $\hat{\xi}(z_t, \hat{\delta}_t) = 1$. West and McCracken (1998) showed that in this case parameter estimation error in δ_t is asymptotically irrelevant (for more details see Section 1.4) so that we can use the simple formulas for the variance estimators reported above. To conduct the $D\sup\Phi_T(K)$ test based on the LM statistic, set $\tilde{\theta}_t = (0_{p\times 1}', \hat{\delta}_t)$ and $T_0 = R$ and use

the variance estimator in (1.15) or (1.17) to compute $\hat{\Sigma}_{ff}$ and use (1.18) to compute $\hat{\Omega}_{T,j}$. Then, compute $D\sup\Phi_T(\bar{K})$ defined in (1.11) by means of the dynamic programming algorithm described in Section 1.5. The null hypothesis of forecast unbiasedness can be rejected if the computed value of the test statistic is larger than the appropriate critical value reported in Section 1.4.

1.3 In-sample inference

This section describes the relevant asymptotic theory to conduct in-sample inference. In the in-sample case, the $\sup\Phi_T(K)$ test statistic defined in equation (1.5) and the $D\sup\Phi_T(\bar{K})$ test statistic in equation (1.11) are constructed setting $T_0 = 1$ and evaluating $\Phi_T(\cdot,\cdot)$ at the relevant GMM estimators defined in equations (1.7) and (1.9) for the LM and Wald case, respectively. This section presents and discusses a set of regularity assumptions which are sufficient to obtain weak convergence of the test statistics under the null hypothesis to a function of Brownian Motions. This result is established in the main theorem of this section.

NOTATION. I introduce some notational conventions that are required for this section and used throughout the rest of the paper. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space on which all of the random elements are defined. Unless specified otherwise, all limits are taken as the sample size $T \to \infty$. The symbol $\xrightarrow{p}$ denotes convergence in probability and $\xrightarrow{d}$ denotes convergence in distribution. Next, $\Rightarrow$ denotes weak convergence for sequences of measurable random elements of a space of bounded Euclidean-valued cadlag functions on the product space $D[0,1]^m$ as defined in Phillips and Durlauf (1986) where each component space $D[0,1]$ is equipped with the Skorohod metric. $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix.

The following regularity assumptions are sufficient to obtain weak convergence of the test statistics under the null hypothesis to the limiting distribution characterized in the theorem below.

ASSUMPTION 1.3.1 (Regularity conditions): *Assume the following regularity conditions hold.*

- *(i)* $\{z_t\}$ *is strong mixing with strong mixing coefficients* $\{\alpha(n)\}$, $\sum_{n=1}^{\infty}\alpha(n)^{1-2/\gamma} < \infty$ *with* $\gamma > 2$.
- *(ii)* $\{z_t\}$ *is weakly stationary. In addition* $\mathbb{E}[f(z_t,\theta_0)] = 0$ *for all* $t = 1,\ldots,T$ *and* $T = 1,2,\ldots$ *and the individual elements of* $f(z_t,\theta_0)$ *have the finite absolute moment* $\mathbb{E}\left[|f^{(i)}(z_t,\theta_0)|^{\gamma}\right] < \infty$ *for* $i = 1,\ldots,m$ *and* $\gamma > 2$.
- *(iii)* $\Sigma_{ff} \equiv \lim_{T\to\infty}\mathbb{E}\left[T^{-1}\left\{\sum_{t=1}^{T}f(z_t,\theta_0)\right\}\left\{\sum_{t=1}^{T}f(z_t,\theta_0)\right\}'\right] \in \mathbb{R}^{m\times m}$ *is positive definite.*

(iv) $f(z,\theta)$ *is continuously partially differentiable in* θ *in a neighborhood of* θ_0 *for every* $\theta_0 \in \Theta^*$ *where* Θ^* *is some convex or open set that contains* Θ*. The functions* $f(z,\theta)$ *and* $\nabla_\theta f(z,\theta) \equiv \partial f(z,\theta)/\partial\theta$ *are measurable functions of* z *for each* $\theta \in \Theta$ *and* $\mathbb{E}\left[\sup_{\theta\in\Theta^*}\|\nabla_\theta f(z_t,\theta)\|\right] < \infty$*.* $\mathbb{E}[f(z_t,\theta_0)'f(z_t,\theta_0)] < \infty$*, and* $\sup_{\theta\in\Theta}\|f(z_t,\theta)\| < \infty$ *for all* $t = 1,\ldots,T$ *and* $T = 1,2,\ldots$*. Each element of* $f(z_t,\theta_0)$ *is uniformly square integrable, for all* $t = 1,\ldots,T$ *and* $T = 1,2,\ldots$*.*

(v) The parameter space Θ *is a compact subset of* $\mathbb{R}^v$*.*

(vi) $\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(z_t,\theta)\right] = 0$*, only when* $\theta = \theta_0$

(vii) The sequence of positive definite weighting matrices $W_T \to_p \Sigma_{ff}^{-1}$*.*

(viii) $M \equiv \lim_{T\to\infty}\mathbb{E}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial f(z_t,\theta)}{\partial\theta'}|_{\theta=\theta_0}\right] \in \mathbb{R}^{m\times v}$ *has full column rank.*

I now discuss Assumption 1.3.1. Assumptions 1.3.1.(i) and 1.3.1.(ii) are asymptotic weak dependence and stationarity conditions on the data which are typical of those found in other literature on nonlinear dynamic models and are closest to the conditions given in Sowell (1996) or Rossi (2005).[11]. Together with Assumption 1.3.1.(iii) which assumes positive definiteness of the long-run variance of the sample moments, the assumptions are sufficient to obtain weak convergence of the partial sample moments to Brownian Motions using the multivariate functional central limit theorem of Phillips and Durlauf (1986). Assumption 1.3.1.(iv) are standard smoothness and boundedness conditions on the sample moment function under the null hypothesis $f(z,\theta)$. An analogue of this assumption is used in Sowell (1996). Together with Assumption 1.3.1.(v) which assumes a compact parameter space and Assumption 1.3.1.(i), the conditions ensure uniform convergence of the GMM objective function via the generic uniform law of large numbers of Andrews (1987). Together with Assumption 1.3.1.(vi) which assumes identification under the null hypothesis, the conditions are sufficient to obtain consistency of the GMM estimators used to construct the test statistic. Assumption 1.3.1.(vii) restricts the choice of weighting matrices used to construct the GMM estimators by requiring that an efficient GMM estimator is used. Finally, Assumption 1.3.1.(viii) which requires the gradient of the sample moment to have full rank ensures that the test statistic has a well-defined asymptotic variance.

The following theorem establishes the asymptotic distribution of the test statistic under the null hypothesis.

[11] The assumptions, in particular the weak stationarity condition, are stronger than necessary and the results presented in this paper are expected to hold if the assumptions are relaxed to the near-epoch-dependence case in Andrews (1993).

THEOREM 1.3.1 (Limiting distribution for in-sample tests): *Assume that the regularity conditions in Assumption 1.3.1 hold. Under the null hypothesis defined in (1.3), it holds that*

$$\sup \Phi_T(K) \;\Rightarrow\; \sup_{\lambda^K \,\in\, \Lambda_\epsilon} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j) - \mathcal{B}_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

$$D\sup \Phi_T(\bar{K}) \Rightarrow \max_{1\leq k\leq \bar{K}} (1/k) \sup_{\lambda^K \,\in\, \Lambda_\epsilon} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j) - \mathcal{B}_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

$$\Lambda_\epsilon \equiv \Big\{ \lambda_j : \;\; \lambda_j \in (\epsilon, 1-\epsilon),\; \lambda_j > \lambda_{j-1} + \epsilon,\;\; j = 1, \ldots, K \Big\}$$

where $\lambda_0 \equiv 0, \lambda_{K+1} \equiv 1$ *and* $\mathcal{B}_p(\cdot)$ *is a* $(p \times 1)$ *vector of independent standard Brownian motions on* $[0, 1]$.

The proof of this theorem is reported in Appendix C.

One can show that the limiting distribution of $\sup \Phi_T(K)$ is equivalent to

$$\sup_{\lambda \,\in\, \Lambda_\epsilon} \sum_{i=1}^{K} \frac{\|\lambda_{i+1}\, \mathcal{B}_p(\lambda_i) - \lambda_i\, \mathcal{B}_p(\lambda_{i+1})\|^2}{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)} \;+\; \mathcal{B}_p(1)'\, \mathcal{B}_p(1)$$

where the first term is the same as $(1/Kp)$ times the limiting distribution of the $\sup F$ test for parameter stability of Bai and Perron (1998) and depends on the number of splitting points K. The second component reflects the additional restrictions on β and does not depend on K. It is equivalent to the χ^2 distribution with p degrees of freedom which is the limiting distribution of a standard LM test conducted on the full sample. Further, note that for $K = 1$, the limiting distribution reduces to the limiting distribution of the QLR^*_T model selection test robust to instabilities which was proposed in Rossi (2005) and imposes one break.

Critical values of the test statistics can be obtained by directly simulating the limiting distributions in Theorem 1.3.1 using a dynamic programming algorithm analog to the one provided in Section 1.5. Table 1.1 reports a selection of critical values for the $D\sup \Phi_t(K)$ test for commonly used significance levels and trimming parameters for $p = 1, 2$. The critical values were obtained by simulating the asymptotic distributions based on 10,000 Monte Carlo replications and an approximation length of $N = 3,600$ for the Brownian Motions. Details and a full tabulation of the critical values for a wide array of values for p, ϵ and significance levels, α, are provided in Appendix A.

Table 1.1: Selected critical values for $D \sup \Phi_T$ tests

	$\epsilon = .05$			$\epsilon = .1$			$\epsilon = .15$		
p	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	10.12	11.57	15.20	9.39	10.99	14.54	8.84	10.48	14.17
2	14.01	15.79	19.96	13.30	15.06	19.20	12.80	14.65	18.62

Notes: This table reports simulated quantiles of the limiting distributions of the $D \sup \Phi_T$ tests. The critical values were obtained based on 10,000 Monte-Carlo replications and an approximation length of $N = 3,600$ observations for the partial sums to simulate the Brownian Motions. Appendix A provides the full table.

1.4 Out-of-sample inference

This section describes the relevant asymptotic theory to conduct out-of-sample inference. The limiting distributions derived in this section apply when the test statistics proposed in Section 1.2 are used in conjunction with a moment condition formulated on out-of-sample forecast errors. I start by discussing the forecasting environment, in particular how to obtain the sequence of estimates $\{\hat{\delta}_t\}_{t=T_0}^T$ which are used to construct the test statistics proposed in Section 1.2. I then present and discuss the required regularity conditions and derive the limiting distribution of the proposed test statistics.

Assume the available sample is of size $T + h$ and that the data $z_t = (y'_{t+h}, x'_t)$ includes a random variable y_t to be predicted h steps ahead as well as a vector of predictors, x_t. The sample is divided into an in-sample portion of length R and an out-of-sample portion of size P such that $R + P = T + h$. Given the sample split, forecasts of y_{t+h} for $t = R, \ldots, T$ are generated using parametric models of the form $y_{t+h} = g(x_t, \delta) + u_{t+h}$ for a known function $g(\cdot, \cdot)$ and an unknown q-dimensional parameter vector δ.

The parameters of the forecasting model are estimated based on a d-dimensional vector of moment equations $\mathbb{E}[h(z_t, \delta)] = 0$. This allows for a variety of estimation methods such as (nonlinear) least squares, maximum likelihood or generalized methods of moments. The parameters of the forecasting model are assumed to be estimated using a recursive scheme where the parameter vector is estimated at each $t = R, \ldots, T$ using all available information which yields a sequence of parameter estimates $\{\hat{\delta}_t\}_{t=R}^T$. The predictors and parameter estimates are used to generate forecasts $\hat{y}_{t+h} = \hat{g}_{t+h}(x_t, \hat{\delta}_t)$ for $t = R, \ldots, T$. which are used in turn to construct a series of forecast errors, $\hat{v}_{t+h} = y_{t+h} - \hat{y}_{t+h}$. Figure 1.1 illustrates the out-of-sample forecasting environment described above.

The following regularity assumptions are sufficient to obtain weak convergence of the test statistics under the null hypothesis to the limiting distributions characterized in the main theorem of this section.

Let $\xi(z_{t+h},\theta) \equiv \left[f(z_{t+h},\beta,\delta)', h_t(\delta)'\right]'$ be an $(m+d\times 1)$ vector stacking the moment functions. Further, let $F \equiv \mathbb{E}\left[\frac{\partial f(z_t,\theta)}{\partial \delta}|_{\theta=\theta_0}\right] \in \mathbb{R}^{m\times q}$. The following regularity conditions are assumed to hold under the null hypothesis.

ASSUMPTION 1.4.1 (Regularity conditions): *Assume the following regularity conditions hold.*

(i) *Assume that $h<\infty$ and that K is fixed while $R\to\infty$, $T\to\infty$ and $\lim_{T\to\infty} R/T = \rho \in (0,1)$.*

(ii) *The estimate $\hat{\delta}_t$ satisfies $\hat{\delta}_t - \delta_0 = B_t H_t$ where B_t is a $(q\times d)$ matrix which satisfies $B_t \xrightarrow{as} B$ where B has rank q and H_t is $(d\times 1)$ with $H_t = t^{-1}\sum_{r=1}^{t} h(z_r,\delta_0)$ (recursive estimation scheme) for a $(d\times 1)$ moment condition $h(z_r,\delta)$.*

(iii) *For some $p>\beta>2$, $\xi(z_{t+h},\theta_0)$ is zero mean, strong mixing with mixing coefficients α_m of size $-p\beta/(p-\beta)$ and it holds that $\sup_{t\geq 1}\|\xi(z_{t+h},\theta_0)\|_p = C<\infty$.*

(iv) *$\{z_{t+h}\}$ is weakly stationary. In addition, $\mathbb{E}\left[\xi(z_{t+h},\theta_0)\right] = 0$ for all $t=1,\ldots,T$ and $T=1,2,\ldots$.*

(v) *$\Sigma \equiv \lim_{T\to\infty} \mathbb{E}\left[T^{-1}\left\{\sum_{t=1}^{T}\xi(z_{t+h},\theta_0)\right\}\left\{\sum_{t=1}^{T}\xi(z_{t+h},\theta_0)\right\}'\right] \in \mathbb{R}^{(m+d)\times(m+d)}$ is positive definite.*

(vi) *$f(z,\theta)$ is continuously partially differentiable in θ in a neighborhood of θ_0 for every $\theta_0 \in \Theta^*$ where Θ^* is some convex or open set that contains Θ. The functions $f(z,\theta)$ and $\nabla_\theta f(z,\theta) \equiv \partial f(z,\theta)/\partial\theta$ are measurable functions of z for each $\theta\in\Theta$ and $\mathbb{E}\left[\sup_{\theta\in\Theta^*}\|\nabla_\theta f(z_t,\theta)\|\right] < \infty$. $\mathbb{E}[f(z_t,\theta_0)' f(z_t,\theta_0)] < \infty$, and $\sup_{\theta\in\Theta}\|f(z_t,\theta)\| < \infty$ for all $t=1,\ldots,T$ and $T=1,2,\ldots$. Each element of $f(z_t,\theta_0)$ is uniformly square integrable, for all $t=1,\ldots,T$ and $T=1,2,\ldots$.*

(vii) *The parameter space Θ is a compact subset of $\mathbb{R}^v$.*

(viii) *$\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(z_t,\theta)\right] = 0$, only when $\theta=\theta_0$*

(ix) *The sequence of positive definite weighting matrices $W_T \to_p \Sigma_{ff}^{-1}$.*

(x) *$M \equiv \lim_{T\to\infty}\mathbb{E}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial f(z_t,\theta)}{\partial\theta'}|_{\theta=\theta_0}\right] \in \mathbb{R}^{m\times v}$ has full column rank.*

(xi) $\lim_{T\to\infty}\sup_{r,s\in(0,1),s>r>\rho}\ T^{-1/2}\ \sum_{t=[rT]+1}^{[sT]}\left(\nabla_\delta f_t(\theta_0,\delta_0) - F\right)BH_t = o_p(1)$

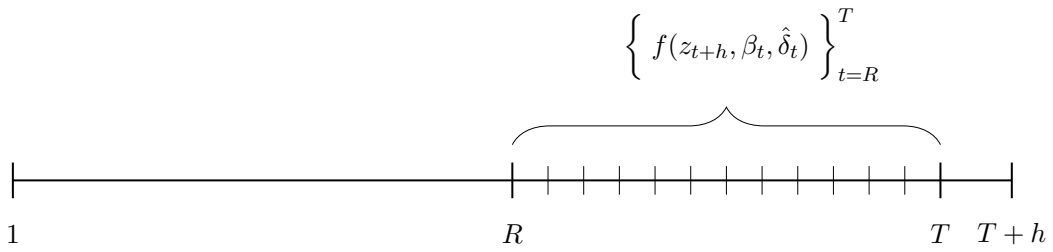
(xii) $\lim_{T\to\infty}\sup_{r,s\in(0,1),s>r>\rho}\ T^{-1/2}\ F\sum_{t=[rT]+1}^{[sT]}(B_t - B)H_t = o_p(1)$

(xiii) $\lim_{T\to\infty}\sup_{r,s\in(0,1),s>r>\rho}\ T^{-1/2}\ \sum_{t=[rT]+1}^{[sT]}\left(\nabla_\delta f_t(\theta_0,\delta_0) - F\right)(B_t - B)H_t = o_p(1)$

(xiv) $\lim_{T\to\infty}\sup_{r,s\in(0,1),s>r>\rho}\ \left[T^{-1}\ \sum_{t=[rT]+1}^{[sT]}\left(\nabla_\theta f_t(\theta_0,\delta_0) - M\right)\right] = o_p(1)$

I now discuss Assumption 1.4.1. Assumption 1.4.1.(i) defines the relevant asymptotic experiment as one where both the size of the in-sample and out-of-sample portions diverge to infinity while the size of the in-sample and out-of-sample portions remains a fixed proportion of the total sample size. Assumption 1.4.1.(ii) is a regularity assumption on the sequence of parameter estimates of the forecasting model. It allows for a variety of estimation methods such as (nonlinear) least squares, maximum likelihood or generalized

Figure 1.1: Illustration of the out-of-sample forecasting environment



Notes: The figure shows how the data sample is partitioned into an in-sample and out-of-sample portion at R. A forecasting model specified by the researcher is used to generate a sequence of parameter estimates $\{\hat{\delta}_t\}$ using data from the in-sample portion of the sample up to a specified point in time. The forecasting model is used to obtain a sequence of forecast errors for the out-of-sample portion of the data which enters the moment condition. The test builds on the sample moments in the out-of-sample portion of the data $t = R, \ldots, T$.

methods of moments. In addition, the assumption describes the recursive parameter estimation scheme. This assumption is typical of the literature on forecast evaluation, see West and McCracken (1998) or Rossi and Sekhposyan (2016). Assumptions 1.4.1.(iii) and 1.4.1.(iv) are weak dependence assumptions equivalent to the ones discussed in the previous section. Together with Assumption 1.4.1.(v) which assumes positive definiteness of the long-run variance, the assumptions are sufficient to obtain weak convergence of the partial sample moments to Brownian Motions using the multivariate functional central limit theorem of Phillips and Durlauf (1986). Assumption 1.4.1.(vi) are standard smoothness and boundedness condition on the sample moment function under the null hypothesis $f(z, \theta)$. An analogue of this assumption is used in Sowell (1996). Together with assumption 1.4.1.(vii) which assumes a compact parameter space and Assumption 1.4.1.(viii) which assumes identification under the null hypothesis, the conditions are sufficient to yield consistency of the GMM estimator used to construct the test statistic. Assumption 1.4.1.(ix) restricts the choice of weighting matrices used to construct the GMM estimators by requiring that an efficient GMM estimator is used. Assumption 1.4.1.(x) ensures that the test statistic has a well-defined asymptotic variance. Finally, Assumptions 1.4.1.(xi)-1.4.1.(xiv) are boundedness conditions which guarantee that the remainder of a mean-value expansion of the sample moments around δ is asymptotically negligible.

Before I derive the limiting distribution, I provide some intuition on why the limiting distribution in the out-of-sample case differs from the limiting distribution in the in-sample case which was derived in Section 1.3. The crucial difference to the in-sample case is that the moment functions $f(z_t, \cdot, \cdot)$ depend on the sequence of estimated parameters $\{\hat{\delta}_t\}_{t=R}^{T}$ which were obtained from a separate moment condition. This makes it necessary to take parameter estimation error in δ explicitly into account when evaluating the limiting distribution of the test statistics proposed in Section 1.2 (see West (1996) for a similar argument).

Under the regularity conditions provided in Assumption 1.4.1, the following mean-value approximation of the partial sample moments evaluated at $\{\hat{\delta}_t\}_{t=R}^{T}$ holds.

LEMMA 1.4.1 (OOS Mean-Value Approximation): *Under the regularity conditions in Assumption 1.4.1 and the null hypothesis defined in (1.3), for any* $r, s \in [0,1]$ *with* $s > r > \rho$ *it holds that*

$$
\begin{aligned}
&P^{-1/2} \sum_{t=[rT]+1}^{[sT]} f(z_{t+h}, \beta_0, \hat{\delta}_t) = \\
&(T/P)^{1/2} \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[sT]} f(z_{t+h}, \beta_0, \delta_0) - \frac{1}{\sqrt{T}} \sum_{t=R}^{[rT]} f(z_{t+h}, \beta_0, \delta_0) \right\} \\
&+ (T/P)^{1/2} \, FB \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[sT]} H_t(\delta_0) - \frac{1}{\sqrt{T}} \sum_{t=R}^{[rT]} H_t(\delta_0) \right\} + o_{p,rs}(1)
\end{aligned}
$$

where H_t, B *are as defined in Assumption 1.4.1.(ii) and* $x_t(r,s) = o_{p,rs}(1)$ *denotes that* $\sup_{r,s \in [0,1],\ s>r>\rho} \|x_t(r,s)\| = o_p(1)$.

The proof of the Lemma is reported in Appendix C.2.

The expansion in Lemma 1.4.1 decomposes the partial sample moment into two terms. The first term on the right hand side represents uncertainty that is present even if δ_0 is known. The second part reflects uncertainty about δ_0 originating from estimating the parameters of the forecasting model δ based on the moment function $\mathbb{E}\left[h(z_t, \theta)\right] = 0$. This is in contrast to the in-sample case where δ is estimated using the same moment condition which is used to construct the test statistic, $\mathbb{E}\left[f(z_t, \theta_t)\right] = 0$. Further, note that whether parameter estimation error in δ needs to be taken into account crucially depends on F having a non-zero value.

The following theorem establishes the asymptotic distribution of the test statistic under the null hypothesis.

THEOREM 1.4.1 (OOS Inference): *Assume that the regularity conditions in Assumption 1.4.1 hold. Under the null hypothesis defined in (1.3), it holds that*

$$\sup \Phi_T(K) \;\Rightarrow\; \sup_{\lambda^K \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \Phi_j(\lambda_{j-1}, \lambda_j)$$

$$D \sup \Phi_T(\bar{K}) \;\Rightarrow\; \max_{1 \leq k \leq \bar{K}} (1/k) \sup_{\lambda^K \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \Phi_j(\lambda_{j-1}, \lambda_j)$$

$$\Lambda_{\epsilon,\rho} = \Big\{ \lambda_j,\ j = 1, \ldots, K :\ \lambda_j \in (\rho + \epsilon, 1 - \epsilon),\ \lambda_j > \lambda_{j-1} + \epsilon \Big\},$$

with $\lambda_0 \equiv \rho,\ \lambda_{K+1} \equiv 1$ *and where*

$$\begin{aligned} \Phi_j(\lambda_{j-1}, \lambda_j) \equiv & \left[\mathcal{B}_m \left(\int_0^{\lambda_j} \omega(u, \lambda_{j-1}, \lambda_j) \omega(u, \lambda_{j-1}, \lambda_j)' \, du \right) \right]' \times \\ & \left\{ \int_0^{\lambda_j} \omega(u, \lambda_{j-1}, \lambda_j)\ \omega(u, \lambda_{j-1}, \lambda_j)' \, du \right\}^{-1} \\ & \times \left[\mathcal{B}_m \left(\int_0^{\lambda_j} \omega(u, \lambda_{j-1}, \lambda_j) \omega(u, \lambda_{j-1}, \lambda_j)' \, du \right) \right] \end{aligned}$$

with

$$\begin{aligned} \omega(u, r, s) \equiv M' \, \Sigma_{ff}^{-1} \, (1 - \rho)^{-1/2} \Big[I_m \quad FB \Big] \times \Big\{ & \Big[\Omega(u, s)^{1/2} - \Omega(u, r)^{1/2} \Big] \mathbb{1}(u \leq r) \\ & + \Omega(u, s)^{1/2} \, \mathbb{1}(r < u \leq s) \Big\} \, \Sigma^{1/2} \end{aligned}$$

and where $\Omega(s, \tau)$ *is as defined as*

$$\Omega(s, \tau)^{1/2} \equiv \begin{pmatrix} \mathbb{1}(s \leq \rho) \cdot I_m & 0_{m \times d} \\ 0_{d \times m} & \big\{ [\ln \tau - \ln \rho] \, \mathbb{1}(s \leq \rho) + [\ln(\tau) - \ln(s)] \, \mathbb{1}(\rho < s \leq \tau) \big\} \cdot I_d \end{pmatrix}$$

The proof of this theorem can be found in Appendix C.2.

Note that in the general case, the limiting distribution in Theorem 1.4.1 depends on nuisance parameters of the data-generating process and has to be simulated for each application.[12]

[12] Critical values for this limiting distribution can be simulated by a dynamic programming algorithm similar to the one discussed in Section 1.5 of this paper where Φ_j is simulated using a modification of the algorithm described Rossi and Sekhposyan (2016).

Forecast unbiasedness, efficiency tests and survey forecasts

The general result presented in Theorem 1.4.1 above simplifies considerably in two cases that are of great interest to practitioners.[13]

The first case applies when parameter estimation error in $\hat{\delta}$ is irrelevant i.e. if it holds that $F = 0$. This case is particularly of interest when the tests are used to evaluate the out-of-sample predictive ability of survey forecasts where the model which generated the forecasts is not available and thus the correction for parameter estimation error can not be applied. Relevant examples of such forecasts are survey and judgemental forecasts produced by central banks such as the Greenbook projections produced by the Federal Reserve Board or private sector forecasts such as the Survey of Professional Forecasters (SPF) or the Blue-Chip Economic Indicators (BCEI).

The second case applies when the parameter estimation error is asymptotically negligible. This case was discussed in West and McCracken (1998) Corollary 5 for the case of regression-based tests of out-of-sample predictive ability based on the full-sample and is also considered in Rossi and Sekhposyan (2016). In such cases, a special condition holds which considerably simplifies the asymptotic distributions of the proposed test statistic. The condition is given in Corollary 1.4.1 below. This condition is satisfied in many applications of interest to empirical researchers, particularly tests for forecast unbiasedness and efficiency under general conditions as well as several other tests under more specific assumptions. These cases are discussed in West and McCracken (1998).

The limiting distribution for the special cases is provided in the following Corollary of Theorem 1.4.1.

COROLLARY 1.4.1 (OOS Inference in Special Cases): *If (a) $F = 0$, that is parameter estimation error is irrelevant, or (b) the following condition holds*

$$\Sigma_{ff} = -\frac{1}{2}(FB\Sigma_{hf} + \Sigma_{fh}B'F') = FB\Sigma_{hh}B'F'$$

then, the result of Theorem 1.4.1 simplifies to

$$\sup \Phi_T(K) \Rightarrow \sup_{\lambda \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j - \rho) - \mathcal{B}_p(\lambda_{j-1} - \rho)\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

$$D\sup \Phi_T(\bar{K}) \Rightarrow \max_{1 \leq k \leq \bar{K}} (1/k) \sup_{\lambda \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j - \rho) - \mathcal{B}_p(\lambda_{j-1} - \rho)\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

The proof of this corollary can be found in Appendix C.2.

[13] These special cases were also considered in Rossi and Sekhposyan (2016).

Note the similarities between the limiting distribution in the in-sample case which was discussed in Section 1.3 and the limiting distribution in the special cases provided above. In particular, the limiting distribution of the in-sample case is obtained when setting $\rho = 0$. Under the special cases, the critical values do not depend on the data-generating process and can be tabulated.[14] In the case where the tests are applied to survey or judgemental forecasts, the only sample available to researchers is $t = R, \ldots, T$. In particular, the researcher cannot specify a value of ρ as the length of the in-sample portion is unknown. In these cases, critical values for the proposed tests can be obtained by setting $\rho = 0$ in the limiting distribution above and the critical values provided in Section 1.6 can be used to conduct the test.

1.5 Implementation

This section gives detailed instructions on how to implement the tests.

1.5.1 Variance estimators

To implement the test statistics defined in (1.5) and (1.11), we require the estimators $\hat{\Sigma}_{ff}$ and $\hat{\Omega}_{T,j}$ that appear in the formulas of the LM and Wald statistics.

Computation of $\hat{\Sigma}$ depends on whether there is serial correlation in the moment conditions. When $f(z_t, \theta_0)$ consists of mean-zero uncorrelated random variables, a consistent estimator is given by

$$\hat{\Sigma}_{ff} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \left[f(z_t, \tilde{\theta}) - \bar{f}_T(\tilde{\theta}) \right] \left[f(z_t, \tilde{\theta}) - \bar{f}_T(\tilde{\theta}) \right]' \tag{1.15}$$

$$\bar{f}_T(\tilde{\theta}) \equiv \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} f(z_t, \tilde{\theta}) \tag{1.16}$$

where $\bar{f}_T(\tilde{\theta})$ is the mean of the sample moments. Alternatively, if $f(z_t, \theta_0)$ consists of mean-zero but serially correlated random variables, then a consistent estimator is given by a kernel-based HAC estimator such as

$$\begin{aligned} \hat{\Sigma}_{ff} = &\sum_{l=0}^{T-1} \left\{ \kappa(l/q_T) \frac{1}{T - T_0 + 1} \sum_{t=l+T_0}^{T} \left(f(z_t, \tilde{\theta}) - \bar{f}_T(\tilde{\theta}) \right) \left(f(z_{t-l}, \tilde{\theta}) - \bar{f}_T(\tilde{\theta}) \right)' \right\} \\ &+ \sum_{l=1}^{T-1} \left\{ \kappa(l/q_T) \frac{1}{T - T_0 + 1} \sum_{t=l+T_0}^{T} \left(f(z_{t-l}, \tilde{\theta}) - \bar{f}_T(\tilde{\theta}) \right) \left(f(z_t, \tilde{\theta}) - \bar{f}_T(\tilde{\theta}) \right)' \right\} \end{aligned} \tag{1.17}$$

where $\kappa(\cdot)$ is a kernel and q_T a bandwidth parameter which can depend on the data. A kernel choice that guarantees that the estimator $\hat{\Sigma}_{ff}$ is positive semi-definite is the Quadratic Spectral Kernel discussed in Andrews (1991).

[14] A table of these critical values is available upon request.

Next, consider the estimator $\hat{\Omega}_{j,T}$. This estimator crucially depends on whether the test is conducted in-sample or out-of-sample. In the case where the test is conducted in-sample, $\hat{\Omega}_{T,j}$ can be computed from simple formulas. Specifically, $\hat{\Omega}_{j,T}$ can be computed as

$$\begin{aligned}
\hat{\Omega}_{j,T} &= (\lambda_j - \lambda_{j-1})^{-1}\, \hat{C}'(\hat{C}\hat{C}')^{-1}\hat{C} \\
\hat{C} &\equiv \bar{M}_\beta'(I_m - \bar{P}_\delta) \\
\bar{P}_\delta &\equiv \bar{M}_\delta(\bar{M}_\delta'\bar{M}_\delta)^{-1}\bar{M}_\delta' \\
\bar{M} &= \hat{\Sigma}_{ff}^{-1/2}\, \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \frac{\partial f_t(z_t, \theta_t)}{\partial \theta'}|_{\theta_t = \tilde{\theta}}
\end{aligned} \tag{1.18}$$

where $\bar{M}_\beta$ and $\bar{M}_\delta$ are obtained from partitioning $\bar{M} = (\bar{M}_\beta, \bar{M}_\delta)$. $\tilde{\theta}$ is the restricted GMM estimator defined in (1.7).

1.5.2 Dynamic programming algorithm

This section discusses how to compute the test statistics described above via an efficient dynamic programming algorithm. First, note that *given a fixed vector of sample splits,* λ, the computation of the $\Phi_{T,j}$ parts of the test statistic defined in equation (1.6) is straightforward. It simply requires computing the restricted GMM estimator defined in equation (1.7) and computing the components of equation (1.6) via the estimators provided in the previous section.

However, to compute the $\sup \Phi_T$ and $D \sup \Phi_T$ test statistics which allow for an *unknown vector of sample splits*, one has to compute the $\sup$ operator over $\lambda \in \Lambda_\epsilon$. This is computationally challenging as it involves computing a series of test statistics $\{\Phi_T(\lambda_{j-1}, \lambda_j)\}_{j=0}^{K+1}$ for *every possible partition of the sample into K segments*, respecting the minimal segment length implicitly defined by the trimming parameter, ϵ. In principle, a grid search procedure could be used, but with $K > 2$ this becomes quickly infeasible as it involves the computation of $\Phi_T(\cdot)$ of order $\mathcal{O}(T^K)$.

To solve the computational problem, I employ a dynamic programming algorithm which efficiently computes the $\sup \Phi_T$ and $D \sup \Phi_T$ statistics in $\mathcal{O}(T^2)$ operations, regardless of the value of K. The algorithm is based on the early work of Hawkins (1976) and extensions by Bai and Perron (1998, 2003) and Qu and Perron (2007).[15]

The basic idea of the algorithm is as follows. For any given number of changes, K, the $\sup \Phi_T$ test statistic in equation (1.5) is given by the sum $\Phi_{T,j}(\cdot)$ statistics for

[15] The dynamic programming algorithm of Bai and Perron (1998, 2003) computes the sum of squared residuals (SSR) of a linear regression model for every possible partition of the sample into K+1 regimes. Qu and Perron (2007) extend this algorithm to compute QMLE estimates assuming a linear pseudo-model with gaussian errors. In contrast, I modify the dynamic programming algorithm to directly compute the sup-Φ_T test statistics which are functions of sums of partial sample moments as well as the restricted GMM estimator.

$j = 1, \ldots K+1$, which are associated with a specific partition of the sample defined by $\lambda = (\lambda_1, \ldots, \lambda_K)$. The problem of computing the sup over all possible values of $\lambda \in \Lambda_\epsilon$ can therefore be transformed into several steps which are described in the following algorithm.

ALGORITHM 1.5.1 (Computation of $\sup \Phi_T(K)$ and $D \sup \Phi_T(\bar{K})$ tests): *The $\sup \Phi_T(K)$ test is computed by implementing Steps 1 and 2 of the following algorithm. The $D \sup \Phi_T$ test is computed by implementing Steps 1-3, setting $K = \bar{K}$ for the first two steps.*

Step 1: *Compute and store all possible segments of the test statistic $\Phi_{T,j}(T_m, T_n) := \Phi_{T,j}([\lambda_m T], [\lambda_n T])$ for $T_m, T_n \in t = 1, \ldots, T$ which satisfy $T_m > T_n$ and $T_r := [\lambda_r T]$. In the case of the LM test, $\Phi_{T,j}$ is as defined in equation (1.6).*

Step 2: *Recursively maximize the sum of k+1 of these partitions for $k = 1, 2, \ldots, K$ using the following Bellman equation.*

$$\Phi_T(\{\lambda_{k,T}\}) = \max_{kh \leq T_j \leq T-h} \left[\Phi_T(\{\lambda_{k-1,T_j}\}) + \Phi_T(T_j + 1, T)\right] \tag{1.19}$$

where $\Phi_T(\{\lambda_{k,T_j}\})$ denotes the value of the $\sup \Phi_T$ statistic associated with an optimal partition based on k changes and using observations $t = 1, \ldots, T_j$ and $h = [\epsilon T]$ is the minimum segment length implied by the trimming parameter. The recursion is initialized with $\Phi_T(\{\lambda_{0,T_j}\}) \equiv \Phi_T(T_j + 1, T)$.

Step 3: *Carrying out Steps 1-2 with $K = \bar{K}$ yields a series of test statistics under k changes, $\left\{\Phi_T(\{\lambda_{k,T}\})\right\}_{k=1}^{\bar{K}}$. Then, compute $D \sup \Phi_T(\bar{K})$ as in equation (1.11) by first normalizing each element of the series, dividing by the respective k, and computing $D \sup \Phi_T(\bar{K})$ as the maximum element of the normalized series.*

An efficient software implementation of the steps above is described in Appendix A. Note that the computation of all possible segments in Step 1 requires computing less than $T(T+1)/2$ times the Φ_T test statistics and is therefore of order $\mathcal{O}(T^2)$.[16]

As stated previously, the test presented in Section 1.2 can be conducted based on a Lagrange-Multiplier form (Φ_T^{LM}) or a Wald-form (Φ_T^W). If all coefficients of the model are to be tested (i.e. $\theta = \beta$ is the full parameter vector), both the LM and Wald tests can be computed via the algorithm described above. However, when the test is carried out on a subvector of θ, the algorithm above needs to be modified to compute the Wald test. One has to augment the steps above with another layer, conditioning on estimates

[16] Depending on the value of the trimming parameter, ϵ, substantially less than $T(T+1)/2$ computations are needed. This is because one only has to consider segments which have a minimum length of $h = [\epsilon T]$ observations. Further, for specific models, additional computational simplifications are possible by using an updating rule to compute Φ_T.

$\hat{\delta}$ under the alternative hypothesis and iterating until convergence.[17] Therefore, the LM test has considerable computational benefits over the Wald versions as it only requires the computation of the restricted GMM estimator under the null hypothesis.

1.6 Simulation studies

To investigate the finite sample performance of the specification tests proposed in this paper, I conduct a series of Monte-Carlo experiments using several data-generating processes. The goal of the simulation exercises is three-fold: First, they illustrate the difference between the proposed model specification test, tests of multiple structural breaks and traditional hypothesis tests assuming a constant parameter. Second, they assess the finite-sample properties of the proposed tests for a variety of data-generating processes. Third, they study under which conditions allowing for multiple shifts in the coefficient vector leads to power gains relative to tests imposing one break. In what follows, I describe each simulation exercise, define the respective data-generating processes and discuss the simulation results.

1.6.1 Asymptotic power illustration

Before studying the finite-sample approximation quality of the limiting distributions, I conduct an asymptotic power exercise to illustrate why it is useful to jointly test the hypotheses in (1.3) rather than relying on traditional hypothesis tests or traditional structural break tests. Recall from the discussion in Section 1.2 that traditional tests are designed to test *either* $H_0^{(1)}$ or $H_0^{(2)}$ and therefore might not reliably detect departures originating in the other part of null hypothesis. In contrast, the tests proposed in this paper, *jointly* test $H_0^{(1)}$ and $H_0^{(2)}$ and reject against any combination of these hypotheses. To illustrate this argument, I conduct a simulation exercise using the following simple data-generating process:

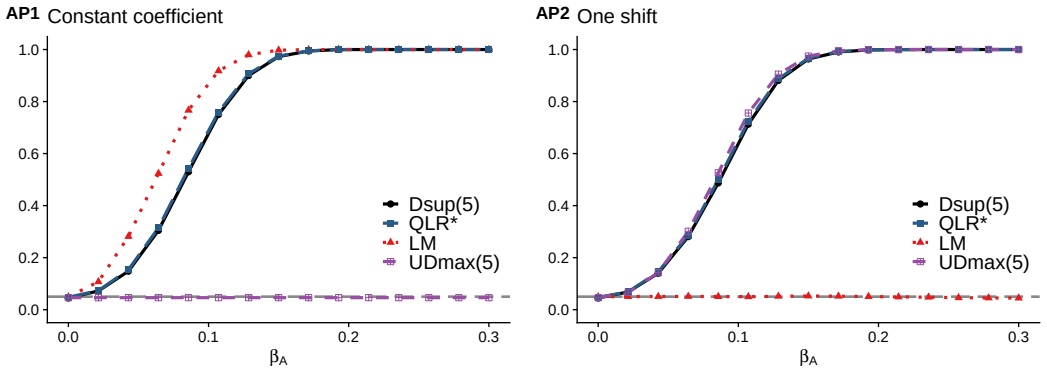
$$y_{t+1} = \beta_{t,T}\, x_t + \eta_t \qquad\qquad x_t, \eta_t \sim iid\ \ \mathcal{N}(0,1) \tag{1.20}$$

I simulate data from this model using a large sample size of $T = 1,000$ and inspect rejection rates of (i) a full-sample LM test, (ii) Bai and Perron (1998)'s UDmax(5) test, (iii) Rossi (2005)'s QLR_T^* test and (iv) the $D\sup(5)$ test proposed in this paper against two stylized data generating-processes capturing a departure from $H_0^{(1)}$ and $H_0^{(2)}$, respectively.[18] The next two paragraphs briefly describe the two considered designs and discuss the simulated power curves which are reported in Figure 1.2.

[17] A similar strategy to compute Wald tests has been used in the structural break literature, see e.g. the algorithms in Qu and Perron (2007) and Bai and Perron (2003). Details on the modified algorithm are available from the author on request.

[18] All tests are conducted at 95% level and a trimming parameter of $\epsilon = 0.05$ is used for tests (ii)-(iv).

Figure 1.2: Asymptotic power illustration



Notes: The figure shows simulated rejection rates for tests of the null hypothesis $H_0 : \beta = 0$ in model (1.20) with 5% nominal size. $D\sup(5)$ denotes the proposed instability-robust test with a maximum of $\bar{K} = 5$ shifts in the parameter vector. LM denotes a traditional full-sample Lagrange-Multiplier test. QLR^* denotes Rossi (2005)'s QLR^*_T test which imposes one break. 'UDmax(5)' denotes Bai and Perron (1998)'s UDmax test with a maximum of 5 breaks. Rejection rates are based on 5,000 replications and $T = 1,000$ observations.

DESIGN AP1. The first design has $\beta_{t,T} = \beta_A \; \forall \; t$ i.e. the parameter which is tested is not time-varying. Figure 1.2 panel AP1 shows the "asymptotic" power of the tests as a function of β_A. It illustrates that when the parameter is not time-varying, the full-sample parameter test (LM) is the most powerful test among the four tests considered. In contrast, the test for a structural break (UDmax(5)) has a flat power function equal to the nominal size 5%. The test robust to multiple instabilities proposed in this paper ($Dsup(5)$) exhibits high power against this alternative.

DESIGN AP2. The second design has $\beta_{t,T} = \beta_A \mathbb{1}(t \leq T/2) - \beta_A \mathbb{1}(t > T/2)$ i.e. there is a single shift in the parameter which is tested. Figure 1.2 panel AP2 shows the "asymptotic power" of the tests as a function of β_A. It illustrates that this shift is not detected by a traditional hypothesis test (LM) which has a flat power function equal to the nominal size 5%. In contrast, the structural break test ($UDmax$) and the instability-robust tests (QLR^* and $D\sup(5)$) exhibit substantial power against this alternative. Finally, it is important to note that while the QLR^* test is optimal for the case of one break, both the structural break test and the proposed $D\sup(5)$ test exhibit virtually the same power.

1.6.2 Finite-sample size

Next, I assess the quality of the finite-sample approximation of the limiting distribution of the test. I first focus on the finite-sample size of the proposed test procedure. Size control in finite-samples is an important feature of any test procedure since a researcher choosing a particular significance level α expects the test to reject only in $(1 - \alpha)\%$ of cases when the null hypothesis is true. I study a data-generating process resembling a linear regression which predicts a scalar series, y_t, with past values of a predictor, x_t, and a control variable,

w_t correlated with the predictor. Both the prediction error and the predictor variable admit serial correlation in the form of an AR(1) process. This class of models has received considerable attention in the predictability literature, both in macroeconomics and finance (see the reviews in Pitarakis and Gonzalo (2019) and Rossi (2013)). The data-generating process is defined as follows

$$y_t = \beta_{t,T}\, x_t + \delta\, w_t \; + \eta_t \qquad \eta_t = \phi_\eta\, \eta_{t-1} + \zeta_t \tag{1.21}$$

$$x_t = \phi_x\, x_{t-1} + \xi_t \qquad w_t = \rho_{xw}\, x_t + \iota_t \tag{1.22}$$

where ζ_t, ξ_t, ι_t are independent and $iid\, \mathcal{N}(0,1)$ and x_0, η_0 are drawn from the unconditional distribution of the respective AR process.

To assess the finite-sample size of the tests, I simulate 10,000 samples from the model above, imposing the null hypothesis $\beta_{t,T} = 0$, and compute rejection rates for (i) the traditional LM test, (ii) Rossi (2005)'s QLR_T^* test and (iii) the $D\sup(5)$ test proposed in this paper. All tests are conducted on the β subvector while leaving δ unspecified. I simulate specifications with sample sizes $T \in \{125, 250, 500, 1000\}$ and various degrees of serial correlation in the predictor and prediction errors, $\phi_\eta, \phi_x \in \{0, 0.25, 0.5\}$ and consider tests both with and without HAC correction.[19] In all specifications, the correlation between the predictor and control variable is fixed at $\rho_{xw} = 0.25$.

RESULTS. Table 1.2 reports the results regarding the empirical size of the model specification tests for the model in equation (1.21) for 5% nominal size and a trimming parameter of $\epsilon = 0.05$. First, consider the case in which no serial correlation is present in the data ($\phi_x = 0, \phi_\eta = 0$). When constructed using a heteroskedasticity-robust variance estimator (upper-left corner of Panel A), we observe that the proposed $D\sup$ test allowing for up to $\bar{K} = 5$ shifts exhibits good size control with finite-sample size being virtually identical to the size of Rossi (2005)' QLR_T test which imposes one break. In comparison to the traditional LM test, both instability-robust tests are slightly undersized in small samples, but size quickly converges to the nominal level as T grows. Using a variance estimator with HAC correction as described above yields similar size results; only in small samples are the $D\sup(5)$ and QLR_T^* more conservative than without the HAC correction (upper-left corner of Panel B). Next, consider the case with serial correlation in the predictor and/or prediction error. If serial correlation is ignored and the tests is constructed using a heteroskedasticity-robust variance estimator, size control crucially depends on the structure and amount of serial correlation in the data. The table illustrates that when serial correlation is present only in the predictor variable or the prediction errors (second and third row/column of Panel A), finite-sample size is barely affected. However, when serial correlation is present in both model components, all three tests become oversized with rejection rates growing both with

[19] The HAC correction is based on AR(1) approximation using Andrews (1991) data-dependent method and a Quadratic-Spectral kernel.

ϕ_x and ϕ_η. In that case, the instability-robust tests have significantly larger size-distortions than the LM tests with the proposed $D\sup(5)$ test exhibiting mildly worse size control than the QLR^* test. However, when the serial correlation is acknowledged and the test is constructed using the variance estimator with HAC correction provided in Section 1.2, the $D\sup(5)$ test recovers good size control and nominal size is close to 5% in medium to large samples. Finally, it is interesting to assess the robustness of these findings to choosing a larger trimming parameter. As Table B.5 in the appendix shows, increasing the trimming parameter to $\epsilon = 0.1$ yields nearly identical results.

To conclude, the simulation exercise shows that, when constructed using the appropriate variance estimator, the $D\sup(K)$ test exhibits good size control across a variety of specifications with finite-sample size comparable to that of Rossi (2005)'s QLR^*_T test. This implies that it is possible to allow for more than one break under the alternative without incurring a penalty with respect to the finite-sample size of the testing procedure.

1.6.3 Finite-sample power

Finally, I examine the finite-sample power of the proposed test. Understanding the power properties of (correctly sized) testing procedures is important as a researcher ideally would like to choose the testing procedure that maximizes the chances of correctly detecting a departure from the null hypothesis based on an available data sample. As power properties typically depend on the data-generating process considered, a careful analysis of finite-sample rejection rates helps to understand under which conditions testing procedures should be used. To assess finite-sample power of the proposed tests, I focus on a class of alternatives where $\beta_{t,T}$ exhibits local departures from the null hypothesis, β_0. Specifically, I focus on alternatives in which the coefficient has a value of zero for the majority of the sample, but there are multiple short episodes during which the coefficient departs from zero. This type of data-generating process has received considerable attention in the predictability literature in recent years (see e.g. the "pockets of predictability hypothesis" in Timmermann (2008) and Farmer et al. (2019) and empirical studies Gonzalo and Pitarakis (2012, 2017) and Rossi (2020).

As in the size simulations discussed in the previous subsection, I employ the following data-generating process

$$y_t = \beta_{t,T}\, x_t + \delta\, w_t \; + \eta_t \qquad \eta_t = \phi_\eta\, \eta_{t-1} + \zeta_t \tag{1.23}$$

$$x_t = \phi_x\, x_{t-1} + \xi_t \qquad w_t = \rho_{xw}\, x_t + \iota_t \tag{1.24}$$

where $\zeta_t, \xi_t, \iota_t \sim iid\;\; \mathcal{N}(0,1)$ and x_0, η_0 are drawn from the unconditional distribution of the respective AR process.

To assess the finite-sample power of the tests, I simulate 5,000 samples from the model above and compute rejection rates for (i) the traditional LM test, (ii) Rossi (2005)'s QLR^*_T

Table 1.2: Finite-sample size for $\epsilon = 0.05$ (nominal size 5%)

		$\phi_\eta = 0$			$\phi_\eta = 0.25$			$\phi_\eta = 0.5$		
T	ϕ_x	Dsup(5)	LM	QLR^*	Dsup(5)	LM	QLR^*	Dsup(5)	LM	QLR^*
Panel A: Heteroskedasticity-robust										
125	0.00	0.043	0.052	0.040	0.047	0.051	0.044	0.057	0.051	0.048
250	0.00	0.042	0.046	0.040	0.043	0.047	0.041	0.049	0.047	0.042
500	0.00	0.044	0.049	0.042	0.045	0.049	0.044	0.051	0.048	0.045
1,000	0.00	0.047	0.050	0.046	0.047	0.050	0.048	0.049	0.052	0.050
125	0.25	0.043	0.051	0.040	0.074	0.065	0.064	0.133	0.081	0.102
250	0.25	0.046	0.045	0.042	0.074	0.061	0.068	0.130	0.078	0.108
500	0.25	0.043	0.051	0.044	0.080	0.066	0.075	0.138	0.079	0.120
1,000	0.25	0.049	0.051	0.046	0.083	0.064	0.078	0.143	0.081	0.128
125	0.50	0.054	0.051	0.045	0.126	0.080	0.101	0.265	0.118	0.194
250	0.50	0.049	0.048	0.044	0.130	0.080	0.110	0.289	0.121	0.224
500	0.50	0.048	0.048	0.045	0.135	0.080	0.119	0.310	0.119	0.247
1,000	0.50	0.048	0.053	0.050	0.141	0.085	0.125	0.330	0.125	0.263
Panel B: HAC - AR(1) approximation, QS kernel, Andrews (1991) bandwidth										
125	0.00	0.032	0.047	0.030	0.031	0.048	0.032	0.036	0.046	0.035
250	0.00	0.036	0.044	0.034	0.036	0.045	0.035	0.039	0.044	0.036
500	0.00	0.043	0.049	0.041	0.041	0.049	0.042	0.046	0.048	0.042
1,000	0.00	0.045	0.049	0.044	0.045	0.050	0.046	0.046	0.051	0.048
125	0.25	0.030	0.046	0.030	0.034	0.052	0.035	0.039	0.053	0.038
250	0.25	0.037	0.044	0.036	0.041	0.050	0.041	0.045	0.052	0.043
500	0.25	0.040	0.049	0.042	0.048	0.054	0.049	0.054	0.054	0.048
1,000	0.25	0.047	0.050	0.045	0.054	0.053	0.052	0.057	0.055	0.054
125	0.50	0.033	0.047	0.034	0.038	0.053	0.039	0.035	0.053	0.036
250	0.50	0.040	0.045	0.038	0.046	0.052	0.046	0.045	0.053	0.044
500	0.50	0.042	0.048	0.040	0.051	0.054	0.048	0.050	0.052	0.049
1,000	0.50	0.047	0.052	0.046	0.055	0.056	0.056	0.055	0.054	0.053

Notes: The table reports simulated finite-sample size for tests of the null hypothesis $H_0 : \beta = 0$ with 5% nominal size. $D\sup(5)$ denotes the proposed instability-robust test with a maximum of $\bar{K} = 5$ breaks, LM denotes a traditional LM test and QLR^* denotes Rossi (2005)'s instability-robust test imposing one break. Rejection rates are based on 10,000 replications from the model in equation (1.21) using a sample of T observations where the serial correlation of the predictor and prediction error is controlled by ϕ_x, ϕ_η, respectively.

test and (iii) the $D\sup(5)$ test proposed in this paper. All tests are conducted on the β subvector while leaving δ unspecified. The instability-robust tests use a trimming parameter of $\epsilon = 0.05$. I focus on a sample size $T = 400$ to ensure all tests have a similar size and raw power can be compared between the tests. As in the size simulations, the correlation between the predictor and control variable is fixed at $\rho_{xw} = 0.25$. For clarity of exposition, the main text presents results for the case without serial correlation and using a heteroskedasticity-robust variance estimator. Appendix B reports additional results for $\phi_x = 0.5, \phi_\eta = 0.5$ and a variance estimator with HAC correction, based on an AR(1) approximation using Andrews (1991) data-dependent method and a Quadratic-Spectral kernel.

I consider power curves based on three designs for the time-varying coefficient vector, $\beta_{t,T}$ which are denoted P1 - P3. Figure 1.3 Panel D illustrates the three designs which differ in the number and location of the local departures from the null hypothesis $\beta_{t,T} = 0$ over the sample as well as the sign of the shifts. The width of the predictability pockets is fixed at 5% of the sample size i.e. each pocket has a duration of 20 observations. The magnitude of the shifts is uniform over the pockets and scaled by a scalar parameter β_A, where $\beta_A = 0$ implies no predictability at any point in time. In what follows, I briefly discuss the considered designs and the corresponding power curves in Figure 1.3.

DESIGN P1. Figure 1.3 Panel P1 shows power curves for the predictability process P1 which features two predictability pockets of opposite signs, located at one-third and three-fourths of the sample, respectively. The simulation illustrates the need for predictability tests that take instabilities into account. It is evident that the traditional LM test exhibits no power against alternatives of this form; the rejection rate of the LM test stays constant at nominal size 5%, regardless of the magnitude of the shift, β_A. In contrast, the predictability tests that allow for instabilities have power in detecting predictability of this form. Further, we note that the power of the $D\sup$ test uniformly dominates that of Rossi (2005)'s test, showing that allowing for multiple shifts in the coefficient vector under the alternative leads to power gains in finite-samples.

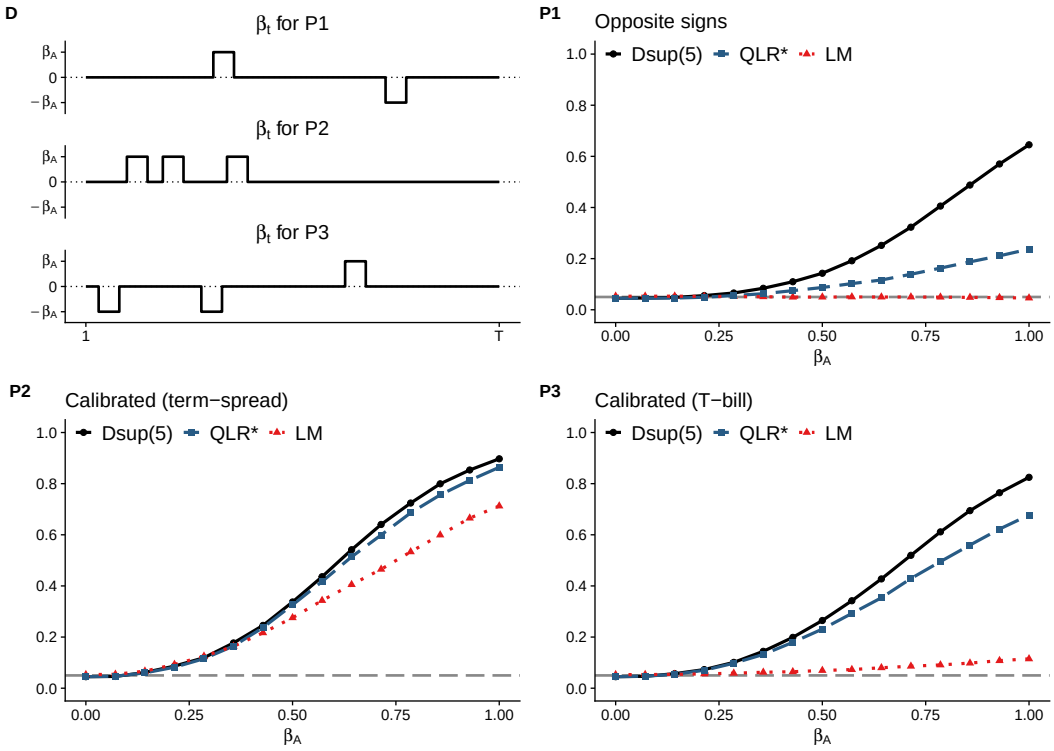
DESIGNS P2 & P3. The bottom two panels in Figure 1.3 show power comparisons for the predictability processes P2 and P3, respectively. These processes are calibrated to empirical results from Farmer et al. (2019) who conduct non-parametric regressions to study the presence of predictability pockets in various commonly considered predictors of the equity premium. When evaluating power of the proposed test in the presence of predictability pockets, there are many choices of processes for $\beta_{t,T}$ that differ with respect to number of the pockets, their location and duration as well sign and magnitude of the implied coefficient shifts. Calibrating $\beta_{t,T}$ allows to assess the test's performance under conditions that could be encountered in "real-world" empirical examples and therefore provides a good

benchmark for assessing the power of the test. I calibrate the processes P2 and P3 to the empirical findings of Farmer et al. (2019) by matching the number of simulated pockets, the relative location of the pockets over the samples as well as the sign of the implied shifts to the results reported by the authors in predicting excess returns from (i) the term spread using daily data (P2) and (ii) the T-bill rate using monthly data (P3).[20] For both processes, the power of the proposed $D\sup$ test uniformly dominates the power of the test imposing one break. This confirms that allowing multiple shifts in the coefficient vector under the alternative leads to power gains in empirically relevant scenarios. Further, as with P1, the traditional LM test shows substantially lower power for P2 and almost no power for P3 which features both negative and positive pockets, reiterating the need for predictability tests that explicitly take instabilities into account.

SENSITIVITY CHECKS. I assess the sensitivity of the conclusions drawn above to variations in the specification of the considered tests. Figure B.5 reports finite-sample power curves for the same data-generating process without serial correlation ($\phi_x = 0, \phi_\eta = 0$), but using variance estimator with a HAC correction is used. All conclusions regarding relative power of the tests discussed above remain unchanged. Finally, I inspect power in the case where both $\phi_x = 0.5$ and $\phi_\eta = 0.5$. Figure B.6 shows that the power of all of the three test decreases considerably relative to the case without serial correlation, but that the proposed $D\sup$ test remains the most powerful among the three tests considered.

[20] Specifically, Farmer et al. (2019) find 3 pockets which have "less than a 5% chance of being spurious" for predicting from the term-spread using daily data. These pockets are located at 12%, 21% and 37% of the sample with all pockets exhibiting positive coefficient shifts. For predicting using the T-bill rate in monthly data, using the same criterion, they find 3 pockets which are located at 6%, 30% and 65% of the sample with the first two pockets exhibiting negative shifts of the coefficient and the third pocket having a positive shift.

Figure 1.3: Finite-sample power for $\epsilon = 0.05$



Notes: The figure shows simulated rejection rates for tests of the null hypothesis $H_0 : \beta = 0$ against different designs for the alternative, $\beta_{t,T}$, denoted P1 - P3. Panel D illustrates the different designs for $\beta_{t,T}$. Power curves are reported for increasing size of the shifts, β_A, under the alternative. All tests are conducted at $\alpha = 5\%$ significance level. The solid black line denotes the proposed $D\sup(5)$ instability-robust test with a maximum of $\overline{K} = 5$ breaks. The blue shaded line denotes Rossi (2005)'s QLR^*_T test imposing one break and the red dotted line denotes a traditional LM test. Rejection rates are based on 5,000 replications for a sample of $T = 400$ observations from the model in equation (1.23) where the serial correlation of the predictor and prediction error $\phi_x = 0, \phi_\eta = 0$, respectively.

1.7 Local stock return predictability

Are stock returns predictable by financial valuation ratios or term-structure variables? This question is at the center of an important research agenda in finance and has been analyzed by a large array of seminal studies.[21] However, despite a significant volume of research being devoted to this question, the predictability debate has not yet reached a consensus. For example, while some studies find evidence of predictability by valuation ratios or consumption ratios (Lettau and Ludvigson, 2001), other studies find that these results are unstable and crucially depend on the stochastic properties of the predictors or the sample period studied (Campbell and Yogo, 2006). Welch and Goyal (2007) came to the conclusion that "[...] the literature has yet to find some variable that has meaningful and robust empirical equity premium forecasting power [...]" (p. 1505).

One explanation for the difficulty of establishing a consensus is that predictability can vary over time. For example, Pesaran and Timmermann (1995) find that the ability of various economic variables to predict stock returns changes with the volatility of returns and Rapach and Wohar (2006) provide evidence of parameter instability in predictive regressions. Recently, some studies have presented evidence that predictability is a local phenomenon and is concentrated in short subsamples of the data. Timmermann (2008) concludes that the "[...] empirical findings suggest that most of the time stock returns are not predictable, but there appear to be pockets in time where there is modest evidence of local predictability." Pesaran and Timmermann (2000) note that oil prices were an important predictor for stock prices during the 1970s but that their importance subsequently vanished. Similarly, Gonzalo and Pitarakis (2012, 2017) and Henkel et al. (2011) find that predictability is linked to measures of business cycle conditions, leading to short episodes of significant predictability. Further theoretical support for local predictability is given in Timmermann (2008) who argues that investors' successful search for good forecasting models itself might generate "pockets of predictability" i.e. short-lived periods of significant predictability that are followed by long periods without predictability. In a recent study, Farmer et al. (2019) argue that such pockets of return predictability are consistent with an asset pricing model featuring incomplete learning and provide ample empirical evidence in support of the predictability pockets hypothesis using non-parametric regressions.

Empirical support for or against predictability crucially relies on specification tests in predictive regressions. However, as argued throughout this paper, when predictability varies over time, traditional hypothesis tests may have low or no power against potentially important alternatives. Simply testing for predictability at each point in time or over a collection of various subsamples does not offer a good alternative as it suffers from a multiple testing problem affecting size and power of the tests. The test proposed in this

[21] For early research on this topic see for example Fama and French (1988) and Campbell and Shiller (1988). Ang and Bekaert (2007), Welch and Goyal (2007) and Timmermann (2008), Cochrane (2008) provide more recent discussions.

paper, however, is robust to multiple shifts in magnitude and signs of the parameter vector and can be applied to a general class of models. This makes it a good choice to investigate the hypothesis of local predictability in return prediction.

This paper is not the first to address the issue of robustifying inference in return prediction to episodic predictability. For example, Gonzalo and Pitarakis (2012, 2017) analyze return predictability in a threshold model that links time variation in predictability to the state of the economy (for example a variable measuring business cycle fluctuations). Henkel et al. (2011) follow a similar approach. In general, however, predictability might be linked to a variety of features of the economic environment that are not necessarily tied to the business cycle. The advantage of using the test proposed in this paper over existing approaches is that the researcher does not need to condition predictability on a set of known variables measuring the state of the economy. Rather, the test can be used as a first step to establish whether there is evidence of local predictability before the researcher comes up with a hypothesis about potential driving forces of the predictability process.

1.7.1 Data

The issue of stock return predictability has been analyzed using a large variety of specifications. To keep the exposition in this paper compact and to ensure comparability with the literature, I focus on the most commonly employed specification which predicts monthly US stock market excess returns over the post-war period 1946-2019 using the set of financial variables considered in Welch and Goyal (2007).[22] In the following, I give details on the construction of the excess return series and the predictor variables.

EQUITY PREMIUM. The dependent variable is constructed based on a CRSP (Center for Research in Security Prices) dataset that is widely used in the literature. Specifically, I construct the excess return on the US stock market (equity premium) as the difference between a measure of the US stock market log return and a risk-free log return. The US stock market return is measured by the value-weighted S&P 500 total stock market return including dividends. The risk-free rate is the three-month T-bill rate from FRED. This measure of the equity premium is widely used in the literature e.g. recently by Welch and Goyal (2007) and Kostakis et al. (2015). For robustness, I also present results using an alternative measure of the excess return where the US stock market return is the value-weighted CRSP stock market return including dividends for NYSE, AMEX and NASDAQ and the risk-free rate is proxied by a 1-month Treasury bill rate from Ibbotson and Associates Inc. This alternative measure of the equity premium has been used recently in Gonzalo and Pitarakis (2012) and Gonzalo and Pitarakis (2017). Figure B.7 Panel A shows the equity premium series used in the empirical analysis.

[22] Studies of excess return prediction often restrict the sample to the post-war period, see e.g. Gonzalo and Pitarakis (2012).

PREDICTOR VARIABLES. The source of the predictor data is an updated version of the monthly dataset used in Welch and Goyal (2007).[23] This predictor dataset has been considered in numerous studies in the literature and has become a benchmark in the predictability literature. I focus on the same set of predictors recently considered in Kostakis et al. (2015), namely the *dividend-payout ratio*, defined as the difference between the log of dividends and the log of earnings, the *earnings-price ratio*, defined as the difference between the log of earnings and the log of stock prices, the *long-term yield*, defined as the long-term US government bond yield from Ibbotson's Stocks, Bonds Bills and Inflation Yearbook, the *T-bill rate* which after 1934 is the 3-month T-bill rate from FRED and before is extracted from the NBER Macrohistory database, the *term-spread* which is the difference between the long-term yield and the T-bill rate, the *dividend-price ratio*, defined as the log of dividends over stock prices, the *dividend-yield*, defined as the log of dividends over lagged prices, the *default yield spread*, defined as the difference between the BAA and AAA-rated corporate bond yields taken from FRED, the *book-to-market ratio* which is the ratio of book value to market value for the DJIA, the *net equity expansion*, defined as the ratio of the twelve month moving sum of net equity issues by NYSE listed stocks divided by the total end-of-year market capitalization of these stocks and the *inflation rate* calculated from the Consumer Price Index of the Bureau of Labor Statistics. Figure B.7 panels B-L show the predictor series used in the empirical analysis.

1.7.2 Predictability tests robust to instabilities

As is customary in the literature, I study the individual predictive ability of each of the financial variables using the following univariate predictive model

$$r^e_{t+1} = \alpha + \beta\, x_t + \eta_{t+1} \tag{1.25}$$

where r^e_{t+1} denotes the one-month-ahead excess return and x_t is the considered predictor. Predictability studies typically conduct tests of the hypothesis $H_0 : \beta = 0$ over the full sample or specific subsamples. However, as illustrated in the previous sections, these tests are not robust to the presence of instabilities and might therefore fail to detect locally occurring predictability. The test proposed in this paper considers the same null hypothesis, but explicitly takes local predictability into account. Using the proposed test, I revisit the specification above and re-evaluate the predictability of the set of predictors described in the previous subsection. As in the simulation studies discussed in Section 1.6, I use a $D\sup(5)$ test with 5% trimming. All tests are conducted based on the HAC variance estimators reported in Section 1.2 based on an AR(1) approximation with Andrews (1991) automatic bandwidth selection procedure and a Quadratic Spectral kernel

[23] The dataset updated until December 2019 was downloaded from Amit Goyal's website and at the time of writing this paper could be found at http://www.hec.unil.ch/agoyal/.

RESULTS. Table 1.3 reports the predictability tests robust to instabilities for each of the potential predictors. To facilitate comparison with the literature, the left panel reports the full-sample least squares estimates ($\hat{\beta}_{OLS}$), the R^2 of the full-sample regression (in percentage points) as well as the traditional predictability tests using a t-ratio with HAC correction for the full-sample (t^{HAC}) and a subsample starting in 1952 (t^{HAC}_{1952}). The right panel reports the results from the instability-robust $D\sup(5)$ predictability tests proposed in Section 1.2 for the same subsamples.

I first discuss the traditional inference approach using test statistics constructed over the full-sample. Note that standard least-squares inference indicates that only few of the financial variables have significant predictive ability for the equity premium at 5% level, namely the T-bill rate, the dividend-price ratio, the dividend-yield and the inflation rate. Further, when considering a slightly different subsample starting in January 1952[24] (t^{HAC}_{1952}), the significance of the dividend-price ratio and the dividend-yield disappears and only the T-bill rate remains significant at 5% level. The results are in line with findings from traditional predictability tests reported in Kostakis et al. (2015) and, more generally, match the conclusion of previous studies documenting that predictive ability of valuation ratios crucially depends on the subsample considered (see e.g. the discussion in Campbell and Yogo (2006) or Welch and Goyal (2007)).[25]

The picture changes considerably when looking at the instability-robust tests. The $D\sup(5)$ test conducted on the 1946 - 2019 sample shows additional significant (local) predictability at 5% level for the dividend-payout ratio, the earnings-price ratio, the default yield spread and net equity expansion; only the long-term yield, the term-spread and the book-to-market ratio are not significant at 5% level. In addition, contrary to the traditional tests, the findings from the $D\sup(5)$ test are robust to moving to the post-1952 subsample.

The test results support the hypothesis that predictability from financial variables is a local phenomenon and therefore difficult to detect with traditional tests which do not take instabilities into account. The findings generalize those found in earlier studies which document evidence of episodic predictability for specific variables (e.g. Gonzalo and Pitarakis (2012) document episodic predictability for the dividend-yield.). Further, they explain why studies that split the sample at different dates have often come to conflicting conclusions regarding the predictive ability of a wide class of predictors and highlight the need for robustifying inference to local instabilities.

[24] The post-1952 subsample is often considered in predictability studies since term structure variables are thought to be more informative after the passing of the 1952 Treasury Accord which separated government debt management from monetary policy.

[25] In particular, the tests results in Gonzalo and Pitarakis (2012) also indicate vanishing excess return predictability for the dividend-yield and Kostakis et al. (2015) find the same variables to be significant in the post-1952 period when using traditional tests. Finally, all regressions have low explanatory power with R^2 below 1%, a feature documented e.g. in Welch and Goyal (2007).

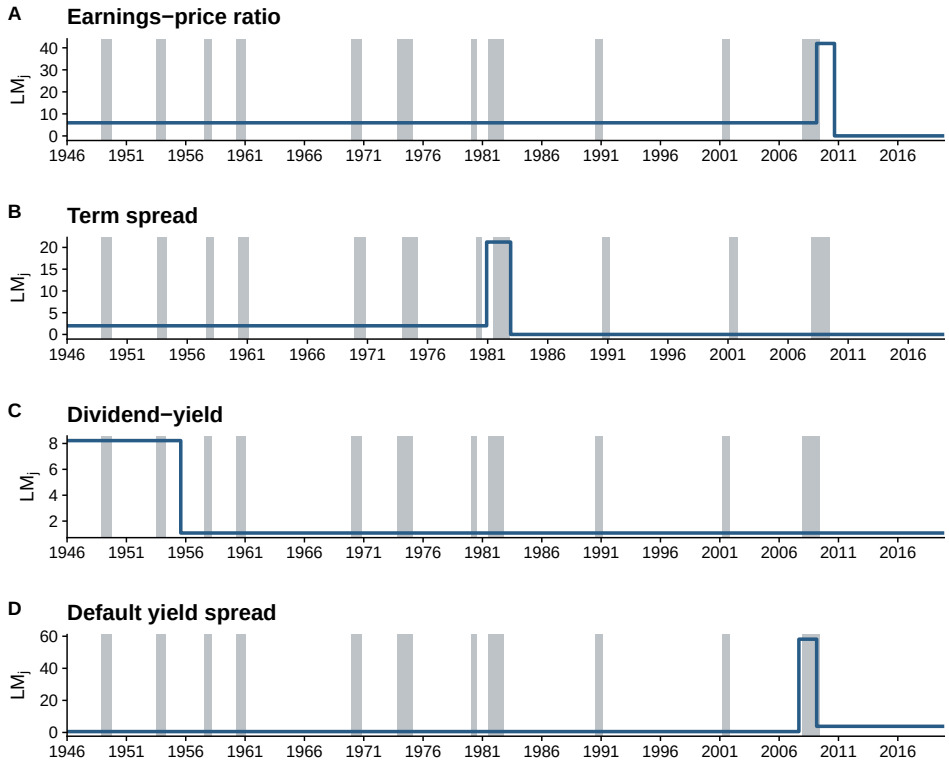
Table 1.3: Predictability tests for the Equity Premium

			Traditional		Robust	
Predictor	$\hat{\beta}_{OLS}$	R^2 (%)	t^{HAC}	t^{HAC}_{1952}	$D\sup(5)$	$D\sup(5)_{1952}$
Dividend payout ratio	0.002	0.03	0.30	0.45	13.72**	14.42**
Earnings-price ratio	0.005	0.28	1.10	0.61	13.86**	12.24**
Long-term yield	-0.086	0.34	-1.62	-1.52	8.57	8.50
T-bill rate	-0.110	0.66	-2.39**	-2.40**	12.07**	11.50*
Term spread	0.192	0.38	1.70*	1.89*	8.40	9.30
Dividend-price ratio	0.006	0.41	1.98**	1.41	16.01***	14.86**
Dividend-yield	0.006	0.46	2.08**	1.53	15.68***	14.85**
Default yield spread	0.154	0.03	0.30	0.36	13.89**	13.54**
Book-to-market ratio	0.004	0.06	0.63	0.29	6.55	5.76
Net equity expansion	-0.041	0.04	-0.37	-0.45	21.43***	21.62***
Inflation rate	-0.915	0.96	-2.58**	-1.85*	17.46***	11.03*

Notes: The table presents the results of conducting predictability tests of the null hypothesis $\beta = 0$ for the post-war sample 1946-2019 in model (1.25) using the S&P 500 Equity Premium. The left panel reports the full-sample least squares estimates, $\hat{\beta}_{OLS}$, the R^2 of the full-sample regression (in percentage points) as well as the traditional predictability tests using a t-ratio with HAC correction for the full-sample, t^{HAC}, and a subsample starting in 1952, t^{HAC}_{1952}. The right panel reports the results from the instability-robust $D\sup LM$ model-specification tests with a maximum of $K = 5$ shifts and trimming parameter set at $\epsilon = 0.05$ for the same subsamples. For all test statistics, the stars denote a rejection the null hypothesis of no predictability at significance levels 1% (***), 5% (**), and 10% (*), respectively.

ROBUSTNESS. To assess the robustness of the results documented above, I repeat the predictability tests using a different measure of the equity premium based on CRSP stock returns including dividends for NYSE, AMEX and NASDAQ with a 1-month treasury bill rate as the risk-free rate (see data section above). Table B.6 reports the test results from this alternative measure of the equity premium. All conclusions discussed in this section are robust to this change. I also address a potential concern regarding the persistence of the considered predictor variables. The predictors typically studied with predictive models such as the one in equation (1.25) are often highly persistent and recently, studies have started to model these predictors as near-unit-root predictors when conducting predictability tests (see e.g. the discussion in Pitarakis and Gonzalo (2019) and the IVX approach proposed by Kostakis et al. (2015)). While it is in principle possible to apply the proposed test to an IVX moment condition, I focus here on exploring the robustness of the results discussed above by using first-differenced values of the predictors, $\Delta x_t = x_t - x_{t-1}$. Table B.7 presents the results from the predictability tests using first-differenced predictors. In comparison to the results discussed above, the evidence of predictability indeed disappears for the earnings-price ratio, the dividend-price ratio, the dividend-yield and the inflation rate, all variables for which there is evidence of near-unit-root behavior, highlighting the importance of carefully assessing the stochastic properties of each predictor before applying predictability tests. However, for the remaining predictors the main conclusions discussed above continue to hold in the first-differenced model. Most importantly, the traditional tests still give conflicting evidence when considering the two subsamples while the proposed instability-robust tests provide stable inference.

Figure 1.4: Predictability paths



The graph shows the $\Phi_{T,j}$ components of the $D\sup LM(10)$ model-specification tests for the S&P 500 Equity Premium based on the model in equation (1.25) with trimming parameter $\epsilon = .02$. The number of coefficient shifts is selected via the BIC criterion discussed in Bai and Perron (2006). Gray vertical bars denote NBER recessions.

PREDICTABILITY PATHS. The evidence discussed above that predictability from financial variables is a local phenomenon, raises an interesting question: Is there heterogeneity in the location of predictability for different predictors? And if yes, how does predictability evolve over the sample? Is predictability really concentrated in short "predictability pockets" such as hypothesized by recent studies or are there larger episodes of predictability? While the tests proposed in this paper do not allow to draw precise inference on which periods over the sample are significant or not at a given significance level,[26] the components of the $D\sup(K)$ test statistic do provide a narrative view on how predictability might evolve over the sample. Figure 1.4 shows the evolution of the $\Phi_{T,j}$ components of the $D\sup(K)$ test over the sample. In contrast to the previous section, I consider a larger upper ceiling of shifts, $\bar{K} = 10$ and a lower trimming parameter, $\epsilon = .02$. I adopt this specification of the tests since Farmer et al. (2019) who assess predictability paths using a

[26] Since the proposed tests are joint hypothesis tests, it is not straightforward to extend the methods to a sequential procedure in the spirit of Bai and Perron (1998)'s $F_T(k+1|k)$ test as the researcher would need to add an intermediate step that tests whether the rejection occurred due to constant predictability or the presence of an additional shift in the coefficient vector. However, developing repeated testing procedures that correct for the multiple testing at each stage is an interesting avenue for future research.

non-parametric testing procedure based on multiple t-tests report evidence of particularly short-lived predictability episodes. Simulation studies available on request show that the tests still exhibits good size control and has the same power properties against the data-generating processes discussed in Section 1.6 when using a sample of the size available here. To choose the number of shifts, I adopt the BIC criterion discussed in Bai and Perron (2006) and only report predictability paths for the variables for which the criterion detects at least one shift. The figure provides evidence that predictability is indeed concentrated in different periods over the sample, depending on the predictor considered. Specifically, while the predictive ability of the dividend-yield seems concentrated in the pre-1956 period, the earnings-price ratio and the default yield spread predictors to have an episode of large predictive ability during the Great Recession period. Finally, the predictive ability of the term spread variable seems concentrated in a brief period during the early 1980s.

1.8 Conclusion

This paper develops a general approach to test whether a parameter should be included in an economic model robust to time-variation in parameters. The hypothesis test can be used to evaluate any economic model described by a set of moment conditions in-sample or out-of-sample. In-sample, the test selects between two nested model specifications in the presence of parameter instabilities. Out-of-sample, the test can be used to evaluate the performance of model forecasts or model-free forecasts such as survey or judgmental forecasts robust to time-variation. The key feature of the proposed test is that it is particularly powerful in the presence of multiple breaks in parameters without imposing a specific form of time-variation. Further, the test statistic provides narrative evidence on which parts of the sample drive the rejection of the null hypothesis.

The approach jointly tests for both parameter instability and a constant non-zero value of the parameter. This allows the test to detect departures from the null hypothesis, even when they only occur over short periods of the sample and makes the test more powerful than traditional hypothesis tests which are based on the full sample. The test statistic jointly considers all possible partitions of the sample up to an upper bound of $\bar{K}$ splits to evaluate whether there is evidence to reject the null hypothesis. It can be constructed based on a Lagrange-Multiplier or a Wald form and can be efficiently implemented via a dynamic programming algorithm provided in the paper.

Extensive Monte-Carlo simulations show that the proposed test is accurately sized in finite samples and is more powerful than tests assuming constant coefficients or a single break if the data-generating process exhibits multiple breaks in parameters. At the same time, the test has high power when model parameters only undergo one shift or are constant. This makes the test particularly useful when the researcher faces uncertainty about whether and how parameters change over time.

The empirical study uses the test to document the presence of local short-horizon predictability in the US equity premium during the 1946-2019 period from a set of financial variables considered in Welch and Goyal (2007). There is significant predictive ability with respect to one-month ahead excess market returns for a large set of predictors, once time-variation is taken into account. In contrast to traditional predictability tests based on the full sample, the conclusions from the proposed test are invariant to changes in the considered sample. Furthermore, the test provides evidence of heterogeneity in the location of predictability episodes across variables. The findings explain why traditional tests often fail to uncover predictability in the full sample and why studies that split the sample at different dates often arrive at conflicting results regarding the predictive ability of a wide class of variables.

Acknowledgements

I am grateful for the support, advice and encouragement I have received from Barbara Rossi. I would also like to thank Geert Mesters for his support and valuable comments. Further, I would like to thank Heather Anderson, Christian Brownlees, Bjarni G. Einarsson, Kirill Evdokimov, Katharina Janezic, Adam Lee, Katerina Petrova, Francesco Ravazzolo, Tatevik Sekhposyan, André Souza, Farshid Vahid-Araghi, as well as participants at the 40th International Symposium on Forecasting, the 8th SIdE Workshop for PhD students in Econometrics and Empirical Economics, the 7th Barcelona GSE PhD Jamboree and the Econometrics seminars at Universitat Pompeu Fabra for their feedback and helpful comments. I gratefully acknowledge financial support from the 7th Economics Job Market Best Paper Award (UniCredit Foundation) and the Best Student Presentation Award of the International Institute of Forecasters sponsored by Amazon.

Appendices

A Critical values & implementation

This section provides additional details on the implementation of the tests. The first section reports the critical values for the tests and describes the simulation procedure used to obtain them. The second section describes a software package accompanying the paper available from the author.

A.1 Asymptotic critical values

Critical values of the test statistics can be obtained by directly simulating the limiting distributions listed in Theorem 1.3.1 using a dynamic programming algorithm.[27] The table below reports critical values for the predictability tests discussed in the paper. The significance levels considered in the tables are 10%, 5%, 2.5% and 1%. The critical values were obtained by simulating the asymptotic distributions based on 10,000 Monte Carlo replications and an approximation length of $N = 3,600$ for the Brownian Motions.[28]

A.2 Software implementation

Accompanying the paper, the author makes available a software package that can be used to conveniently use the tests for applied work and for simulating the critical values reported in Table A.4. The code is mainly written in C++ using the Armadillo C++ linear algebra library (Sanderson and Curtin, 2016) and offers an R interface provided in the form of an R package using the Rcpp library (Eddelbuettel and François, 2011). The routines carrying out the dynamic programming algorithm have been parallelized using the OpenMP application programming interface (Dagum and Menon, 1998). The software package can be obtained from the author on request.

[27] Alternatively, one could obtain the critical values from an approximation strategy for functions of Brownian motions discussed in Bai (1999).

[28] Simulations of critical values were carried out on the Amazon Web Services Elastic Compute Cloud (AWS EC2) using a *c5.xlarge* instance type (4 vCPUs, 8 GB memory) running Amazon Linux 2.

Table A.4: Asymptotic critical values for $\sup \Phi_T(K)$ and $D \sup \Phi_T(\bar{K})$ tests

ϵ	p	α	$D\sup(5)$	$\sup(1)$	$\sup(2)$	$\sup(3)$	$\sup(4)$	$\sup(5)$
0.05	1	0.100	10.115	9.644	17.542	23.162	28.471	33.017
0.05	1	0.050	11.566	11.293	19.593	25.721	30.974	35.638
0.05	1	0.025	13.180	13.048	21.542	27.837	33.124	38.112
0.05	1	0.010	15.196	15.155	24.182	30.330	36.257	41.303
0.05	2	0.100	14.008	13.810	23.920	31.926	39.282	46.025
0.05	2	0.050	15.789	15.647	26.053	34.516	42.134	49.096
0.05	2	0.025	17.468	17.361	28.146	36.901	44.720	52.020
0.05	2	0.010	19.953	19.915	30.832	39.694	47.824	55.510
0.10	1	0.100	9.386	9.131	15.605	20.002	23.664	26.368
0.10	1	0.050	10.985	10.813	17.555	22.215	26.101	28.879
0.10	1	0.025	12.555	12.462	19.423	24.315	28.143	31.150
0.10	1	0.010	14.528	14.528	21.754	26.888	31.063	34.225
0.10	2	0.100	13.301	13.185	21.621	28.170	33.857	38.278
0.10	2	0.050	15.054	15.007	24.040	30.691	36.529	41.426
0.10	2	0.025	16.703	16.690	25.863	33.087	39.379	44.537
0.10	2	0.010	19.189	19.162	28.168	36.201	42.386	47.949
0.15	1	0.100	8.833	8.670	14.007	17.367	19.426	18.791
0.15	1	0.050	10.474	10.335	16.057	19.726	21.857	21.349
0.15	1	0.025	12.113	12.075	18.036	21.777	23.950	23.486
0.15	1	0.010	14.163	14.163	20.391	24.511	27.025	26.211
0.15	2	0.100	12.800	12.734	19.997	25.229	28.851	29.077
0.15	2	0.050	14.651	14.627	22.291	27.870	31.651	31.824
0.15	2	0.025	16.262	16.230	24.231	30.194	34.195	34.773
0.15	2	0.010	18.613	18.581	26.794	32.879	37.592	37.952

Notes: This table reports simulated quantiles of the limiting distributions of the $\sup \Phi_T(K)$ and $D \sup \Phi_T(\bar{K})$ tests. The critical values were obtained based on 10,000 Monte-Carlo replications and an approximation length of $N = 3,600$ observations for the partial sums to simulate the Brownian Motions.

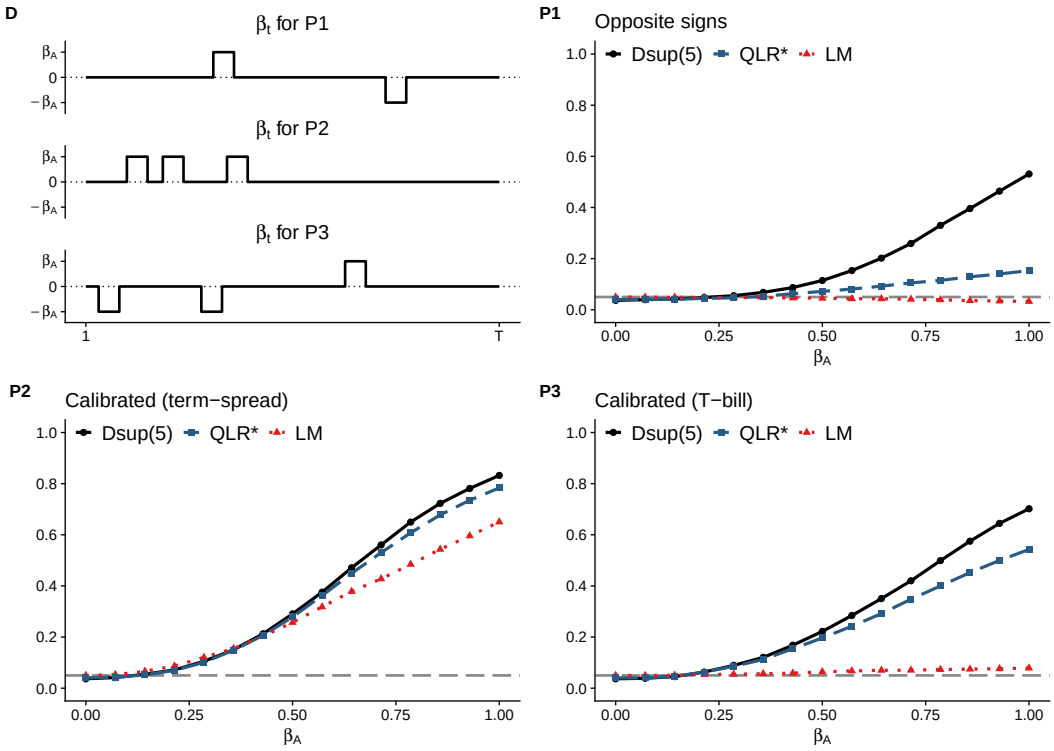
B Additional tables & figures

Table B.5: Finite-sample size for $\epsilon = 0.1$ (nominal size 5%)

		$\phi_\eta = 0$			$\phi_\eta = 0.25$			$\phi_\eta = 0.5$		
T	ϕ_x	Dsup(5)	LM	QLR^*	Dsup(5)	LM	QLR^*	Dsup(5)	LM	QLR^*
Panel A: Heteroskedasticity-robust										
125	0.00	0.032	0.052	0.034	0.032	0.051	0.035	0.036	0.051	0.035
250	0.00	0.037	0.046	0.039	0.040	0.047	0.038	0.040	0.047	0.040
500	0.00	0.042	0.049	0.042	0.042	0.049	0.040	0.042	0.048	0.044
1,000	0.00	0.045	0.050	0.045	0.045	0.050	0.045	0.046	0.052	0.048
125	0.25	0.031	0.051	0.034	0.054	0.065	0.054	0.086	0.081	0.084
250	0.25	0.038	0.045	0.041	0.066	0.061	0.065	0.105	0.078	0.104
500	0.25	0.041	0.051	0.042	0.071	0.066	0.071	0.114	0.079	0.113
1,000	0.25	0.046	0.051	0.046	0.078	0.064	0.076	0.122	0.081	0.118
125	0.50	0.036	0.051	0.037	0.085	0.080	0.084	0.176	0.118	0.172
250	0.50	0.043	0.048	0.041	0.105	0.080	0.102	0.227	0.121	0.210
500	0.50	0.043	0.048	0.043	0.110	0.080	0.111	0.244	0.119	0.230
1,000	0.50	0.046	0.053	0.047	0.120	0.085	0.117	0.260	0.125	0.240
Panel B: HAC - AR(1) approximation, QS kernel, Andrews (1991) bandwidth										
125	0.00	0.024	0.047	0.025	0.022	0.048	0.025	0.024	0.046	0.025
250	0.00	0.033	0.044	0.032	0.033	0.045	0.032	0.032	0.044	0.032
500	0.00	0.040	0.049	0.041	0.040	0.049	0.040	0.039	0.048	0.041
1,000	0.00	0.044	0.049	0.043	0.044	0.050	0.045	0.042	0.051	0.046
125	0.25	0.023	0.046	0.024	0.026	0.052	0.028	0.027	0.053	0.028
250	0.25	0.034	0.044	0.034	0.037	0.050	0.039	0.038	0.052	0.039
500	0.25	0.039	0.049	0.040	0.044	0.054	0.046	0.045	0.054	0.045
1,000	0.25	0.044	0.050	0.044	0.051	0.053	0.050	0.051	0.055	0.052
125	0.50	0.024	0.047	0.028	0.028	0.053	0.031	0.025	0.053	0.027
250	0.50	0.037	0.045	0.037	0.040	0.052	0.044	0.037	0.053	0.039
500	0.50	0.039	0.048	0.039	0.045	0.054	0.047	0.043	0.052	0.045
1,000	0.50	0.045	0.052	0.044	0.051	0.056	0.053	0.050	0.054	0.052

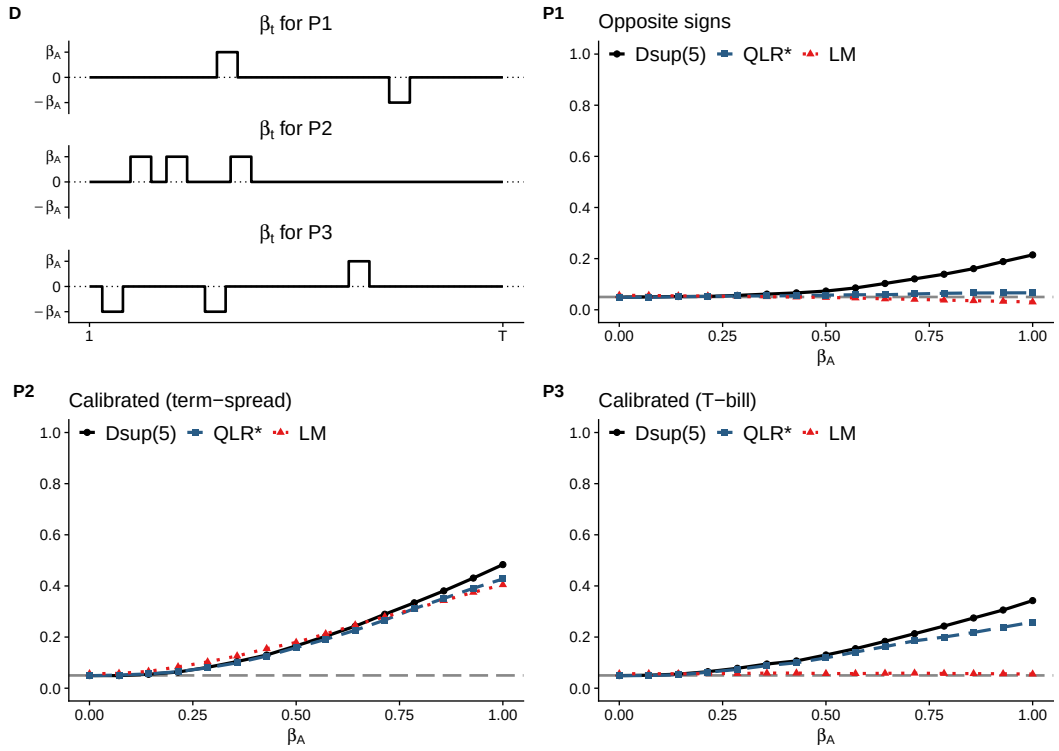
Notes: The table reports simulated finite-sample size for tests of the null hypothesis $H_0 : \beta = 0$ with 5% nominal size. $D\sup(5)$ denotes the proposed instability-robust test with a maximum of $\bar{K} = 5$ breaks, LM denotes a traditional LM test and QLR^* denotes Rossi (2005)'s instability-robust test imposing one break. Rejection rates are based on 10,000 replications from the model in equation (1.21) using a sample of T observations where the serial correlation of the predictor and prediction error is controlled by ϕ_x, ϕ_η, respectively.

Figure B.5: Finite-sample power for $\epsilon = 0.05$ (HAC correction)



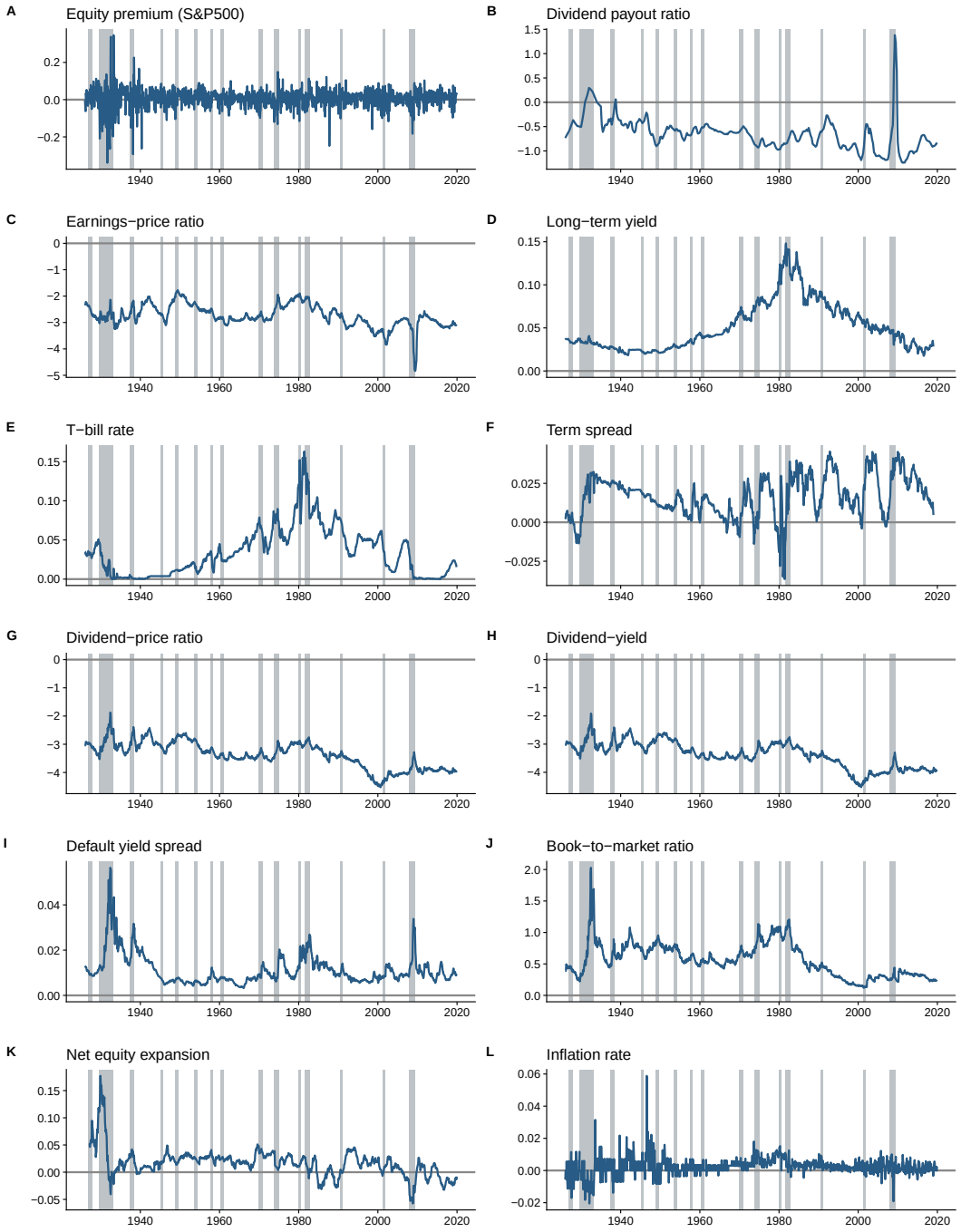
Notes: The figure shows simulated rejection rates for tests of the null hypothesis $H_0 : \beta = 0$ against different designs for the alternative, $\beta_{t,T}$, denoted P1 - P3. Panel D illustrates the different designs for $\beta_{t,T}$. Power curves are reported for increasing size of the shifts, β_A, under the alternative. All tests are conducted at $\alpha = 5\%$ significance level. The solid black line denotes the proposed $D\sup(5)$ instability-robust test with a maximum of $\overline{K} = 5$ breaks. The blue shaded line denotes Rossi (2005)'s QLR^*_T test imposing one break and the red dotted line denotes a traditional LM test. Rejection rates are based on 5,000 replications for a sample of $T = 400$ observations from the model in equation (1.23) where the serial correlation of the predictor and prediction error $\phi_x = 0, \phi_\eta = 0$, respectively.

Figure B.6: Finite-sample power for $\epsilon = 0.05$ (HAC correction, Serial correlation)



Notes: The figure shows simulated rejection rates for tests of the null hypothesis $H_0 : \beta = 0$ against different designs for the alternative, $\beta_{t,T}$, denoted P1 - P3. Panel D illustrates the different designs for $\beta_{t,T}$. Power curves are reported for increasing size of the shifts, β_A, under the alternative. All tests are conducted at $\alpha = 5\%$ significance level. The solid black line denotes the proposed $D\sup(5)$ instability-robust test with a maximum of $\bar{K} = 5$ breaks. The blue shaded line denotes Rossi (2005)'s QLR^*_T test imposing one break and the red dotted line denotes a traditional LM test. Rejection rates are based on 5,000 replications for a sample of $T = 400$ observations from the model in equation (1.23) where the serial correlation of the predictor and prediction error $\phi_x = 0.5, \phi_\eta = 0.5$, respectively.

Figure B.7: Data used in Equity Premium prediction



Notes: The figure shows the equity premium series y_{t+1} in panel A and the raw predictor data used in the empirical application, x_t in panels B to L. Shaded gray bars denote recessions as measured by the NBER indicator.

Table B.6: Predictability tests for the Equity Premium

			Traditional		Robust	
Predictor	$\hat{\beta}_{OLS}$	R^2 (%)	t^{HAC}	t^{HAC}_{1952}	$D\sup(5)$	$D\sup(5)_{1952}$
Dividend payout ratio	0.003	0.05	0.43	0.58	14.58**	15.47***
Earnings-price ratio	0.005	0.24	1.02	0.57	11.80**	10.71*
Long-term yield	-0.086	0.32	-1.57	-1.48	7.25	7.18
T-bill rate	-0.112	0.65	-2.22**	-2.23**	11.29*	10.66*
Term spread	0.205	0.41	1.70*	1.87*	8.09	8.71
Dividend-price ratio	0.006	0.42	1.90*	1.42	13.83**	12.62**
Dividend-yield	0.007	0.49	2.06**	1.61	13.60**	12.58**
Default yield spread	0.271	0.07	0.49	0.54	12.71**	12.40**
Book-to-market ratio	0.005	0.08	0.74	0.43	5.66	5.72
Net equity expansion	-0.038	0.03	-0.32	-0.39	18.86***	19.12***
Inflation rate	-0.940	0.97	-2.67***	-1.89*	17.37***	10.86*

Notes: The table presents the results of conducting predictability tests of the null hypothesis $\beta = 0$ for the post-war sample 1946-2019 in model (1.25) using the CRSP Equity Premium. The left panel reports the full-sample least squares estimates, $\hat{\beta}_{OLS}$, the R^2 of the full-sample regression (in percentage points) as well as the traditional predictability tests using a t-ratio with HAC correction for the full-sample, t^{HAC}, and a subsample starting in 1952, t^{HAC}_{1952}. The right panel reports the results from the instability-robust $D\sup LM$ model-specification tests with a maximum of $\bar{K} = 5$ shifts and trimming parameter set at $\epsilon = 0.05$ for the same subsamples. For all test statistics, the stars denote a rejection the null hypothesis of no predictability at significance levels 1% (***), 5% (**), and 10% (*), respectively.

Table B.7: Predictability tests for the Equity Premium (first differences)

			Traditional		Robust	
Predictor	$\hat{\beta}_{OLS}$	R^2 (%)	t^{HAC}	t^{HAC}_{1952}	$D\sup(5)$	$D\sup(5)_{1952}$
Dividend payout ratio	-0.061	0.45	-2.14**	-2.29**	20.21***	20.00***
Earnings-price ratio	0.018	0.07	0.88	0.96	4.87	4.72
Long-term yield	-1.590	1.04	-3.02***	-2.92***	11.89**	12.86**
T-bill rate	-1.102	1.06	-3.30***	-3.26***	19.37***	17.81***
Term spread	0.350	0.11	0.96	1.00	11.37*	10.42*
Dividend-price ratio	-0.037	0.14	-0.91	-1.04	9.11	9.65
Dividend-yield	0.027	0.07	0.81	0.72	3.65	3.82
Default yield spread	0.261	0.00	0.15	0.19	13.50**	13.32**
Book-to-market ratio	-0.069	0.20	-1.33	-1.64	7.69	7.90
Net equity expansion	-0.674	0.35	-1.52	-1.68*	11.25*	10.82*
Inflation rate	-0.265	0.08	-1.22	-1.98**	3.14	4.87

Notes: The table presents the results of conducting predictability tests of the null hypothesis $\beta = 0$ for the post-war sample 1946-2019 in model (1.25) using the S&P 500 Equity Premium where all variables are transformed to first-differences. The left panel reports the full-sample least squares estimates, $\hat{\beta}_{OLS}$, the R^2 of the full-sample regression (in percentage points) as well as the traditional predictability tests using a t-ratio with HAC correction for the full-sample, t^{HAC}, and a subsample starting in 1952, t^{HAC}_{1952}. The right panel reports the results from the instability-robust $D\sup LM$ model-specification tests with a maximum of $\bar{K} = 5$ shifts and trimming parameter set at $\epsilon = 0.05$ for the same subsamples. For all test statistics, the stars denote a rejection the null hypothesis of no predictability at significance levels 1% (***), 5% (**), and 10% (*), respectively.

C Mathematical derivations

NOTATION. Before presenting the derivations, recall some notational conventions that are used throughout the rest of the appendix. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space on which all of the random elements are defined. Unless specified otherwise, all limits are taken as the sample size $T \to \infty$. The symbol $\xrightarrow{p}$ denotes convergence in probability and $\xrightarrow{d}$ denotes convergence in distribution. Next, $\Rightarrow$ denotes weak convergence for sequences of measurable random elements of a space of bounded Euclidean-valued cadlag functions on the product space $D[0,1]^T$ as defined in Phillips and Durlauf (1986) where each component space $D[0,1]$ is equipped with the Skorohod metric. $\|\cdot\|$ denotes the Euclidean norm of a vector of matrix and $[\cdot]$ is the integer part operator. For notational simplicity, I say that $x_t(r,s) = o_{p,rs}(1)$ if it holds that $\sup_{r,s\in[0,1],s>r\rho} \|x_t(r,s)\| = o_p(1)$.

C.1 In-sample inference

THEOREM 1.3.1 (Limiting distribution for in-sample tests): *Assume that the regularity conditions in Assumption 1.3.1 hold. Under the null hypothesis defined in (1.3), it holds that*

$$\begin{aligned}
\sup \Phi_T(K) \;&\Rightarrow\; \sup_{\lambda^K \in \Lambda_\epsilon} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j) - \mathcal{B}_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}} \right\} \\
D \sup \Phi_T(\bar{K}) &\Rightarrow \max_{1\le k\le \bar{K}} (1/k) \sup_{\lambda^K \in \Lambda_\epsilon} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j) - \mathcal{B}_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}} \right\} \\
\Lambda_\epsilon &\equiv \Big\{ \lambda_j : \;\; \lambda_j \in (\epsilon, 1-\epsilon),\; \lambda_j > \lambda_{j-1} + \epsilon,\;\; j = 1, \ldots, K \Big\}
\end{aligned}$$

where $\lambda_0 \equiv 0, \lambda_{K+1} \equiv 1$ *and* $\mathcal{B}_p(\cdot)$ *is a* $(p \times 1)$ *vector of independent standard Brownian motions on* $[0,1]$.

Proof. Note that under the null hypothesis in (1.3), it holds that $\theta_t = \theta_0 = (0_{p\times 1}, \delta)\ \forall\ t$. Therefore, under the null hypothesis, the function defining the moment condition $f(z_t, \theta_t) = f(z_t, \beta_t, \delta)$ can be written as a function of a constant parameter $f(z_t, \theta)$. This notation will be used in the subsequent derivations.

To prove the weak convergence results stated in the theorem, I start by showing that the partial sample moments satisfy the following invariance principle under the null hypothesis.

$$T^{-1/2}\, W_T^{1/2} \sum_{t=1}^{[sT]} f(z_t, \theta_0) \;\Rightarrow\; \mathcal{B}_m(s) \qquad s \in (0,1]$$

Recall the following result from Corollary 2.2 of Phillips and Durlauf (1986) which generalizes the stationary version of the univariate invariance principle by McLeish (1975).

> Corollary 2.2, Phillips and Durlauf (1986): *Let* $\{u_t\}_{t=1}^{\infty}$ *be a weakly stationary sequence of random* $n \times 1$ *vectors satisfying* $\mathbb{E}[u_t] = 0 \ \ \forall\ t$. *If (a)* $\mathbb{E}|u_{i1}|^{\beta} < \infty$ $(i = 1, ..., n)$ *for some* $2 \leq \beta < \infty$ *and (b) either* $\sum_{n=1}^{\infty} \ \varphi_n^{1-1/\beta} < \infty$ *or,* $\beta > 2$ *and* $\sum_{n=1}^{\infty} \alpha_n^{1-2/\beta} < \infty$, *then*
>
> $$\Sigma = \lim_{T\to\infty} \mathbb{E}\left[T^{-1} S_T S_T'\right] = \mathbb{E}[u_1 u_1'] + \sum_{k=2}^{\infty} \left\{\mathbb{E}[u_1 u_k'] + \mathbb{E}[u_k u_1']\right\}$$
>
> *where* $S_t = \sum_{j=1}^{[Tt]} u_j$. *If* Σ *is positive definite, then* $X_T(t) = \frac{1}{\sqrt{T}} \Sigma^{-1/2} S_{[Tt]} \Rightarrow W(t)$ *as* $T \to \infty$.

Choose $u_t := f(z_t, \theta_0)$ and verify the conditions of the Corollary. The first requirement and condition (a) follow from Assumption 1.3.1.(ii) and Assumption 1.3.1.(iv). Condition (b) follows from 1.3.1.(i). The last requirements follows from Assumption 1.3.1.(iii). Applying the Corollary, using Assumption 1.3.1.(vii) and Slutsky's Theorem, it follows that

$$T^{-1/2}\ W_T^{1/2} \sum_{t=1}^{[sT]} f(z_t, \theta_0) \ \Rightarrow\ \mathcal{B}_m(s) \tag{26}$$

LAGRANGE-MULTIPLIER FORM, Φ_T^{LM}

The Lagrange-Multiplier form builds on the restricted GMM estimator defined in (1.7). I start by proving that this estimator is consistent under the null hypothesis in (1.3) i.e. that $\tilde{\theta} \xrightarrow{p} \theta_0$. Recall from equation (1.7) the definition of $\tilde{\theta}$.

$$\begin{aligned} \tilde{\theta} &:= \arg\max_{\theta\in\Theta} \ \hat{Q}_T(\theta) \qquad \text{subject to } \ A\tilde{\theta} = 0 \\ \hat{Q}_T(\theta) &:= \hat{F}_T(\theta)'\ W_T\ \hat{F}_T(\theta) \\ \hat{F}_T(\theta) &\equiv \frac{1}{T}\sum_{t=1}^{T} f(z_t, \theta) \end{aligned}$$

where $T_0 = 1$ and $A \equiv \begin{bmatrix} I_{p\times p} & 0_{p\times q} \end{bmatrix}$.

To prove consistency of $\tilde{\theta}$, I first show consistency of the unrestricted estimator $\hat{\theta}$ which is defined as the estimator above, but ignores the constraint $A\tilde{\theta} = 0$. Define the limiting objective function $Q_0(\theta) \equiv \mathbb{E}[F_T(\theta)]'\ \Sigma_{ff}^{-1}\ \mathbb{E}[F_T(\theta)]$ and apply Theorem 2.1 of Newey and McFadden (1994) to show $\hat{\theta} \xrightarrow{p} \theta_0$. The theorem requires that (i) $Q_0(\theta)$ is uniquely

maximized at θ_0; (ii) Θ is compact; (iii) $Q_0(\theta)$ is continuous and (iv) $\hat{Q}_T(\theta)$ converges uniformly in probability to $Q_0(\theta)$. Requirement (i) is satisfied by the identification assumption in 1.3.1.(vi) and positive definiteness of Σ_{ff} in 1.3.1.(iii). Requirement (ii) is satisfied by 1.3.1.(v). Requirement (iii) is satisfied by 1.3.1.(iv). The uniform convergence requirement in (iv) follows from verifying Assumptions A1, B1, and A5 in Andrews (1987) and applying the main theorem. Assumption A1 of Andrews (1987) follows from Assumption 1.3.1.(v), Assumption B1 follows from Assumption 1.3.1.(i) and Assumption A5 follows from 1.3.1.(v) and 1.3.1.(iv). Having shown that all requirements are satisfied, we apply Theorem 2.1 of Newey and McFadden (1994) and get $\hat{\theta} \xrightarrow{p} \theta_0$. Consistency of the restricted estimator $\tilde{\theta} \to \theta_0$ then follows from the argument of Theorem 9.1 of Newey and McFadden (1994).

Next, I derive a preliminary asymptotic result characterizing the limiting distribution of the normalized partial sample moment for any block of the sample with $t = [rT]+1, \ldots, [sT]$, $r, s \in [0,1]$ and $s > r$. Start again from the constrained GMM estimator $\tilde{\theta}$ defined in equation (1.7). Define the following Lagrangian for $\tilde{\theta}$:

$$\begin{aligned} \tilde{\theta} &= \arg\max_{\theta\in\Theta} \mathcal{L}_T(\theta,\mu) & \mathcal{L}_T(\theta,\mu) &= \frac{1}{2}\, F_T(\theta)' W_T F_T(\theta) + a(\theta)'\, \mu_T \\ a(\theta) &:= A\,\theta = \beta & A &= \begin{bmatrix} I_p & 0_{p\times q} \end{bmatrix} \end{aligned}$$

where μ_T is a $(p \times 1)$ vector of Lagrangian multipliers which will be non-zero if the constraints are binding. The first-order conditions of this optimization problem are

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{T}\, \nabla_\theta F_T(\tilde{\theta})' W_T F_T(\tilde{\theta}) - \nabla_\theta a(\tilde{\theta})' \sqrt{T}\tilde{\mu}_T \\ a(\tilde{\theta}) \end{bmatrix} \tag{27}$$

An element-by-element mean value expansion of $f(z_t,\theta)$ around θ_0, evaluated at $\tilde{\theta}$ yields

$$f(z_t,\tilde{\theta}) = f(z_t,\theta_0) + \frac{\partial f(z_t,\bar{\theta})}{\partial\theta}(\tilde{\theta}-\theta_0)$$

where $\bar{\theta} = [\bar{\theta}^{(1)},\ldots,\bar{\theta}^{(v)}]'$ and $\bar{\theta}^{(i)} = \alpha^{(i)}\tilde{\theta}^{(i)} + (1-\alpha^{(i)})\theta_0^{(i)}$ for some $\alpha^{(i)} \in [0,1]$ and each $t = 1,\ldots,T$ and $i = 1,\ldots,k$. Summing these terms from 1 to T, dividing by T and pre-multiplying by $\sqrt{T}$ gives

$$\sqrt{T}\; F_T(\tilde{\theta}) = \sqrt{T} F_T(\theta_0) + \nabla_\theta F_T(\bar{\theta})\; \sqrt{T}(\tilde{\theta}-\theta_0) \tag{28}$$

A similar mean-value expansion of $a(\theta)$ about θ_0, evaluated at $\tilde{\theta}$ gives

$$\sqrt{T}\; a(\tilde{\theta}) = \sqrt{T} a(\theta_0) + A\; \sqrt{T}(\tilde{\theta}-\theta_0) \tag{29}$$

Substituting the expansions into the first order conditions and rearranging yields

$$\begin{bmatrix} -\sqrt{T}\,\nabla_\theta F_T(\bar{\theta})' W_T F_T(\theta_0) \\ -\sqrt{T} a(\theta_0) \end{bmatrix} = \begin{bmatrix} \nabla_\theta F_T(\tilde{\theta})' W_T \nabla_\theta F_T(\bar{\theta}) & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} \sqrt{T}(\tilde{\theta} - \theta_0) \\ \sqrt{T}\tilde{\mu}_T \end{bmatrix} \tag{30}$$

Using the consistency result proved above that $\tilde{\theta} \xrightarrow{p} \theta_0$ and uniform convergence of $\nabla_\theta F_T(\theta)$ following from the assumptions of the theorem, we get

$$\begin{bmatrix} -M'\Sigma^{-1/2}\,\Sigma_{ff}^{-1/2}\,\sqrt{T}F_T(\theta_0) \\ -\sqrt{T}a(\theta_0) \end{bmatrix} = \begin{bmatrix} D & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} \sqrt{T}(\tilde{\theta} - \theta_0) \\ \sqrt{T}\tilde{\mu}_T \end{bmatrix} + o_p \tag{31}$$

where $D \equiv M'\Sigma^{-1}M = \bar{M}'\bar{M}$.

From the formula for inverses of block matrices, we have

$$\begin{bmatrix} D & A' \\ A & 0 \end{bmatrix}^{-1} = \begin{bmatrix} D^{-1/2}(I-P)D^{-1/2} & D^{-1}A'(AD^{-1}A')^{-1} \\ (AD^{-1}A')^{-1}AD^{-1} & -(AD^{-1}A')^{-1} \end{bmatrix} \tag{32}$$

where $P \equiv D^{-1/2}A'(AD^{-1}A')^{-1}AD^{-1/2}$ is an $m \times m$ idempotent matrix of rank q. Solving for $\sqrt{T}(\tilde{\theta} - \theta_0)$ using the formula for the block-inverse and re-arranging, we get

$$\begin{aligned} \sqrt{T}(\tilde{\theta} - \theta_0) = &- D^{-1/2}(I-P)D^{-1/2}\bar{M}'\,\Sigma_{ff}^{-1/2}\,\sqrt{T}F_T(\theta_0) \\ &- D^{-1}A'(AD^{-1}A')^{-1}\sqrt{T}a(\theta_0) + o_p \end{aligned} \tag{33}$$

Next, we calculate an alternative form for $f(z_t, \tilde{\theta})$. An element-by-element mean-value expansion of $f(z_t, \tilde{\theta})$ around θ_0 gives

$$\sqrt{T}f(z_t, \tilde{\theta}) = \sqrt{T}f(z_t, \theta_0) + \nabla_\theta f(z_t, \bar{\theta})\,\sqrt{T}(\tilde{\theta} - \theta_0)$$

where again $\bar{\theta} = [\bar{\theta}^{(1)}, \ldots, \bar{\theta}^{(v)}]'$ and $\bar{\theta}^{(i)} = \alpha^{(i)}\tilde{\theta}^{(i)} + (1-\alpha^{(i)})\theta_0^{(i)}$ for some $\alpha^{(i)} \in [0,1]$ and each $t = 1, \ldots, T$ and $i = 1, \ldots, k$.

Take any $r, s \in [0,1]$ with $s > r$. Summing the expansion above between $[rT]+1$ and $[sT]$, multiplying by $\frac{1}{\sqrt{T}}W_T^{1/2}$ and using the expression for $\sqrt{T}(\tilde{\theta} - \theta_0)$ (33), we get

$$\begin{aligned}\frac{1}{\sqrt{T}}\, W_T^{1/2} \sum_{t=[rT]+1}^{[sT]} f(z_t,\tilde{\theta}) = \frac{1}{\sqrt{T}}\, W_T^{1/2} \sum_{t=[rT]+1}^{[sT]} f(z_t,\theta_0) \\ - W_T^{1/2}\,(1/T) \sum_{t=[rT]+1}^{[sT]} \nabla_\theta f_t(\bar{\theta}) D^{-1/2}(I-P)D^{-1/2}\bar{M}' \\ \times \sqrt{T}\,\Sigma_{ff}^{-1/2}\, F_T(\theta_0) \\ - W_T^{1/2}\,(1/T) \sum_{t=[rT]+1}^{[sT]} \nabla_\theta f_t(\bar{\theta}) D^{-1} A'(AD^{-1}A')^{-1}\sqrt{T}a(\theta_0) \\ + o_p \end{aligned} \tag{34}$$

where under the null hypothesis $a(\theta_0) = 0_{p\times 1}$ so that the third term disappears.

To derive the limiting distribution, we inspect the convergence of each component of the sum above. First, note that since $\tilde{\theta} \xrightarrow{p} \theta_0$, it follows that $\bar{\theta} \xrightarrow{p} \theta_0$ and under the assumptions of the theorem we have that

$$(1/T) \sum_{t=[rT]+1}^{[sT]} \nabla_\theta f_t(\bar{\theta}) \xrightarrow{p} (s-r)\cdot M \tag{35}$$

Further, one can show that

$$\bar{M} D^{-1/2}(I-P)D^{1/2}\bar{M}' = \bar{P}_\delta \tag{36}$$

where $\bar{M}$ and $\bar{P}_\delta$ are as defined in the main text of Section 1.3.

Inspect the first term of the expression in (34). Rewriting the partial sum as a difference of two partial sums, applying the result proved in 26 above as well as the continuous mapping theorem, we get

$$\begin{aligned}\frac{1}{\sqrt{T}}\, W_T^{1/2} \sum_{t=[rT]+1}^{[sT]} f(z_t,\theta_0) = \frac{1}{\sqrt{T}}\, W_T^{1/2} \sum_{t=1}^{[sT]} f(z_t,\theta_0) - \frac{1}{\sqrt{T}}\, W_T^{1/2} \sum_{t=1}^{[rT]} f(z_t,\theta_0) \\ \Rightarrow \mathcal{B}_m(s) - \mathcal{B}_m(r) \end{aligned} \tag{37}$$

Next, inspect the second term of the sum in (34). By the same result in (26), we have that $\sqrt{T}\,\Sigma_{ff}^{-1/2} F_T(\theta_0) \Rightarrow \mathcal{B}_m(1)$. Then, using the results in equations (35) and (36) as well as

the continuous mapping theorem, we have that

$$\begin{aligned} & W_T^{1/2}\ (1/T) \sum_{t=[rT]+1}^{[sT]} \nabla_\theta f_t(\bar{\theta}) D^{-1/2}(I-P)D^{-1/2}\bar{M}'\ \sqrt{T}\ \Sigma_{ff}^{-1/2}\ F_T(\theta_0) \\ & \Rightarrow (s-r)\ \bar{P}_\delta\, \mathcal{B}_m(1) \end{aligned} \tag{38}$$

Using the two convergence results in (37) and (38), the continuous mapping theorem and regrouping terms, we get

$$T^{-1/2}\ W_T^{1/2} \sum_{t=[rT]+1}^{[sT]} f(z_t, \tilde{\theta})\ \Rightarrow\ \mathcal{Z}(r,s) \tag{39}$$

$$\mathcal{Z}(r,s) \equiv \bar{P}_\delta\big[\,\mathcal{B}\,\mathcal{B}_m(s) - \mathcal{B}\,\mathcal{B}_m(r)\big] + (I_m - \bar{P}_\delta)\big[\,\mathcal{B}_m(s) - \mathcal{B}_m(r)\big] \tag{40}$$

for any $r, s \in [0,1]$ with $s > r$ where $\mathcal{B}\,\mathcal{B}_m(l) := \mathcal{B}_m(l) - l\,\mathcal{B}_m(1)$ denotes a $m \times 1$ vector of independent Brownian bridges for $l \in [0,1]$.

Finally, I derive the limiting distribution of the $\sup \Phi_T^{LM}(K)$ test statistic. To characterize the limiting distribution, I follow the strategy employed in Sowell (1996) by deriving a continuous functional[29] mapping from $D[0,1]^m$ to $\mathbb{R}$ which defines the test statistic when applied to the normalized partial sum of the sample moments between $t = [rT]+1, \ldots, [sT]$ for some $r, s \in [0,1], s > r$ in (39). The same continuous functional is then applied to the limiting stochastic process $\mathcal{Z}(r,s)$ defined in (39) to characterize the limiting distribution of the test statistic under the null hypothesis, which follows from the continuous mapping theorem.

Consider the following functional defining the $\sup \Phi_T^{LM}(K)$ test statistic for given K and given $\lambda^K \in \Lambda_\epsilon$ where $\lambda^K \equiv (\lambda_1, \ldots, \lambda_K)$, $\lambda_0 \equiv 0$ and $\lambda_{K+1} \equiv 1$ where the consistent variance estimators $\hat{\Sigma}_{ff}$ and $\hat{\Omega}_{T,j}$ have been replaced by their limits.

$$\sup \Phi_T^{LM}(K) := \sup_{\lambda^K \in \Lambda_\epsilon} \sum_{j=1}^{K+1} \mathcal{F}_{T,j}(\lambda_{j-1}, \lambda_j)'\ \ \Omega_{T,j}(\lambda_{j-1}, \lambda_j)\ \ \mathcal{F}_{T,j}(\lambda_{j-1}, \lambda_j)$$

$$\mathcal{F}_{j,T}(\lambda_{j-1}, \lambda_j) := \bar{M}_\beta'(I_m - \bar{P}_\delta)\ \ \times \frac{1}{\sqrt{T}}\, \Sigma_{ff}^{-1/2} \sum_{t=[\lambda_{j-1}T]+1}^{[\lambda_j T]} f(z_t, \tilde{\theta})$$

$$\Omega_{j,T}(\lambda_{j-1}, \lambda_j) := (\lambda_j - \lambda_{j-1})^{-1}\ \big[\bar{M}_\beta'\ (I_m - \bar{P}_\delta)\ \bar{M}_\beta\big]^{-1}$$

Apply this functional to the limiting stochastic process $\mathcal{Z}(r,s)$ defined in (39) to characterize the limiting distribution of the test statistic. The limiting stochastic process

[29] Continuous with respect to the uniform metric.

is defined by the mapping

$$\begin{aligned}\sup \Phi^{LM}(K) &:= \sup_{\lambda^K \in \Lambda_\epsilon} \sum_{j=1}^{K+1} A_j(\lambda)' \, V_j(\lambda)^{-1} \, A_j(\lambda) \\ A_j(\lambda) &:= \bar{M}_\beta'(I_m - \bar{P}_\delta) \, Z(\lambda_{j-1}, \lambda_j) \\ V_j(\lambda) &:= (\lambda_j - \lambda_{j-1}) \, \bar{M}_\beta' \, (I_m - \bar{P}_\delta) \, \bar{M}_\beta\end{aligned}$$

Consider first $A_j(\lambda)$. Using the properties of the projection matrix, $\bar{P}_\delta$, we have

$$\begin{aligned}A_j(\lambda) &= \bar{M}_\beta'(I_m - \bar{P}_\delta) \; \mathcal{Z}(\lambda_{j-1}, \lambda_j) \\ &= \bar{M}_\beta'(I_m - \bar{P}_\delta) \Big\{ \bar{P}_\delta \big[\mathcal{B}\mathcal{B}_m(\lambda_j) - \mathcal{B}\mathcal{B}_m(\lambda_{j-1}) \big] \\ &+ (I_m - \bar{P}_\delta) \big[\mathcal{B}_m(\lambda_j) - \mathcal{B}_m(\lambda_{j-1}) \big] \Big\} \\ &= \bar{M}_\beta'(I_m - \bar{P}_\delta) \, \big[\mathcal{B}_m(\lambda_j) - \mathcal{B}_m(\lambda_{j-1}) \big]\end{aligned}$$

Define $C := \bar{M}_\beta'(I_m - \bar{P}_\delta) \in \mathbb{R}^{p \times m}$. Using the result above, we have that

$$\begin{aligned}A_j(\lambda)' \, V_j(\lambda)^{-1} \, A_j(\lambda) &= \big[C \; \mathcal{Z}(\lambda_{j-1}, \lambda_j) \big]' \times (\lambda_j - \lambda_{j-1})^{-1} \\ &\times (CC')^{-1} \; \times \big[C \; \mathcal{Z}(\lambda_{j-1}, \lambda_j) \big]' \\ &= \Big\{ (\lambda_j - \lambda_{j-1})^{-1/2} \, (CC')^{-1/2} C \, \big[\mathcal{B}_m(\lambda_j) - \mathcal{B}_m(\lambda_{j-1}) \big] \Big\}' \times \\ &\Big\{ (\lambda_j - \lambda_{j-1})^{-1/2} \, (CC')^{-1/2} C \, \big[\mathcal{B}_m(\lambda_j) - \mathcal{B}_m(\lambda_{j-1}) \big] \Big\}\end{aligned}$$

where the last step follows since $(CC')^{-1}$ is a square, symmetric and positive-semidefinite matrix and therefore has a matrix square root $(CC')^{-1} = (CC')^{-1/2} \, (CC')^{-1/2}$ (Newey and McFadden, 1994, Lemma 9.6).

Next, it is easy to verify that $(CC')^{-1/2}C$ is an orthonormal $p \times m$ matrix so that it holds that $\big\{(CC')^{-1/2}C\big\} \; \big\{(CC')^{-1/2}C\big\}' = I_p$. Since $(CC')^{-1/2}C$ is orthonormal, $(CC')^{-1/2}C \, \mathcal{B}_m(s)$ has the same distribution as $\mathcal{B}_p(s)$ and we have

$$\begin{aligned}A_j(\lambda)' \, V_j(\lambda)^{-1} \, A_j(\lambda) &= [B_p(\lambda_j) - B_p(\lambda_{j-1})]' \, (\lambda_j - \lambda_{j-1})^{-1} \, [B_p(\lambda_j) - B_p(\lambda_{j-1})] \\ &= \frac{\|B_p(\lambda_j) - B_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}}\end{aligned}$$

and therefore

$$\sup \Phi^{LM}(K) = \sup_{\lambda^K \in \Lambda_\epsilon} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j) - \mathcal{B}_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

so that in conclusion we have shown that

$$\sup \Phi_T^{LM}(K) \;\Rightarrow\; \sup_{\lambda^K \in \Lambda_\epsilon} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j) - \mathcal{B}_p(\lambda_{j-1})\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

The limiting distribution of the $D \sup \Phi_T^{LM}$ statistic follows then from the continuity of the $\max$ operator in (1.11) and the continuous mapping theorem. This concludes the proof of Theorem 1.3.1 for the Lagrange-Multiplier form.

WALD FORM

The proof for the Wald form follows the same strategy as above, first showing consistency of $\hat{\beta}_j \xrightarrow{p} \beta_0$ under the null hypothesis, then deriving the limiting stochastic process of $\sqrt{T}(\hat{\beta}_j - \beta_0)$ based on the estimator defined in (1.9) and finally applying a continuous mapping to form the test statistic and to characterize its limiting distribution. The full proof is available on request.

□

C.2 Out-of-sample inference

LEMMA 1.4.1 (OOS Mean-Value Approximation): *Under the regularity conditions in Assumption 1.4.1 and the null hypothesis defined in (1.3), for any* $r, s \in [0,1]$ *with* $s > r > \rho$ *it holds that*

$$
\begin{aligned}
&P^{-1/2} \sum_{t=[rT]+1}^{[sT]} f(z_{t+h}, \beta_0, \hat{\delta}_t) = \\
&(T/P)^{1/2} \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[sT]} f(z_{t+h}, \beta_0, \delta_0) - \frac{1}{\sqrt{T}} \sum_{t=R}^{[rT]} f(z_{t+h}, \beta_0, \delta_0) \right\} \\
&+ (T/P)^{1/2} \, FB \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[sT]} H_t(\delta_0) - \frac{1}{\sqrt{T}} \sum_{t=R}^{[rT]} H_t(\delta_0) \right\} + o_{p,rs}(1)
\end{aligned}
$$

where H_t, B *are as defined in Assumption 1.4.1.(ii) and* $x_t(r,s) = o_{p,rs}(1)$ *denotes that* $\sup_{r,s \in [0,1],\ s>r>\rho} \|x_t(r,s)\| = o_p(1)$.

Proof. A second-order element-by-element mean value expansion of $f(z_{t+h}, \beta, \delta)$ around δ_0 and evaluated at $\hat{\delta}_t$ for $t = R, \ldots, T$ yields.

$$
f(z_{t+h}, \beta_0, \hat{\delta}_t) = f(z_{t+h}, \beta_0, \delta_0) + \nabla_\delta f(z_{t+h}, \beta_0, \delta_0)(\hat{\delta}_t - \delta_0) + w_{t+h} \tag{41}
$$

where the i-th element of w_{t+h} is

$$
w_{t+h,i} \equiv \frac{1}{2} (\hat{\delta}_t - \delta_0)' \, \frac{\partial^2 f_i(z_{t+h}, \beta_0, \tilde{\delta}_{t,i})}{\partial \delta \, \partial \delta'} \, (\hat{\delta}_t - \delta_0) \tag{42}
$$

and $\tilde{\delta}_{t,i}$ lies between $\hat{\delta}_t$ and δ_0.

Summing between $t = [rT] + 1, \ldots, [sT]$ and pre-multiplying by $P^{-1/2}$ gives

$$
\begin{aligned}
P^{-1/2} \sum_{t=[rT]+1}^{[sT]} f(z_{t+h}, \beta_0, \hat{\delta}_t) = \; & P^{-1/2} \sum_{t=[rT]+1}^{[sT]} f(z_{t+h}, \beta_0, \delta_0) \\
& + P^{-1/2} \sum_{t=[rT]+1}^{[sT]} \nabla_\delta f(z_{t+h}, \beta_0, \delta_0)(\hat{\delta}_t - \delta_0) \\
& + P^{-1/2} \sum_{t=[rT]+1}^{[sT]} w_{t+h}
\end{aligned} \tag{43}
$$

The second term in (43) can be written as

$$
\begin{aligned}
&P^{-1/2}\sum_{t=[rT]+1}^{[sT]}\nabla_\delta f(z_{t+h},\beta_0,\delta_0)(\hat{\delta}_t-\delta_0)= \\
&P^{-1/2}\sum_{t=[rT]+1}^{[sT]}\nabla_\delta f(z_{t+h},\beta_0,\delta_0)B_tH_t(\delta_0) \\
&=P^{-1/2}FB\sum_{t=[rT]+1}^{[sT]}H_t(\delta_0) \\
&+P^{-1/2}\sum_{t=[rT]+1}^{[sT]}\big(\nabla_\delta f(z_{t+h},\beta_0,\delta_0)-F\big)BH_t(\delta_0) \\
&+P^{-1/2}\sum_{t=[rT]+1}^{[sT]}F\big(B_t-B\big)H_t(\delta_0) \\
&+P^{-1/2}\sum_{t=[rT]+1}^{[sT]}\big(\nabla_\delta f(z_{t+h},\beta_0,\delta_0)-F\big)\big(B_t-B\big)H_t(\delta_0)
\end{aligned}
\tag{44}
$$

where the first step follows from Assumption 1.4.1.(ii) and the second step by adding and subtracting the relevant terms involving F and B. The second term after the last equality is $o_{p,rs}(1)$ by Assumption 1.4.1.(xi), the third by part 1.4.1.(xii) and the fourth by 1.4.1.(xiii). Further, it can be shown that the remainder term $P^{-1/2}\sum_{t=[rT]+1}^{[sT]}w_{t+h}(r,s)=o_{p,rs}(1)$ from an argument similar to the one in the proof of equation (4.1) in West (1996).

Substituting into (43) gives

$$
\begin{aligned}
P^{-1/2}\sum_{t=[rT]+1}^{[sT]}f(z_{t+h},\beta_0,\hat{\delta}_t) &= P^{-1/2}\sum_{t=[rT]+1}^{[sT]}f(z_{t+h},\beta_0,\delta_0) \\
&\quad+P^{-1/2}FB\sum_{t=[rT]+1}^{[sT]}H_t(\delta_0)+o_{p,rs}(1)
\end{aligned}
\tag{45}
$$

Finally, multiplying and dividing by $\sqrt{T}$, splitting the sums and re-arranging terms gives

$$
\begin{aligned}
&P^{-1/2}\sum_{t=[rT]+1}^{[sT]}f(z_{t+h},\beta_0,\hat{\delta}_t)= \\
&(T/P)^{1/2}\left\{\frac{1}{\sqrt{T}}\sum_{t=R}^{[sT]}f(z_{t+h},\beta_0,\delta_0)-\frac{1}{\sqrt{T}}\sum_{t=R}^{[rT]}f(z_{t+h},\beta_0,\delta_0)\right\} \\
&+(T/P)^{1/2}\,FB\left\{\frac{1}{\sqrt{T}}\sum_{t=R}^{[sT]}H_t(\delta_0)-\frac{1}{\sqrt{T}}\sum_{t=R}^{[rT]}H_t(\delta_0)\right\}+o_{p,rs}(1)
\end{aligned}
$$

which proves the Lemma. □

THEOREM 1.4.1 (OOS Inference): *Assume that the regularity conditions in Assumption 1.4.1 hold. Under the null hypothesis defined in (1.3), it holds that*

$$\sup \Phi_T(K) \;\Rightarrow\; \sup_{\lambda^K \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \Phi_j(\lambda_{j-1}, \lambda_j)$$

$$D \sup \Phi_T(\bar{K}) \;\Rightarrow\; \max_{1 \leq k \leq \bar{K}} (1/k) \sup_{\lambda^K \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \Phi_j(\lambda_{j-1}, \lambda_j)$$

$$\Lambda_{\epsilon,\rho} = \Big\{ \lambda_j,\ j = 1, \ldots, K :\ \lambda_j \in (\rho + \epsilon, 1 - \epsilon),\ \lambda_j > \lambda_{j-1} + \epsilon \Big\},$$

with $\lambda_0 \equiv \rho,\ \lambda_{K+1} \equiv 1$ *and where*

$$\begin{aligned} \Phi_j(\lambda_{j-1}, \lambda_j) \equiv & \Big[\mathcal{B}_m \left(\int_0^{\lambda_j} \omega(u, \lambda_{j-1}, \lambda_j) \omega(u, \lambda_{j-1}, \lambda_j)' \, du \right) \Big]' \times \\ & \left\{ \int_0^{\lambda_j} \omega(u, \lambda_{j-1}, \lambda_j)\, \omega(u, \lambda_{j-1}, \lambda_j)' \, du \right\}^{-1} \\ & \times \Big[\mathcal{B}_m \left(\int_0^{\lambda_j} \omega(u, \lambda_{j-1}, \lambda_j) \omega(u, \lambda_{j-1}, \lambda_j)' \, du \right) \Big] \end{aligned}$$

with

$$\begin{aligned} \omega(u, r, s) \equiv M' \, \Sigma_{ff}^{-1} \, (1 - \rho)^{-1/2} \Big[I_m \quad FB \Big] \times \Big\{ & \Big[\Omega(u, s)^{1/2} - \Omega(u, r)^{1/2} \Big] \mathbb{1}(u \leq r) \\ & + \Omega(u, s)^{1/2} \, \mathbb{1}(r < u \leq s) \Big\} \, \Sigma^{1/2} \end{aligned}$$

and where $\Omega(s, \tau)$ *is as defined as*

$$\Omega(s, \tau)^{1/2} \equiv \begin{pmatrix} \mathbb{1}(s \leq \rho) \cdot I_m & 0_{m \times d} \\ 0_{d \times m} & \big\{ [\ln \tau - \ln \rho] \, \mathbb{1}(s \leq \rho) + [\ln(\tau) - \ln(s)] \, \mathbb{1}(\rho < s \leq \tau) \big\} \cdot I_d \end{pmatrix}$$

Proof. Note, as in the in-sample case above, that under the null hypothesis in (1.3), it holds that $\theta_t = \theta_0 = (0_{p \times 1}, \delta) \ \forall \ t$. Therefore, under the null hypothesis, the function defining the moment condition $f(z_t, \theta_t) = f(z_t, \beta_t, \delta)$ can be written as a function of a constant parameter $f(z_t, \theta)$. This notation will be used in the subsequent derivations.

To prove the weak convergence results stated in the theorem, I start by showing that the partial sample moments satisfy the following invariance principle under the null hypothesis.

$$\frac{1}{\sqrt{T}} \sum_{t=R}^{[sT]} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ H_t(\delta_0) \end{pmatrix} \;\Rightarrow\; \int_0^s \Omega(u, s)^{1/2} \, d\xi(u) \tag{46}$$

where $\Omega(u,\tau)^{1/2}$ is as defined in the theorem and $\xi(u) \equiv \Sigma^{1/2}\,\mathcal{B}_{m+d}(u)$ where $\mathcal{B}_{m+d}(u)$ is an $(m+d)\times 1$ vector of independent standard Brownian motions.

This result can be proven by following the same reasoning as in the proof of Proposition 1 of Rossi and Sekhposyan (2016). Start by defining $b_{R,t,j} \equiv \mathbb{1}(t \geq R)$. Then, by direct calculations, for any $j \geq R$, it holds that

$$\sum_{t=R}^{j} f(z_{t+h},\beta_0,\delta_0) = \sum_{t=1}^{j} b_{R,t,j}\; f(z_{t+h},\beta_0,\delta_0) \tag{47}$$

Under the recursive estimation scheme in Assumption 1.4.1.(ii) and defining

$$a_{R,t,j} \equiv \left(R^{-1}+\ldots+j^{-1}\right)\cdot \mathbb{1}(t\leq R) \;+\; \left(t^{-1}+\ldots+j^{-1}\right)\cdot \mathbb{1}(R<t\leq j) \tag{48}$$

for any $j \geq R$ it holds by direct calculation that

$$\sum_{t=R}^{j} H_t(\delta_0) = \sum_{t=R}^{j} t^{-1}\left(\sum_{r=1}^{t} h(z_r,\delta_0)\right) = \sum_{t=1}^{j} a_{R,t,j}\; h(z_t,\delta_0) \tag{49}$$

Using (47) and (49) it holds that

$$\frac{1}{\sqrt{T}}\sum_{t=R}^{[sT]}\begin{pmatrix} f(z_{t+h},\beta_0,\delta_0) \\ H_t(\delta_0)\end{pmatrix} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\begin{pmatrix} b_{R,t,[sT]}\cdot I_m & 0_{m\times d} \\ 0_{d\times m} & a_{R,t,[sT]}\cdot I_d\end{pmatrix}\begin{pmatrix} f(z_{t+h},\beta_0,\delta_0) \\ h_t(\delta_0)\end{pmatrix} \tag{50}$$

To derive the limiting distribution, as in Rossi and Sekhposyan (2016), I consider an asymptotic approximation for the weights $a_{R,t,j}$ and $b_{R,t,j}$. From Assumption 1.4.1.(i), we have $\rho := \lim_{T\to\infty} R/T$ and thus

$$b_{R,t,j} \equiv \mathbb{1}(t\geq R) \approxeq \mathbb{1}(s\geq\rho) \qquad s \equiv \lim_{T\to\infty} t/T \tag{51}$$

Following West (1996) and Rossi and Sekhposyan (2016) it can further be shown that

$$\begin{aligned} a_{R,t,j} &\approxeq \left(\int_R^j \frac{1}{k}\,dk\right)\mathbb{1}(t\leq R) + \left(\int_t^j \frac{1}{k}\,dk\right)\mathbb{1}(R<t\leq j) \\ &\approxeq [\ln(\tau)-\ln(\rho)]\,\mathbb{1}(s\leq\rho) + [\ln(\tau)-\ln(s)]\,\mathbb{1}(\rho<s\leq\tau) \end{aligned} \tag{52}$$

To prove the weak convergence in (46), I employ the result for weak convergence of stochastic integrals based on mixing sequences of Hansen (1992). In particular, define $\{\xi_{j,T}\}$ to be the following normalized stochastic sum process

$$\xi_j \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{j}\xi(z_{t+h},\beta_0,\delta_0) \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{j}\begin{pmatrix} f(z_{t+h},\beta_0,\delta_0) \\ h_t(\delta_0)\end{pmatrix} \tag{53}$$

where $\xi(z_{t+h}, \theta)$ is defined in the main text of Section 1.4. Further, define the stochastic integral of interest as

$$\int_0^\tau \begin{pmatrix} \sigma_f(s) \cdot I_m & 0_{m\times d} \\ 0_{d\times m} & \sigma_h(s,\tau) \cdot I_d \end{pmatrix} d\,\xi_T = \tag{54}$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{j} \begin{pmatrix} b_{R,t,j} \cdot I_m & 0 \\ 0 & a_{R,t,j} \cdot I_d \end{pmatrix} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ h_t(\delta_0) \end{pmatrix} \tag{55}$$

To apply Theorem 3.1 of Hansen (1992) we need to verify its conditions. The first requirement is Assumption 1 of Hansen (1992) which is satisfied by the mixing condition in Assumption 1.4.1.(iii) and Assumption 1.4.1.(v). To satisfy the second requirement, we need to show that $T^{1/2}\ \xi_T \Rightarrow \Sigma^{-1/2}\xi$ where $\xi(s) \equiv \Sigma^{1/2}\,\mathcal{B}_{m+d}(s)$. This follows from applying Corollary 2.2 of Phillips and Durlauf (1986) under Assumptions 1.4.1.(iii), 1.4.1.(iv), 1.4.1.(v) and 1.4.1.(vi).

Applying Theorem 3.1 of Hansen (1992), we get

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{j} \begin{pmatrix} b_{R,t,j} \cdot I_m & 0 \\ 0 & a_{R,t,j} \cdot I_d \end{pmatrix} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ h_t(\delta_0) \end{pmatrix} - C_T^*(\tau) \tag{56}$$

$$\Rightarrow \int_0^\tau \Omega(s,\tau)^{1/2}\, d\xi(s) \tag{57}$$

where

$$C_T^*(\tau) = \Bigg\{ T^{-1/2} \sum_{t=1}^{[\tau T]} \Bigg[\begin{pmatrix} b_{R,t,j} \cdot I_m & 0 \\ 0 & a_{R,t,j} \cdot I_d \end{pmatrix} - \begin{pmatrix} b_{R,t-1,j} \cdot I_m & 0 \\ 0 & a_{R,t-1,j} \cdot I_d \end{pmatrix} \Bigg] \zeta_t$$
$$- T^{-1/2} \begin{pmatrix} b_{R,t-1,j} \cdot I_m & 0 \\ 0 & a_{R,t-1,j} \cdot I_d \end{pmatrix} \zeta_{j+1} \Bigg\}$$

with $j := [\tau T]$ and $\zeta_t = \sum_{k=1}^{\infty} \mathbb{E}_t \left(\left[f(z_{t+h+k}, \theta_0)', h(z_t, \delta_0)' \right]' \right)$. Using the same reasoning as in Rossi and Sekhposyan (2016), based on the steps in the proof of Cavaliere (2005), Theorem 4 and the fact that the variances $\sigma_f(s), \sigma_h(s,\tau)$ are square integrable and bounded, we get

$$\frac{1}{\sqrt{T}} \sum_{t=R}^{[sT]} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ H_t(\delta_0) \end{pmatrix} \Rightarrow \int_0^\tau \Omega(s,\tau)^{1/2}\, d\xi(s) \tag{58}$$

LAGRANGE-MULTIPLIER FORM, Φ_T^{LM}

To derive the limiting distribution of the $\sup \Phi_T(K)$ test statistic, we follow the same strategy as in the proof of Theorem 1.3.1. The test statistic is formed by applying a continuous functional to the stochastic process derived above. We then derive the associated limiting stochastic process under the null hypothesis and apply the same functional to characterize the limiting distribution of the test statistic. Given K, define the continuous mapping

$$\sup \Phi_T^{LM}(K) := \sup_{\lambda^K \in \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} A_{j,T}(\lambda)' \left\{V_{j,T}(\lambda)\right\}^{-1} A_{j,T}(\lambda) \tag{59}$$

where

$$A_{j,T}(\lambda) := M' \Sigma_{ff}^{-1} P^{-1/2} \sum_{t=[\lambda_{j-1}T]+1}^{[\lambda_j T]} f(z_{t+h}, \beta_0, \hat{\delta}_t) \tag{60}$$

To characterize the limiting stochastic process of $A_{j,T}(\lambda)$, note that under the regularity conditions in Assumptions 1.4.1, we can apply Lemma 1.4.1 to get

$$\begin{aligned}
&P^{-1/2} \sum_{t=[\lambda_{j-1}T]+1}^{[\lambda_j T]} f(z_{t+h}, \beta_0, \hat{\delta}_t) = \\
&(T/P)^{1/2} \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_j T]} f(z_{t+h}, \beta_0, \delta_0) - \frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_{j-1} T]} f(z_{t+h}, \beta_0, \delta_0) \right\} \\
&+ (T/P)^{1/2} \, FB \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_j T]} H_t(\delta_0) - \frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_{j-1} T]} H_t(\delta_0) \right\} + o_{p,rs}(1)
\end{aligned}$$

Plugging into (60) and grouping terms and omitting the o_p term, we get

$$\begin{aligned}
A_{j,T}(\lambda) =& M' \Sigma_{ff}^{-1} (T/P)^{1/2} \begin{bmatrix} I_m & FB \end{bmatrix} \left\{ \frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_j T]} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ h_t(\delta_0) \end{pmatrix} \right. \\
&\left. - \frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_{j-1} T]} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ h_t(\delta_0) \end{pmatrix} \right\}
\end{aligned} \tag{61}$$

Using the weak convergence result we derived in (58) above, we get

$$\frac{1}{\sqrt{T}} \sum_{t=R}^{[\lambda_j T]} \begin{pmatrix} f(z_{t+h}, \beta_0, \delta_0) \\ H_t(\delta_0) \end{pmatrix} \Rightarrow \int_0^{\lambda_j} \Omega(s, \lambda_j)^{1/2} S^{1/2} \, d\mathcal{B}_{m+d}(s) \tag{62}$$

and applying the continuous mapping theorem,

$$
\begin{aligned}
A_{j,T} \;\Rightarrow\; & M'\Sigma_{ff}^{-1}(1-\rho)^{-1/2}\left[I_m \quad FB\right]\left\{\int_0^{\lambda_j}\Omega(s,\lambda_j)^{1/2}S^{1/2}\,d\mathcal{B}_{m+d}(s)\right. \\
& \left.-\int_0^{\lambda_{j-1}}\Omega(s,\lambda_{j-1})^{1/2}S^{1/2}\,d\mathcal{B}_{m+d}(s)\right\} \\
& = M'\Sigma_{ff}^{-1}(1-\rho)^{-1/2}\left[I_m \quad FB\right]\left\{\int_0^{\lambda_{j-1}}\left[\Omega(s,\lambda_j)^{1/2}-\Omega(s,\lambda_{j-1})^{1/2}\right]\right. \\
& \left.\times S^{1/2}\,d\mathcal{B}_{m+d}(s)+\int_{\lambda_{j-1}}^{\lambda_j}\Omega(s,\lambda_j)^{1/2}S^{1/2}\,d\mathcal{B}_{m+d}(s)\right\} \\
& = M'\Sigma_{ff}^{-1}(1-\rho)^{-1/2}\left[I_m \quad FB\right]\times \\
& \int_0^{\lambda_j}\left\{\left[\Omega(s,\lambda_j)^{1/2}-\Omega(s,\lambda_{j-1})^{1/2}\right]\mathbb{1}(s\le\lambda_{j-1})\right. \\
& \left.+\,\Omega(s,\lambda_j)^{1/2}\,\mathbb{1}(\lambda_{j-1}<s\le\lambda_j)\right\} \\
& \times S^{1/2}\,d\mathcal{B}_{m+d}(s) \\
& = \int_0^{\lambda_j}\omega(s,\lambda_j,\lambda_{j-1})\,d\mathcal{B}_{m+d}(s) \\
& = \mathcal{B}_m\left(\int_0^{\lambda_j}\omega(s,\lambda_j,\lambda_{j-1})\omega(s,\lambda_j,\lambda_{j-1})'\,ds\right)
\end{aligned}
\tag{63}
$$

where

$$
\begin{aligned}
& \omega(s,\lambda_{j-1},\lambda_j)\equiv \\
& M'\Sigma_{ff}^{-1}(1-\rho)^{-1/2}\left[I_m \quad FB\right]\left\{\left[\Omega(s,\lambda_j)^{1/2}-\Omega(s,\lambda_{j-1})^{1/2}\right]\mathbb{1}(s\le\lambda_{j-1})\right. \\
& \left.+\,\Omega(s,\lambda_j)^{1/2}\,\mathbb{1}(\lambda_{j-1}<s\le\lambda_j)\right\}S^{1/2}
\end{aligned}
\tag{64}
$$

The result of the theorem then follows from applying the continuous mapping theorem to get the distribution of $\Phi_T(\lambda)$ and the distributions of $\sup\Phi_T(K)$ and $D\sup\Phi_T(\bar{K})$ as defined in the theorem and analogue to the proof of Theorem 1.3.1.

WALD FORM

The proof of the Wald form follows similar steps and is available on request. □

COROLLARY 1.4.1 (OOS Inference in Special Cases): *If (a)* $F = 0$*, that is parameter estimation error is irrelevant, or (b) the following condition holds*

$$\Sigma_{ff} = -\frac{1}{2}(FB\Sigma_{hf} + \Sigma_{fh}B'F') = FB\Sigma_{hh}B'F'$$

then, the result of Theorem 1.4.1 simplifies to

$$\sup \Phi_T(K) \;\Rightarrow\; \sup_{\lambda \,\in\, \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j - \rho) - \mathcal{B}_p(\lambda_{j-1} - \rho)\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

$$D\sup \Phi_T(\bar{K}) \;\Rightarrow\; \max_{1\le k\le \bar{K}} (1/k) \sup_{\lambda \,\in\, \Lambda_{\epsilon,\rho}} \sum_{j=1}^{K+1} \left\{ \frac{\|\mathcal{B}_p(\lambda_j - \rho) - \mathcal{B}_p(\lambda_{j-1} - \rho)\|^2}{\lambda_j - \lambda_{j-1}} \right\}$$

Proof. The proof follows from directly calculating $\int_0^{\lambda_j} \omega(s, \lambda_j, \lambda_{j-1})\omega(s, \lambda_j, \lambda_{j-1})' \; ds$, imposing the condition given in the corollary. The proof follows similar steps to the proofs of Proposition 3,4, & 7 of Rossi and Sekhposyan (2016). □

Chapter 2

Has the Information Channel of Monetary Policy Disappeared? Revisiting the Empirical Evidence

This paper is co-authored with Barbara Rossi and Tatevik Sekhposyan.

2.1 Introduction

The recent literature has found that, in response to unexpected increases in interest rates, survey-based estimates of expected output growth rise while those of inflation decline (Campbell et al., 2012, 2017). This is contrary to the common New Keynesian wisdom that contractionary monetary policy causes a decline in output growth and inflation as well as their expectations. An explanation for this puzzling behavior is the so-called "information channel" of monetary policy. According to the information channel, agents update their beliefs after an unexpected monetary policy action not only because they learn about the current and future path of monetary policy, but also because they learn new information about economic fundamentals. The intuition is that the Federal Reserve communicates not only the future path of monetary policy but also how optimistic it is about the current and future state of the economy: if the Federal Reserve's expectation of future fundamentals is different from the state of the economy perceived by market participants, market participants will update their expectations. In this case, the responses to a monetary policy shock may not be estimated correctly. In fact, if the monetary policy tightening is the endogenous reaction to a future state of the economy that is more positive than markets anticipate, market participants might expect an increase in future output and inflation and update their expectations accordingly.

A sufficient, but not necessary, condition for the information channel is that the central bank has superior knowledge about the state of the economy relative to market participants; that is, it has an "information advantage". When that is the case, it is likely that market participants will update their information about the state of the economy based on the new information contained in central banks' announcements.[1]

How important is the information channel empirically? Does the Federal Reserve indeed have an information advantage in forecasting macroeconomic variables beyond what is known to private forecasters? And does this matter when estimating the response of the economy to monetary policy shocks?

We revisit the empirical evidence by making an important departure from the previous literature; namely, we allow for instabilities. That is, we allow the information advantage of the Federal Reserve relative to market participants to change over time. As we show, this is an important empirical feature of the data. Furthermore, we also allow the nature of monetary policy shocks to vary over time, depending on whether the information advantage is present in the data. We show that the latter matters when estimating the effects of a monetary policy shock in the economy. When the central bank has an informational advantage, the information channel is at work: the macroeconomic responses are confounded and the researcher may estimate an expansionary response to a contractionary monetary policy shock. On the other hand, when the information advantage disappears, the information channel loses importance: the perils of confoundedness disappear and researchers are able to correctly recover the response in the data.

Once we take instabilities into account, we find substantially weaker evidence for the empirical relevance of the information channel in the most recent period: (i) The Federal Reserve lost its short horizon information advantage regarding the state of the economy relative to market participants; (ii) market surprises are no longer predictable by the Federal Reserve's internal forecasts; and (iii) macroeconomic responses to a monetary policy shock are no longer confounded and private forecasters' responses are less sensitive to monetary policy shocks. Our results are consistent with the hypothesis that the decline in the relevance of the information channel is linked to the improved communication strategies of the Federal Reserve in recent years.

RELATED LITERATURE. Our paper is related to various strands of the literature. First, our paper is related to the large literature estimating the effects of monetary policy shocks – see Christiano et al. (1999, 2005), among others. Traditional Vector Autoregressions (VAR) conventionally estimate a positive, hump-shaped response of output and inflation to

[1] Note that this is a sufficient but not necessary condition. The private sector might be updating its expectations on the state of the economy even when the central bank does not have an information advantage. For example, Timmermann (2006) has shown that forecast combination may improve forecast accuracy even if the combination uses noisy or biased forecasts as long as they are not perfectly correlated with the forecasts already included in the pool. Furthermore, Morris and Shin (2002) develop a theoretical model where public information can affect the equilibrium outcomes even if it is noisier than private information.

an expansionary monetary policy shock measured by an exogenous increase in the Federal Funds Rate (FFR). However, as Campbell et al. (2012) and Nakamura and Steinsson (2018) show, survey expectations of output growth typically rise after an unexpected monetary policy tightening, thus contradicting the predictions of standard economic models.

Melosi (2017) and Nakamura and Steinsson (2018) develop theoretical models to rationalize this "real activity puzzle". In Melosi (2017)'s model, policy actions are publicly observable, but private information about the economy's fundamentals is "dispersed" across market participants and policymakers. Thus, a change in the current policy rate not only affects real interest rates but also provides the public with information on the central bank's view about macroeconomic developments. Melosi (2017) refers to the latter channel as the "signaling channel" of monetary transmission and provides the first econometric analysis of such "signaling effects". He finds that his model can explain the forecast errors observed in the data while perfect information models cannot. Nakamura and Steinsson (2018) suggest a similar explanation based on the information channel of monetary policy. In their model, monetary policy shocks affect not only the real interest rate but also the private sector's belief about the path of the natural rate of interest; this happens because, as the central bank tracks the natural rate of interest, its announcements are likely to contain news about it. They find strong empirical support for both the Federal Reserve (Fed) information channel of monetary policy and the conventional one.

The presence of the information channel and its *evolution over time* is the objective of our paper. Miranda-Agrippino and Ricco (2020) investigate the responses of core macroeconomic aggregates to monetary policy announcements using an identification strategy robust to information frictions. Their instrument is the component of high-frequency market-based monetary surprises at the time of a policy announcement that is orthogonal to both the central bank's own forecasts as well as previous market surprises. In a similar spirit, Jarociński and Karadi (2020) use sign restrictions to separately identify monetary policy and "information shocks" from stock price dynamics. Andrade and Ferroni (2021) investigate the information channel in forward guidance announcements in the euro area and find evidence of information effects.[2]

A series of papers have investigated the empirical importance of the information channel of monetary policy, including Cieslak and Schrimpf (2019), Lunsford (2020), Paul (2020) and Bauer and Swanson (2020). Cieslak and Schrimpf (2019) and Lunsford (2020) classify central banks' announcements according to their characteristics. Cieslak and Schrimpf (2019) distinguish among different types of central bank communication news: news about monetary policy, economic growth and financial risk premia. Their analysis results in a comprehensive database of international monetary policy events classified according to their information content. They show that news about monetary policy

[2]On the other hand, Bundick and Smith (2020) and Inoue and Rossi (2021) show, using completely different methodologies, that the response of the economy to forward guidance shocks is indeed consistent with standard New Keynesian models' predictions.

prevails in announcements about monetary policy decisions; news about economic growth prevails in press conferences; and the importance of risk premium shocks increases in the unconventional monetary policy period. However, they do not investigate time variation in the effects of monetary policy news. Lunsford (2020) investigates whether the type of forward guidance language used by the Fed can influence information effects. He shows that forward guidance shapes the private sector's responses to Federal Open Market Committee (FOMC) monetary policy statements and that forward guidance on the economic outlook has stronger information effects than communication on the policy inclination. Similarly to us, he finds time variation in the transmission of monetary policy shocks. In particular, analyzing the period from February 2000 to May 2006, he finds evidence of a structural break in the magnitude of the Federal Funds Rate surprises in August 2003. He argues that, before this break, FOMC statements only included forward guidance about the economic outlook, while they also included information on the FOMC policy inclinations after the break. He concludes that financial markets, survey forecasters as well as the macroeconomy react differently depending on the type of forward guidance. Our analysis differs from Lunsford (2020) in that we formally evaluate the relative importance of changes in the economic outlook versus pure monetary policy in a larger sample rather than focusing on different forms of forward guidance in a specific sample period. Paul (2020) and Bauer and Swanson (2020) instead focus on alternative explanations for the puzzling response of survey forecasters to monetary policy statements. Paul (2020) finds that the puzzling increase in private forecasters' expectations of output growth to unexpected increases in interest rates is present only when including unscheduled meetings. Bauer and Swanson (2020) find that the puzzling estimates are consistent not only with the central bank's information channel of monetary policy but also with the central bank's response to macroeconomic news. According to the latter, economic news simultaneously causes changes in the Fed's monetary policy as well as in the private sector's forecasts and there is little role for an information effect. Differently from Paul (2020) and Bauer and Swanson (2020), we focus on the interrelation between the information advantage and the information channel of monetary policy and investigate their evolution over time.

Second, there is a large literature that focuses on the evaluation of central banks' forecasts as well as the quality of the Fed's internal forecasts relative to the private sector's. In a seminal contribution, Romer and Romer (2000) showed that the Federal Reserve has an information advantage relative to the private sector when forecasting inflation. On the other hand, both D'Agostino and Whelan (2008) and Rossi and Sekhposyan (2016) find evidence of instabilities in the information advantage. D'Agostino and Whelan (2008) show that the Federal Reserve's superior forecasting performance relative to survey forecasts deteriorated since the early 1990s across medium to long forecast horizons. Our contribution is instead to document that even the central bank's short-horizon forecast advantage disappeared in the most recent period. In addition, while the analysis in D'Agostino and Whelan (2008) is based on ad-hoc sub-samples, we consider general patterns of time variation and let

the data uncover the time period when the forecast advantage appears/disappears. Rossi and Sekhposyan (2016) also show that the evidence of central banks' forecast advantage in predicting inflation depends on the time period, and has deteriorated over time. Our paper considers instead a wider range of macroeconomic variables (such as real output growth, unemployment and, especially, interest rates) and links the forecast advantage to the information channel of monetary policy. In relation to this literature, our paper conducts a comprehensive analysis of the information channel of monetary policy, looking at various dimensions considered in the literature beyond forecast evaluation.

Lastly, our work is more distantly related to the literature on forecast rationality, in particular Faust and Wright (2009), Patton and Timmermann (2012), Croushore (2012) and Rossi and Sekhposyan (2016). Faust and Wright (2009), Patton and Timmermann (2012) and Croushore (2012) note that rationality of inflation forecasts depends on the sample period used for forecast evaluation, while Rossi and Sekhposyan (2016) formally investigate the rationality of the central bank as well as the private sector inflation forecasts in the presence of instabilities.

The remainder of the paper proceeds as follows. Section 2.2 presents our analysis of the Federal Reserve's information advantage. Section 2.3 investigates the time-varying information content of high-frequency market-based monetary surprises typically used in the literature to identify monetary policy shocks. Section 2.4 investigates the empirical relevance of the information channel for determining the economy's response to monetary policy while Section 2.5 investigates the reaction of private forecasters. Section 2.6 concludes.

2.2 Does the Federal Reserve have an information advantage?

This section revisits the empirical evidence on whether the Federal Reserve has more information than the private sector when forecasting key macroeconomic variables. We establish that, while the Federal Reserve historically had an information advantage when forecasting real GDP growth and inflation, at least at short horizons, this advantage disappeared in the recent period. We also estimate several important change-points in the information advantage that coincide with changes in FOMC communication policy. Finally, we discuss the relationship between the information advantage and relative forecast accuracy (measured by a mean squared error loss function, MSFE) and show that, based on the latter criterion, the Federal Reserve's advantage deteriorated as well.

2.2.1 The evolution of the Federal Reserve's information advantage

To assess whether the Federal Reserve has an information advantage over the private sector in forecasting a macroeconomic variable, x, we consider the following information

advantage regression:

$$x_{t+h} - x^{BCEI}_{t+h|t} = \delta + \beta_{GB}\, x^{GB}_{t+h|t} + \beta_{BCEI}\, x^{BCEI}_{t+h|t} + \eta_{t+h} \tag{2.1}$$

where $x^{GB}_{t+h|t}$ is the Greenbook/Tealbook forecast at horizon h, $x^{BCEI}_{t+h|t}$ is the consensus forecast from the Blue Chip Economic Indicators' (BCEI) survey at horizon h, x_{t+h} denotes the real-time realization of the variable of interest and η_{t+h} is an unforecastable error term. Our goal is to test whether β_{GB} equals zero. In fact, the Fed's forecasts provide additional information above and beyond that in the private sector's forecasts if β_{GB} is different from zero. In other words, a value of β_{GB} different from zero indicates that forecasters would prefer to put weight on both Greenbook/Tealbook forecasts as well as BCEI forecasts if they had a choice.

Because the coefficients in the above regression could be time-varying, we investigate whether β_{GB} is different from zero using tests robust to instabilities. Tests based on the full sample characterize the average out-of-sample performance, which might mask the evolution of the information advantage over time (Rossi, 2006a). Instead, we base our analysis on what we label as the "Information-Advantage Fluctuation test" using the general framework in Rossi and Sekhposyan (2016). Specifically, we estimate the information advantage regression in eq. (2.1) in rolling windows of m forecasts.[3] Let $\beta_{GB,t}$ be the time-varying parameter and let $\widehat{\beta}_{GB,t}$ denote the parameter estimated sequentially in regression (2.1) for $t = m/2, ..., T - m/2$ using observations centered around time t – that is, the most recent $m/2$ observations as well as the following $m/2$ ones. We then construct a t-statistic at each point in time t:

$$\tau_{GB,t} = \widehat{\beta}_{GB,t} / \sqrt{\widehat{\sigma}^2_{GB}/m} \tag{2.2}$$

where $\widehat{\sigma}^2_{GB}$ is the Newey and West (1987) HAC estimator of the asymptotic variance of the parameter estimate in the rolling window centered at time t.[4] The Information-Advantage Fluctuation test statistic is:

$$\mathcal{F}_{GB} = max_t \left| \tau_{GB,t} \right|, \tag{2.3}$$

which we use to test the null hypothesis that $\beta_{GB,t} = 0$ at every point in time t against the alternative that $\beta_{GB,t} \neq 0$ at some point in time t.

Figures 2.1 and 2.2 depict $\tau_{GB,t}$ over time. The largest (absolute) value in the sequence of $\tau_{GB,t}$ is the Information-Advantage Fluctuation test statistic, $\mathcal{F}_{GB}$. The (red) dashed horizontal line denotes the corresponding five percent critical value.[5] When $\mathcal{F}_{GB}$ is outside

[3] Without loss of generality, m is an even integer.

[4] The variance estimate is based on a Newey and West (1987) HAC estimator using a truncation lag equal to $m^{1/4}$. For details on the variance estimator, see Rossi and Sekhposyan (2016). The results presented in this section are robust to using a heteroskedasticity-consistent variance estimator instead of the HAC estimator.

[5] The relevant critical value is the t-statistic analog to the Wald test critical values for the survey and model-free

the critical value lines, the test rejects the null hypothesis that $\beta_{GB,t} = 0$ for every t, and we conclude that the central bank had an information advantage at some point in time. Importantly, the critical value properly controls size and guards against sequential testing bias.[6]

The path of $\tau_{GB,t}$ is a local measure of forecast advantage over the rolling window, which we attribute to the center point of the window itself (similarly to usual non-parametric approaches).[7] Since the rolling-window approach involves smoothing, by construction the date is only indicative of when the forecast advantage started/ended; later in this section, we complement our results with Bai and Perron (1998)'s test of multiple discrete breaks. Also, for completeness, we report the coefficient estimates ($\widehat{\beta}_{GB,t}$), which have the same sign as $\tau_{GB,t}$, in the Not-for-Publication Appendix.

DATA. To implement the Information-Advantage Fluctuation test in eq. (2.1), we require data on the central bank's and private sector's forecasts as well as the corresponding real-time realizations for key macroeconomic variables. In our analysis, we consider forecasts of inflation, GDP growth, unemployment and the interest rate. We also require a strategy to match the forecasts and realizations so that their targets align. As a measure of central bank forecasts, we use the Greenbook/Tealbook forecasts between February 1984 and December 2015, which are prepared by the staff of the Federal Reserve prior to each regular FOMC meeting (eight times per year). These forecasts are based on a maintained assumption about monetary policy and are made available to the public after a five year lag. This lag constrains the end of our sample period. For inflation, we use the forecasts of the annualized, chain-weighted quarter-over-quarter growth in the GDP deflator. For GDP growth, we use the forecasts of the annualized, chain-weighted quarter-over-quarter real GDP growth rate. For the unemployment rate, we use the Greenbook/Tealbook projections for the quarterly average unemployment rate in percentage points. Finally, for the interest rate, we use the projections of the three month Treasury bill rate.

As a measure of private sector forecasts, we use the Blue Chip Economic Indicators (BCEI), which is a monthly commercial survey-based dataset containing consensus (average) forecasts for 16 macroeconomic variables collected from approximately 50 business economists. We consider only forecasts up to four quarters since the series exhibit missing values occurring systematically beyond this horizon. Since the BCEI forecasts are for fixed events, i.e. for selected quarters in reference years, forecasts beyond four quarters are not available for every FOMC round.[8] For inflation, we use the annualized quarter-over-

forecasts reported in Table II Panel C of Rossi and Sekhposyan (2016).

[6]Note that a rejection does not simply indicate time-variation: the test rejects the hypothesis that the central bank never had an information advantage relative to the survey forecasters. Importantly, the test would also reject if there was no time variation, but the central bank had a constant information advantage. Thus, the path of $\tau_{GB,t}$ contains valuable information on the reason behind the rejection.

[7]Therefore, note that the last year in the figures is 2011 only because that is the center point in the last window we consider: our sample in fact ends in December 2015.

[8]For example, for the five-quarter-ahead forecasts of all the variables considered in our analysis, the BCEI

quarter BCEI consensus forecasts of the GDP deflator price index. For GDP growth, we use the annualized quarter-over-quarter consensus forecasts of real GDP growth and for the unemployment rate the consensus forecasts of the quarterly average of the unemployment rate in percentage points. Finally, for the interest rate we use the forecasts of the quarterly average yield on a three month Treasury bill in percentage points. The forecasts are available for our entire sample from February 1984 - December 2015. At the beginning of the sample, the BCEI survey was conducted over three days, beginning on the first working day of each month, and was subsequently shortened to two days starting in December 2000. The BCEI consensus forecasts are released on the 10th of each month.

To implement the regression in eq. (2.1), we match the BCEI forecasts to the Greenbook/Tealbook forecasts so that the BCEI forecast is always strictly before the FOMC meeting associated with each Greenbook/Tealbook forecast. This results in the BCEI forecasts sometimes being published before the Greenbook/Tealbook forecasts and sometimes after, but both forecasts are made strictly before the FOMC announcement. Appendix A reports sensitivity analyses using an alternative timing assumption that strictly orders BCEI forecasts before the Greenbook/Tealbook forecasts.

For the realizations, we use real-time data from the Philadelphia Fed's "Real-Time Data Set for Macroeconomists". We use the quarterly first-release values where available and, in a handful of cases, we impute any missing values using second-release values. We use the first-releases because of their timeliness. The Not-for-Publication Appendix shows that our results are robust to using second and third releases for output growth and inflation (interest rates are never revised and revisions to the unemployment rate are negligible). We do not consider final releases since they include ex-post re-definitions and major classification changes that the forecasters would not have known at the time the forecasts were made. For realized inflation, we use the annualized quarter-over-quarter growth rate in the GNP/GDP deflator price index in percentage points; for realized GDP, we use the annualized quarter-over-quarter growth rate in real GNP/GDP in percentage points; for the unemployment rate, we use the quarterly average of the monthly history of (quarterly) vintages provided by the Federal Reserve Bank of Philadelphia. Interest rate data is not revised, thus we use the average quarterly secondary market rate of the three month Treasury bill (TB3MS), which we obtain from the FRED database maintained by the Federal Reserve Bank of St. Louis.

Details on the Greenbook/Tealbook forecasts, the BCEI forecasts, real-time data as well as the data sources are documented in the Not-for-Publication Appendix.

forecasts for any FOMC meeting occurring in the last quarter of each year is systematically missing, as survey respondents were only asked to forecast until the last quarter of the next year. This corresponds to two or three meetings out of the eight regular FOMC meetings per year.

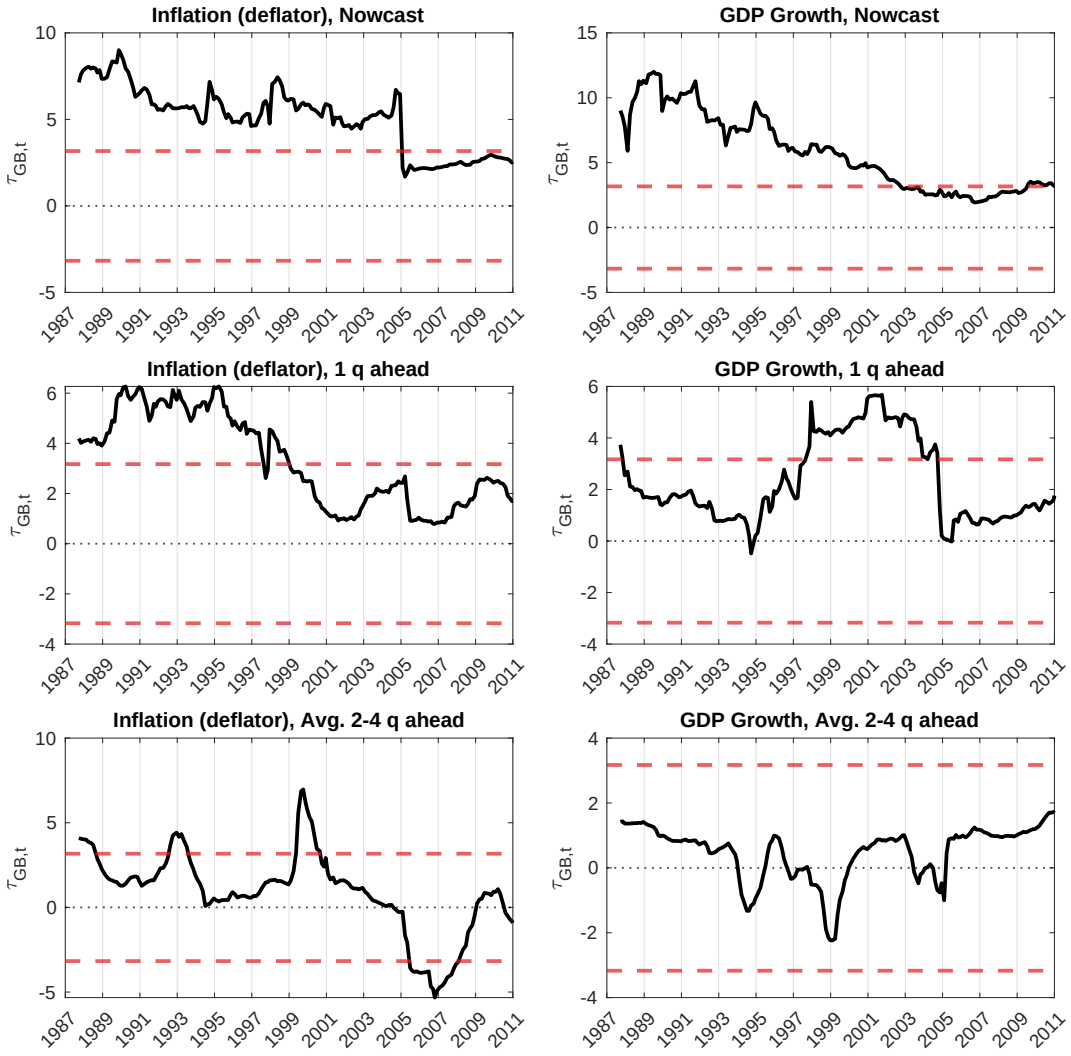


Figure 2.1: Information advantage fluctuation test: Inflation and GDP growth

Notes: The figure shows $\tau_{GB,t}$ from eq. (2.1) based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. Dashed (red) lines denote 5% critical value lines based on Rossi and Sekhposyan (2016)'s two-sided Fluctuation test.

RESULTS. Figures 2.1 and 2.2 plot $\tau_{GB,t}$ on the y-axis for the nowcast, the one-quarter-ahead forecast and an average of forecasts from two to four quarters ahead. Recall that the largest absolute value of $\tau_{GB,t}$ is the Fluctuation test statistic, $\mathcal{F}_{GB}$, and the timing reported on the x-axis is the mid-point of the rolling sample used to estimate $\tau_{GB,t}$ over time. The figure also reports the five percent critical value lines for the Information-Advantage Fluctuation test.

First, consider the results for inflation and real GDP growth reported in Figure 2.1. The figure shows that, for both variables, the Greenbook/Tealbook had an information advantage for the nowcast and one-quarter-ahead forecasts, which deteriorated in the early 2000s. At longer horizons (two-to-four-quarter-ahead quarter ahead), the information advantage is only sporadic and, for most of the sample, Greenbook/Tealbook forecasts do not seem to

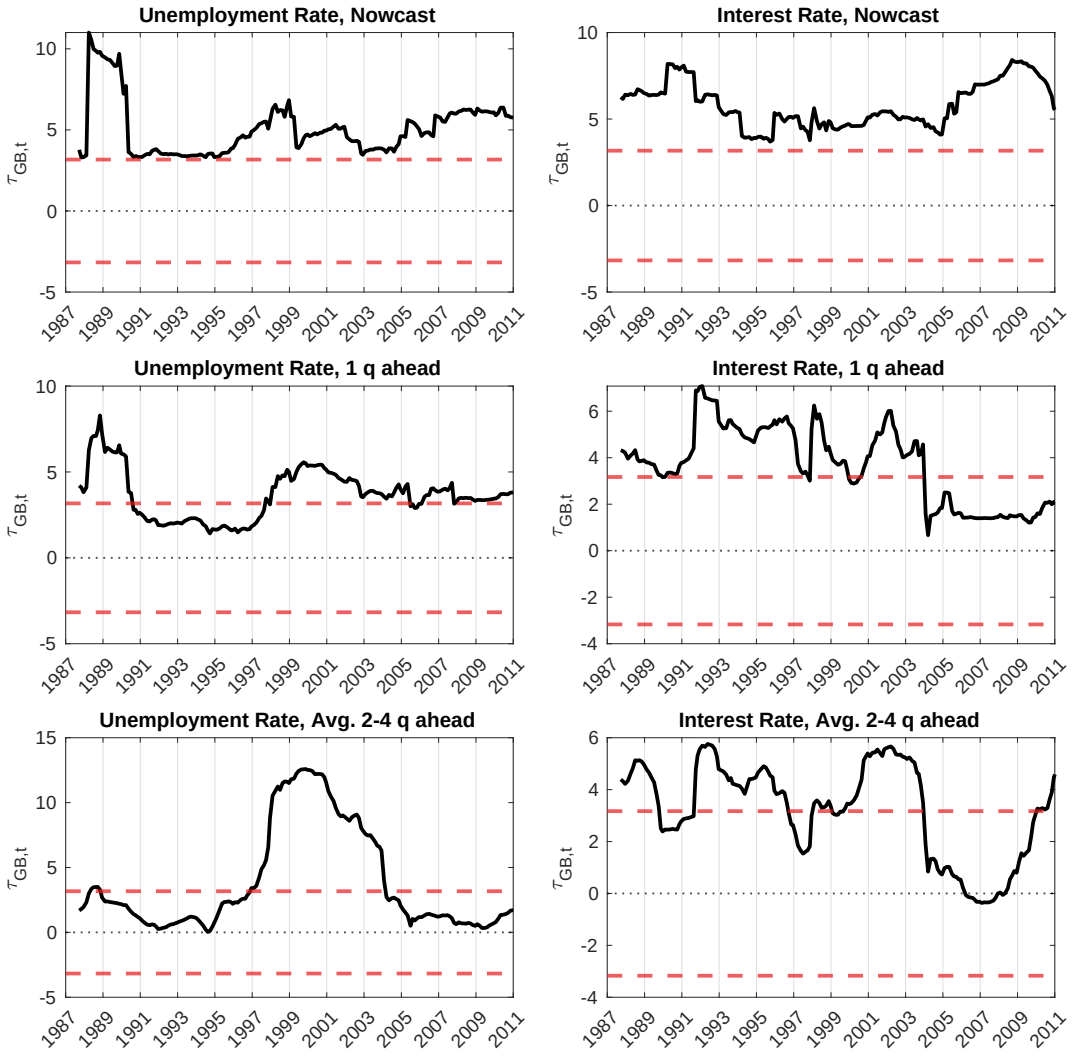


Figure 2.2: Information advantage fluctuation test: Unemployment and interest rates

Notes: The figure shows $\tau_{GB,t}$ from eq. (2.1) based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. Dashed (red) lines denote 5% critical value lines based on Rossi and Sekhposyan (2016)'s two-sided Fluctuation test.

provide additional information relative to the BCEI consensus forecasts. For unemployment and interest rate forecasts in Figure 2.2, the information advantage weakened substantially in one- and two-to-four-quarter-ahead forecasts, yet persists throughout the whole sample only for the nowcast.

REMARKS. Note that our regressions shed light on whether the central bank has an information advantage relative to survey participants from a historical point of view: finding that the central bank lost its information advantage in a given year does not imply that survey participants were aware of it in real time, as Greenbook/Tealbook forecasts become public with a delay of five years. However, the private sector might have been able to gauge the relative accuracy of the Greenbook/Tealbook forecasts in other (informal) ways. For

example, Ericsson (2016) shows that FOMC minutes contain useful information which can help infer the staff's Greenbook/Tealbook forecasts of real GDP growth rate years before the public release of these forecasts.

Finally, it is important to note that the Greenbook/Tealbook projections condition on a hypothetical, counterfactual policy path that is not supposed to be a monetary policy forecast. As discussed in Faust and Wright (2008), the conditional nature of the forecasts can be neglected when "the conditioning paths are not too far from the central bank's unconditional expectation for policy and/or that policy feedback is not too large over the relevant horizon". As discussed in Faust and Wright (2008), these assumptions may be reasonable at the very short forecast horizons we focus on. On the other hand, as information on the current interest rate and projections of monetary policy are readily available to private forecasters (e.g. through summaries/analyses/projections published by the FOMC), there is reason to believe that BCEI forecasts might be conditional forecasts as well. Berge et al. (2019) provide a framework to study the conditionality of survey forecasts and analyze both Greenbook/Tealbook and BCEI consensus forecasts. They report that interest rate projections were incorporated efficiently into both central banks' and private sector's forecasts of common macroeconomic variables, leading to the conclusion that both forecasts are conditional.

ROBUSTNESS. Appendix A investigates the robustness to the relative timing assumptions of Greenbook/Tealbook and BCEI forecasts. Specifically, we report Information-Advantage Fluctuation tests for the case where BCEI forecasts are always published before the corresponding Greenbook/Tealbook forecasts. The Not-for-Publication Appendix reports additional sensitivity analyses to different window sizes as well as the second and third vintages for the real-time realizations (instead of the first release). Our results remain robust to these changes.

2.2.2 Discrete breaks and changes in FOMC communication

The analysis in the previous section establishes that the information advantage of the Federal Reserve weakened in recent years. This section sharpens the evidence by testing for structural breaks and estimating break dates in the information advantage regressions.

While the Information-Advantage Fluctuation test robustly shows that there is little evidence to reject the hypothesis that β_{GB} is zero in recent years, the rolling-window nature of the test makes it difficult to precisely identify the exact point in time in which the departure from zero has taken place. In fact, the non-parametric approach we adopt is designed for smooth changes. When the changes are of a discrete nature, our approach may smooth them out over the rolling windows, thus making it difficult to identify exactly when the change happened. Therefore, we present complementary evidence based on Bai and Perron (1998)'s test, which is designed to identify multiple sharp breaks in parameters.

However, in the context of our analysis, the Bai and Perron (1998) test has several drawbacks. First, it requires that all the parameters change discretely and at the same time. The Fluctuation test, instead, is a non-parametric test that summarizes the time path of the parameter of interest (in this case the coefficient on the Greenbook/Tealbook forecast), while allowing other regression parameters to change (or not) in a data-driven way. Second, and most importantly, while the Bai and Perron (1998) test identifies multiple and discrete change-points, it cannot be used to test the joint hypothesis we are interested in, namely $\beta_{GB,t} = 0$ at every point in time.[9] Therefore, the Information-Advantage Fluctuation and the Bai and Perron (1998) tests complement each other: we report the latter here to shed additional light on the estimated break dates.

We conduct Bai and Perron (1998)'s test on $\beta_{GB,t}$ in the information advantage regressions in eq. (2.1).[10] Table 2.1 reports results for forecasts of inflation, real GDP growth, unemployment and the interest rate for several forecast horizons: the nowcast, one-quarter-ahead and the average over two-, three- and four-quarter-ahead forecasts. For each variable and forecast horizon, consistently with Bai and Perron (1998)'s notation, the table reports the $UDmax$ test statistic as well as the estimated number of breakpoints according to two criteria: the $\sup F(k+1|k)$ (denoted by k_{supF}) and the BIC (denoted by k_{BIC}). In most cases, the two criteria agree on the number of breakpoints, although there are some differences. We follow Bai and Perron (1998)'s recommendation and base our inference on the k_{supF} criterion. The last column in the table reports the estimated break dates, together with their 90 percent confidence intervals.

For inflation, there are broadly three break-points, dating to the early 1990s, 2002/2003, and 2008. For real GDP growth, there are breaks in the early 1990s and late 1990s/early 2000s for the one-quarter and the average of two-to-four-quarter-ahead forecasts. For the nowcast, on the other hand, the break date appears to be in 2010. For the unemployment rate, there are no detected breaks in the nowcast and one-quarter-ahead forecasts, consistent with the Fluctuation test results, while 2007 shows up as a break date at longer horizons. For the interest rate, 2007 is detected as a robust break-point across horizons; however, there are also breaks in 1992 and 2000 at the longer horizons.

As we emphasized earlier, Bai and Perron (1998) test for parameter stability, while the Fluctuation test jointly evaluates parameter stability and whether the parameter equals to zero. Nevertheless, the 2002/2003 break date is robustly detected by both tests. This date is related to a major change in the communication strategy of the FOMC, which in August

[9] For example, in the case that the true β_{GB} coefficient is constant but different from zero, the Fluctuation test will reject, while the Bai and Perron (1998) test will not reject as there is no instability.

[10] We follow the recommendations in Bai and Perron (2006) and Bai and Perron (2003), and use their $UDmax$ statistic with a maximum of $K = 5$ breaks. When the test rejects the null hypothesis of no break, we estimate the number of breaks by the sequence of $\sup F(k+1|k)$ tests and estimate the break dates by globally minimizing the sum of squared residuals in eq. (2.1). Confidence intervals are constructed based on the asymptotic approach provided in Bai and Perron (1998), assuming serially correlated, but homogeneous residuals across segments.

Table 2.1: Results from multiple break tests

Horizon	UDmax	k_{supF}	k_{BIC}	Break date [90 % CI]
GDP Deflator Inflation				
Nowcast	11.99	3	3	10/1991 [10/1990 - 07/1992] 12/2002 [08/2000 - 08/2005] 09/2008 [12/2006 - 11/2009]
1 q ahead	12.16	3	3	03/1991 [05/1990 - 12/1991] 09/2003 [06/2001 - 09/2005] 06/2008 [05/2006 - 11/2009]
Avg. 2-4 q ahead	19.53	3	4	10/1990 [05/1989 - 08/1991] 03/2003 [11/1995 - 08/2008] 03/2008 [08/2003 - 04/2011]
GDP Growth				
Nowcast	13.17	1	0	03/2010 [03/2005 - 10/2013]
1 q ahead	17.98	2	2	07/1993 [10/1991 - 03/1995] 03/2000 [11/1998 - 12/2001]
Avg. 2-4 q ahead	13.95	2	2	10/1992 [02/1991 - 02/1994] 10/1999 [11/1998 - 08/2001]
Unemployment Rate				
Nowcast	4.44	0	0	
1 q ahead	4.11	0	3	
Avg. 2-4 q ahead	22.19	1	3	03/2007 [07/1990 - 12/2015]
Interest Rate				
Nowcast	84.25	1	1	06/2007 [06/2005 - 09/2007]
1 q ahead	89.76	1	1	06/2007 [02/2005 - 09/2007]
Avg. 2-4 q ahead	238.65	3	3	08/1992 [08/1984 - 07/1996] 02/2000 [02/1989 - 06/2006] 03/2007 [08/1992 - 12/2009]

Notes: The trimming parameter is 0.15 and the maximum number of potential breaks is five. The HAC covariance is estimated based on Andrews (1991)'s AR(1) bandwidth selection (no prewhitening). The break dates are based on k_{supF}.

2003 started including time-dependent forward guidance in its post-meeting statement. This is an important break-point, documented in several studies (see Lunsford 2020).

The break dates in the early 1990s overall appear to coincide with many changes introduced by the FOMC in its communication strategy, starting with the decision to publish minutes in March 1993 and subsequently issuing statements following every meeting in May 1999. Bai and Perron (1998)'s tests also detect break-dates in the period between 2007 to 2009. These appear to be related to the release of the quarterly Summary of Economic Projections (SEP) in November 2007 (which reported ranges and central tendencies of participants' forecasts for up to three years ahead) and the subsequent quarterly press conferences related to the SEP in April 2011.

2.2.3 Relationship to forecast accuracy

In addition to information advantage regressions, one could consider other, potentially different test statistics. In this section, we provide complementary evidence based on MSFE comparisons. Just as a preview, the results confirm our main finding: the forecasting performance is time-varying, and Greenbook/Tealbook forecasts were not significantly more accurate than the BCEI in the latest part of the sample.

It is important to clarify at the onset that information advantage regressions and MSFE comparisons are two very different ways of comparing forecasts: in general, finding that a forecast has an information advantage over a competitor does not imply that the former has a lower MSFE. In fact, information advantage regressions investigate whether a forecaster that has access to both Greenbook/Tealbook as well as BCEI forecasts will use both or will prefer only one of them. Hence, there is a tight link between information advantage regressions and the forecast combination literature. As we discuss in Section II.A in the Not-for-Publication Appendix (see also Winkler and Clemen, 1992), the optimal forecast combination (based on the MSFE measure of accuracy) weights each forecast proportionally to its forecast accuracy when the underlying forecasts are unbiased and uncorrelated. When forecasts are correlated, instead, the most accurate model still gets a higher weight in the combination, yet the weight does not reflect its accuracy because it is "distorted" by the correlation. In addition, when the forecasts are biased, it is possible that, depending on the direction and magnitude of the biases, the most accurate forecast in terms of MSFE gets a lower weight than the less accurate forecast. These insights suggest caution in interpreting the magnitude of the coefficients (that is, the weights) in the information advantage regression (see also Sims, 2002), as they cannot always be interpreted as a measure of forecast accuracy.

As the discrepancy between the forecast advantage and the relative MSFE measures may depend on the correlation between the forecasts as well as their bias, we empirically investigate them in our data. Section II.B in the Not-for-Publication Appendix shows the existence of a time-varying (and, at times, strong) cross-correlation between the central bank

and private sector forecasts. The cross-correlations are particularly large for the nowcasts of the unemployment rate, real GDP growth and interest rates and have been growing over time for the latter two. In addition, Sections II.C and II.D in the Not-for-Publication Appendix show that both Greenbook/Tealbook forecasts as well as the BCEI exhibit time-varying biases and failures of rationality (see also Rossi and Sekhposyan, 2016).

Given this evidence, we complement our information advantage regressions with a time-varying analysis of forecast accuracy. Figures 2.3 and 2.4 depict rolling estimates of relative predictive accuracy, measured by the difference between the MSFEs of the BCEI consensus forecasts and the Greenbook/Tealbook forecasts, scaled by its standard deviation (labeled $\tau_{GR,t}$). We implement a Fluctuation test in a regression similar to that in eq. (2.1), where the left-hand side is the MSFE difference and the only regressor is the constant. The null hypothesis is that the Greenbook/Tealbook and the BCEI forecasts have the same predictive accuracy; under the alternative, positive values of the test statistic indicate that the Greenbook/Tealbook predictive performance is more accurate. The dashed (red) line indicates the five percent critical value of the Giacomini and Rossi (2010) test.

The figures show that the equal predictive accuracy of these forecasts is rejected for all the variables and all horizons. The time path of the test statistic indicates that the forecast accuracy of Greenbook/Tealbook in predicting inflation worsened in the early 1990s for the nowcast and the one-quarter-ahead forecasts, while the deterioration dates to the late 1990s for the average two-to-four-quarter-ahead forecasts. For the real GDP growth rate, the central bank either had no comparative advantage in forecasting or, in the case of the nowcast, the information advantage disappeared in the late 1990s. For unemployment, the performance of Greenbook/Tealbook and BCEI forecasts is broadly the same, with some sporadic outperformance of Greenbook/Tealbook at the two-to-four-quarter-ahead horizon. In terms of the interest rate, the central bank appears to have more accurate forecasts only for the nowcast and that advantage disappears around 2003. Thus, even based on the relative forecast accuracy criterion, either the central bank had no forecasting advantage or, in some cases, the advantage disappeared even earlier than our estimates suggest, namely in the 1990s.

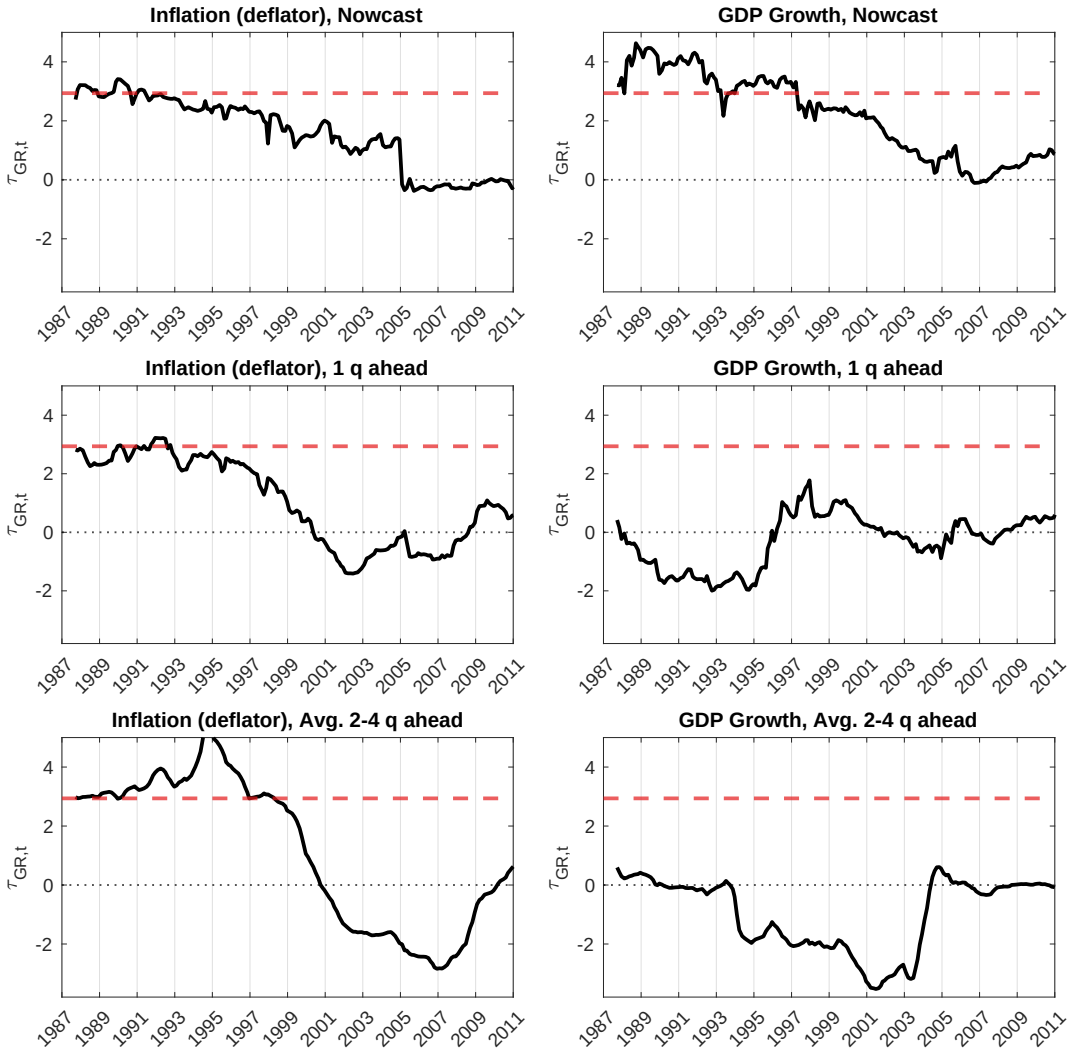


Figure 2.3: Forecasting performance fluctuation test: GDP growth and inflation

Notes: The figure shows $\tau_{GR,t}$ based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. Dashed (red) lines denote 5% critical value lines based on Giacomini and Rossi (2010)'s one-sided Fluctuation test.

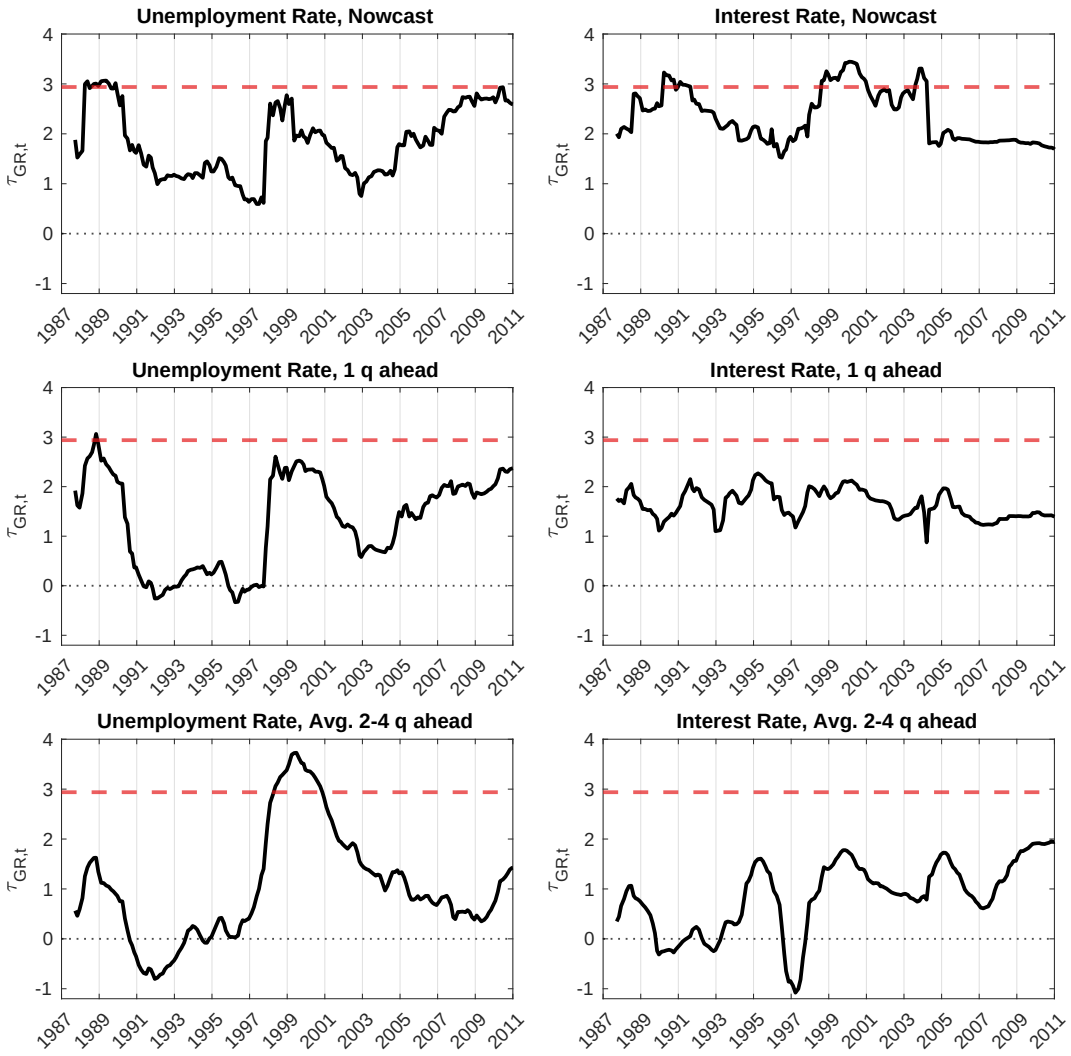


Figure 2.4: Forecasting performance fluctuation test: Unemployment and interest rates

Notes: The figure shows $\tau_{GR,t}$ based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. Dashed (red) lines denote 5% critical value lines based on Giacomini and Rossi (2010)'s one-sided Fluctuation test.

2.3 Do monetary policy surprises contain information effects?

The previous section provided empirical evidence that the Federal Reserve had an information advantage in predicting key macroeconomic variables historically, but lost such an advantage recently. A related and important question is whether the importance of the information channel of monetary policy has similarly decreased over time. As discussed in the introduction, the existence of the central bank information advantage is a sufficient, but not a necessary condition for the presence of the information channel of monetary policy. This is because the private sector might update its expectations on the state of the economy even if the central bank does not have an information advantage (see e.g. Morris and Shin, 2002). When studying the reaction of the economy to information effects, it is therefore important to recognize that potential time-variation in information effects might exist independently from the results we already established in Section 2.2. Therefore, we investigate whether high-frequency market-based surprises can be explained by the Federal Reserve's economic outlook and whether this information content has changed over time.

Similarly to our information advantage analysis in Section 2.2, we find that, while high-frequency surprises can be historically predicted by Federal Reserve staff forecasts, this relationship has become insignificant in the recent period. Building on this finding, we then construct an updated version of Miranda-Agrippino and Ricco (2020)'s policy instrument, which is a monetary policy surprise cleaned from information available to the Federal Reserve. We explicitly take time variation into account when constructing the instrument, which is used in Section 2.4 to estimate the response of the economy to monetary policy shocks as well as in Section 2.5 to establish whether professional forecasters revise their forecasts in response to monetary policy.

2.3.1 The information content of market surprises

We start by investigating the information content of high-frequency market-based surprises identified in a short window of time around FOMC announcements. Our main objective is to establish whether market surprises are predictable by the information available to the Federal Reserve staff and whether this predictability has changed over time. High-frequency surprises are widely used in the literature on information effects and several papers highlight the importance of controlling for the Fed's private information. For example, Romer and Romer (2004) show that Greenbook/Tealbook forecasts significantly predict changes in the intended Fed Funds Rate around FOMC meetings. Campbell et al. (2017) demonstrate that the "Delphic component" of high-frequency surprises (i.e. the component that reflects the Federal Reserve's private information on current and future macroeconomic conditions) can explain the puzzling decrease in unemployment survey expectations after a monetary policy tightening (Campbell et al., 2012). Jarociński and Karadi (2020) construct an instrument for monetary policy shocks by controlling for the central bank's assessment of the economic outlook, revealed by stock market surprises. Finally, Miranda-Agrippino

and Ricco (2020) construct a monetary policy shock instrument that controls for the Federal Reserve information by extending the approach of Romer and Romer (2004) to high-frequency-identified market surprises; they show that their instrument accounts for information effects across a large number of Structural VAR (SVAR) specifications. In our analysis, we follow Miranda-Agrippino and Ricco (2020)'s and Romer and Romer (2004)'s approach and study the correlation of high-frequency market-based monetary surprises with the Federal Reserve's internal forecasts at different horizons. Specifically, we project surprises in the three month Federal Funds Futures Rate ($FF4$) on Greenbook/Tealbook forecasts of real GDP growth, GDP deflator inflation and the unemployment rate as well as their revisions for different forecast horizons, via the following regression:

$$\mathbb{S}_t = \alpha^{(h)} + \sum_j \theta_j^{(h)\prime} \begin{pmatrix} F_t^{GB}(x_{j,q+h}) \\ F_t^{GB}(x_{j,q+h}) - F_{t-1}^{GB}(x_{j,q+h}) \end{pmatrix} + \varepsilon_t^{(h)} \tag{2.4}$$

where $\mathbb{S}_t$ is the high-frequency market-based surprise, $F_t^{GB}(x_{j,q+h})$ denotes the h-quarter-ahead Greenbook/Tealbook forecast of variable x_j associated with the FOMC meeting at time t; $F_t^{GB}(x_{j,q+h}) - F_{t-1}^{GB}(x_{j,q+h})$ denotes the forecast revision; and $\theta_j^{(h)}$ collects the coefficients associated with the forecasts and forecast revisions of variable x_j, where j denotes the variable to be forecasted (j = GDP growth, inflation and unemployment). We estimate eq. (2.4) for one quarter backcasts, nowcasts as well as one- and two-quarter-ahead forecasts (i.e. $h = -1, 0, 1, 2$).

Miranda-Agrippino and Ricco (2020) find, in a similar specification over the full sample, that the null hypothesis of no correlation between the market surprises and Fed forecasts ($\theta_j^{(h)} = 0$) can be rejected. They interpret this result as evidence of an information channel via FOMC announcements. Given our findings from Section 2.2, however, we are interested instead in studying how this correlation evolves over time. Therefore, similarly to the approach in Section 2.2, we define the Fluctuation test statistic for the regression in eq. (2.4) as:

$$\mathcal{F}_{FED} = \max_t |W_t|, \tag{2.5}$$

with

$$W_t = m\ \hat{\theta}_t^{(h)\prime}\ \left\{\hat{V}_\theta^{(h)}\right\}^{-1}\ \hat{\theta}_t^{(h)}, \quad \text{for } t = m/2, \ldots, T - m/2, \tag{2.6}$$

where $\hat{\theta}_t^{(h)}$ and $\hat{V}_\theta^{(h)}$ are computed in rolling windows of m observations.

DATA. To implement the regression in eq. (2.4), we choose as our baseline measure of interest rate surprises the change in the three month Fed Funds Futures in a half-hour window starting 10 minutes before and ending 20 minutes after the announcement.[11] This

[11] FF4 contracts exchange a constant interest for the average Federal Funds Rate over the course of the third calendar month. In most of our sample, regular policy meetings are spaced roughly six weeks apart. Therefore, the three month futures rate can be interpreted as the shift in the expected Federal Funds Rate

surprise measure was used by Gertler and Karadi (2015), Jarociński and Karadi (2020), Paul (2020) and Miranda-Agrippino and Ricco (2020). We study surprises around 234 FOMC meetings from February 1990 to December 2015 using an updated version of the Gürkaynak et al. (2005) dataset. While market-based monetary surprises are also available for more recent dates, our dataset is constrained by the availability of Greenbook/Tealbook forecasts, which are only released with a five-year lag.[12] The dataset contains both scheduled as well as unscheduled FOMC meetings and other important announcements. We associate each Greenbook/Tealbook forecast to the relevant FOMC announcement. For scheduled meetings, these forecasts have a direct mapping to the announcements, as they were prepared specifically for the respective FOMC meeting. For unscheduled announcements, we use the latest available Greenbook/Tealbook forecast made before the announcement and correct the forecast horizon when the target quarter of the forecasts changed. We then compute the revision of each Greenbook/Tealbook forecast as the difference between the forecast associated with the current FOMC meeting and the previous meeting, correcting the forecast horizon of the earlier forecast when necessary.[13]

RESULTS. Figure 2.5 reports W_t for the regression in eq. (2.4) for the one-quarter backcast, the nowcast, as well as the one- and two-quarter-ahead forecasts together with the 5% critical value line for the Fluctuation test. The largest (absolute) value of W_t is the Fluctuation statistic, $\mathcal{F}_{FED}$. When $\mathcal{F}_{FED}$ is above the critical value line, the test rejects the null hypothesis that market surprises were never predictable by the Greenbook/Tealbook forecasts. As in the previous sections, the timing reported on the x-axis is the mid-point of the rolling sample used to estimate W_t over time. The figure illustrates that high-frequency market-based surprises were significantly predictable by the Fed staff before the mid-2000s, but that the predictability disappeared in the most recent period, regardless of the forecast horizon. Importantly, the results hold for both scheduled as well as unscheduled announcements. Thus, it is important to account for information effects in the first part of the sample, but less so in the most recent period.[14]

2.3.2 An information-robust instrument of monetary policy

In the next two sections, we study the impact of monetary policy announcements on the macroeconomy and forecasters' expectations using an information-robust instrument. Given our previous findings, namely that accounting for the Federal Reserve information is more important in the earlier than in the later part of the sample, we modify their instrument

following the next policy meeting.

[12] Greenbook/Tealbook forecasts are currently available up to December 2015.

[13] Note that by definition, the forecast revision associated with unscheduled FOMC meetings is zero as the forecasts have not been updated since the last scheduled FOMC announcement.

[14] In addition to regressions at individual horizons, Miranda-Agrippino and Ricco (2020) also consider F-tests in a regression including all variables and horizons, following the specification in Romer and Romer (2004). Table B.4 in the Appendix replicates the full-sample results in Miranda-Agrippino and Ricco (2020) and shows that our conclusions are robust.

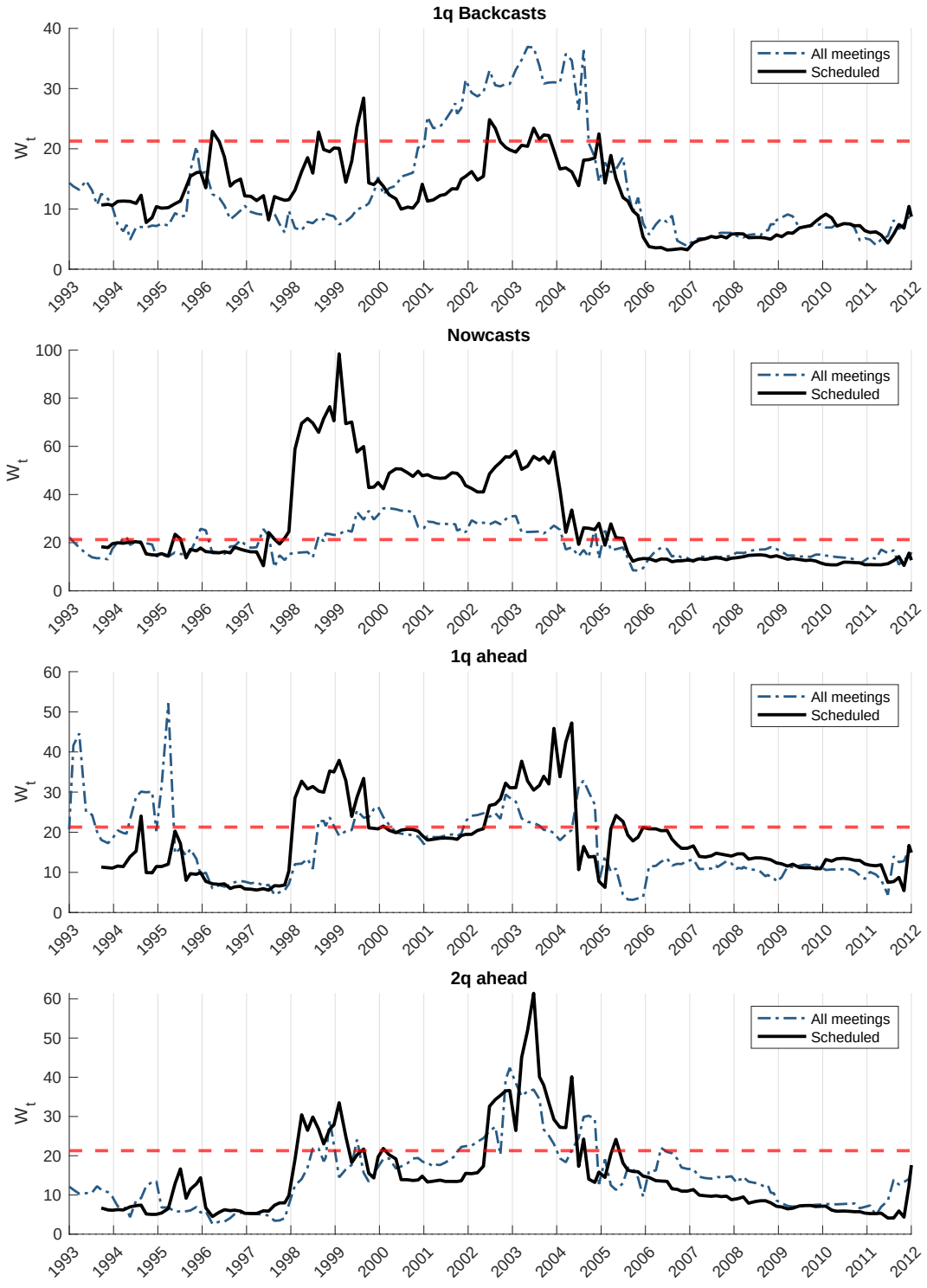


Figure 2.5: Information content of market-based monetary surprises

Notes: The figure shows W_t from eq. (2.6) based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. The dashed (red) line denotes the 5% critical value based on Rossi and Sekhposyan (2016)'s Fluctuation test.

series by: (i) extending the sample to the latest available Greenbook/Tealbook data;[15] and (ii) taking the time-variation in the information content of surprises into account by estimating the instrument separately in the relevant sub-samples.

To construct the informationally-robust instrument for monetary policy shocks, we closely follow Miranda-Agrippino and Ricco (2020). First, to control for the central bank's private information, we project the FF4 surprises ($\mathbb{S}_t$) on Greenbook/Tealbook forecasts as well as their revisions for macroeconomic variables at the meeting-level frequency:

$$\begin{aligned}\mathbb{S}_t &= \alpha + \sum_{h=-1}^{3} \theta_j' F_t^{GB}(x_{q+h}) + \sum_{h=-1}^{2} \delta_j' \left[F_t^{GB}(x_{q+h}) - F_{t-1}^{GB}(x_{q+h})\right] + \mathbb{S}_t^{MPI} \\ &= \mathbb{S}_t^{CBINFO} + \mathbb{S}_t^{MPI},\end{aligned} \tag{2.7}$$

where $F_t^{GB}(x_{q+h})$ is a vector containing the central bank's forecasts of output, inflation and unemployment. The residual of this projection ($\mathbb{S}_t^{MPI}$) is the monetary policy shock "cleaned" from the Federal Reserve's information on the economic outlook, whereas its orthogonal component ($\mathbb{S}_t^{CBINFO}$) measures the Federal Reserve's own information. Motivated by our analysis from the previous sub-section, we explicitly take instabilities into account by separately estimating the regression above in the two sub-samples before and after August 2003.

Second, we aggregate the resulting meeting-level series to a monthly frequency, as the meeting-level series are irregularly spaced and the analyses in Sections 2.4 and 2.5 are conducted at the monthly frequency. Therefore, as in Miranda-Agrippino and Ricco (2020), we transform $\mathbb{S}_m^{MPI}$ and $\mathbb{S}_m^{CBINFO}$ to monthly series by summing the individual surprises occurring in each month and setting them to zero in months in which there is no FOMC announcement.[16]

Figure 2.6 reports our updated instrument series. The top panel shows the FF4 surprises, $\mathbb{S}_t$, at the monthly frequency while the bottom panel shows their decomposition into the information robust monetary policy instrument ($\mathbb{S}_t^{MPI}$) and the central bank's information shock ($\mathbb{S}_t^{CBINFO}$). The (red) vertical line in the bottom panel separates the two sub-samples. The correlation with the original MPI shock of Miranda-Agrippino and Ricco (2020) is 0.912 for the full sample, 0.936 for the first sub-sample and 0.868 for the second sub-sample. In addition, our series preserves almost all of the large realizations of their original shock.[17]

[15] The original Miranda-Agrippino and Ricco (2020) series is available from February 1990 to December 2009 whereas, at the time of writing this paper, Greenbook/Tealbook forecasts are available until December 2015.

[16] Miranda-Agrippino and Ricco (2020) also adjust the resulting monthly instrument series to account for potential serial correlation by estimating an $AR(12)$ on the $\mathbb{S}_t^{MPI}$ series. As we construct our instrument in two sub-samples, to mitigate small sample concerns we rely on the BIC criterion to select the lag length for the $AR(p)$ process. The BIC selects $\hat{p} = 0$ for the $\mathbb{S}_t^{MPI}$ series, which is what we use to obtain our information-robust instrument. The results are robust if the lag length were selected in individual sub-samples.

[17] Note that the original shock is only available until December 2009 so that the second sub-sample only partially overlaps with Miranda-Agrippino and Ricco (2020)'s original sample.

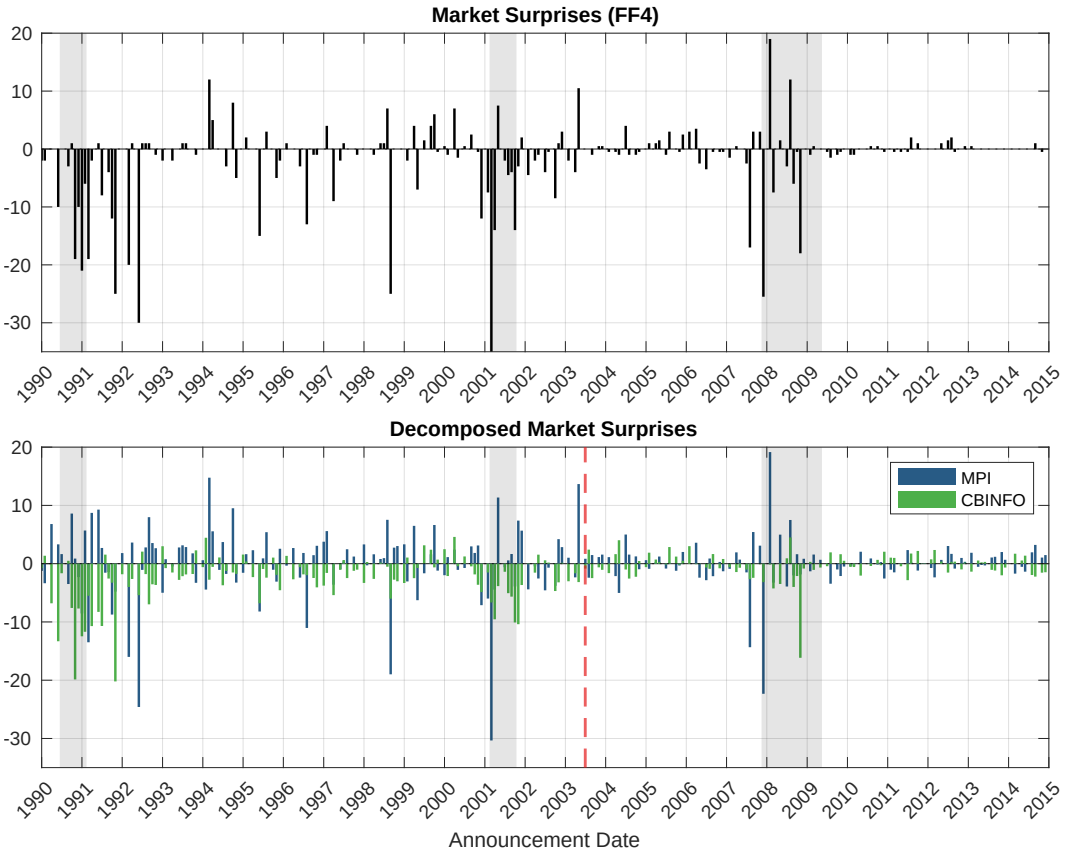


Figure 2.6: Contributions to the surprises in the three-month Fed Funds Futures

Notes: High-frequency market-based surprises are aggregated to monthly frequency, expressed in basis points. The decomposition is based on eq. (2.7).

2.4 The impact of information effects on the macroeconomy

After having established that the information content of market surprises has disappeared over time, we next turn to assessing whether the information channel played a role in the transmission of monetary policy in the U.S. and how its importance evolved over time. According to the information channel theory (Nakamura and Steinsson, 2018), in the presence of informational rigidities, informationally constrained private agents could infer from an increase in the central bank's policy rate not only that the central bank is deviating from its rule, but also that it is endogenously responding to stronger than expected future fundamentals. If the latter component is not correctly taken into account, the estimated responses to the monetary policy shock could potentially mix the response to the actual monetary policy shock and to the signal about the future state of the economy. Several papers have shown that impulse responses to high-frequency market-based surprises can be contaminated by information effects and that model-consistent impulse responses can be obtained by using an instrument that controls for the information channel (see Miranda-Agrippino and Ricco, 2020 and Jarociński and Karadi, 2020).

To investigate the information channel of monetary policy, we study impulse responses of several key macroeconomic aggregates to a monetary policy shock using a SVAR with high-frequency instruments. We consider both high-frequency surprises in the three month Federal Funds Futures Rate (FF4), as well as our updated version of the information-robust instrument discussed in Section 2.3. Since FF4 market surprises do not control for information effects, the implied impulse responses reflect both the effect of the change in the policy rate as well as any potential response of the economy to information effects. In contrast, the impulse responses obtained using the information-robust instrument control for the information set of the central bank, and therefore only reflect changes in the policy rate.[18] The only difference between the two impulse responses is thus whether we control for information effects or not. By comparing the impulse responses in sub-samples, we can thus assess how the information effect has changed in the recent period relative to the earlier part of the sample.

2.4.1 The VAR model

We estimate a six-variable SVAR model where the vector of endogenous variables are the industrial production index, the unemployment rate, the consumer price index, the commodity price index, the excess bond premium by Gilchrist and Zakrajšek (2012) and the one year nominal policy rate.[19] This is the same SVAR used in Miranda-Agrippino and Ricco (2020) and it is similar to that in Coibion (2012) and Gertler and Karadi (2015). Details on the data series and their sources can be found in the Not-for-Publication Appendix. All variables are monthly from January 1979 to December 2019. The SVAR is estimated in (log) levels with 12 lags. As discussed above, the impulse responses are identified using two external instruments: (i) the FF4 surprises ($\mathbb{S}_t$) and (ii) our updated version of the information-robust instrument of Miranda-Agrippino and Ricco (2020), $\mathbb{S}_t^{MPI}$. In both cases, the impulse responses are normalized such that the monetary policy shock increases the policy rate by one percent on impact. As discussed in the previous section, due to the lagged release of the Greenbook/Tealbook forecasts, the external instrument series is only available from February 1990 to December 2015. The impact responses in the SVAR are therefore identified from a proxy regression over the common sub-sample, February 1990 to December 2015. The VAR is estimated with standard Bayesian Normal Inverse-Wishart priors and the tightness of the prior is set as in Giannone et al. (2015).

2.4.2 The role of information effects

Figure 2.7 shows the BVAR impulse responses identified using: (i) the FF4 surprises ($\mathbb{S}_t$, dashed blue line) and (ii) the information-robust series ($\mathbb{S}_t^{MPI}$, solid black line). We report

[18] Because the FF4 horizon is three months, it can also capture some near-term forward guidance in addition to policy rate changes. However, it also mitigates the effect of the zero lower bound.

[19] Both the commodity price index as well as the one year nominal rate series are end-of-month values.

impulse responses for two sub-samples: January 1979 - July 2003 and August 2003 - December 2019.[20]

First, focus on the sub-sample from January 1979 - July 2003 (left panel of Figure 2.7). Using the FF4 market surprises as instruments, in response to a contractionary monetary policy shock industrial production increases and unemployment decreases. However, using the information-robust instrument, we recover impulse responses that are consistent with economic theory: output decreases and unemployment increases. The large discrepancy between the responses thus corresponds to the economy's reaction to the information effects. These findings are similar to Miranda-Agrippino and Ricco (2020) for their full sample (1979 - 2014).

However, these conclusions change when we consider the second sub-sample, August 2003 - December 2019 (right panel in Figure 2.7). In this case, the responses based on the FF4 surprises are more in line with economic theory than in the earlier sample: output decreases and unemployment increases in the short run. Overall, the discrepancy between the impulse responses based on FF4 and those based on the information-robust instrument is negligible in the later part of the sample, while it was substantial in the first sub-sample. Thus, we conclude that, while information effects were important historically, they are much less important in the most recent period.

ROBUSTNESS. Figure C.11 in the Appendix shows that our conclusions are robust to comparing the information-robust instrument ($\mathbb{S}_t^{MPI}$) with the associated information component, ($\mathbb{S}_t^{CBINFO}$). We also address potential misspecification concerns by repeating the analysis using a local projection approach. If the VAR is correctly specified, local projections are less efficient than BVARs but have the advantage of being robust against dynamic misspecification. Figure C.12 in the Appendix shows that, although the local projection-based response bands are much larger than BVAR-based ones, our main conclusions remain robust. Finally, the Not-for-Publication Appendix shows that our conclusions remain robust regardless of whether we estimate $\mathbb{S}_t^{MPI}$ using scheduled meetings only or if we remove potential serial correlation in the shock series via an AR(12). We also show that, in our full sample, the responses based on our updated instrument are consistent with the full sample results in Miranda-Agrippino and Ricco (2020).

[20] Specifically, note that while the impact parameters are identified in each sub-sample, the lag parameters are estimated on the full sample to improve the efficiency of the parameter estimates. Our results are robust to re-estimating the SVAR in each sub-sample separately.

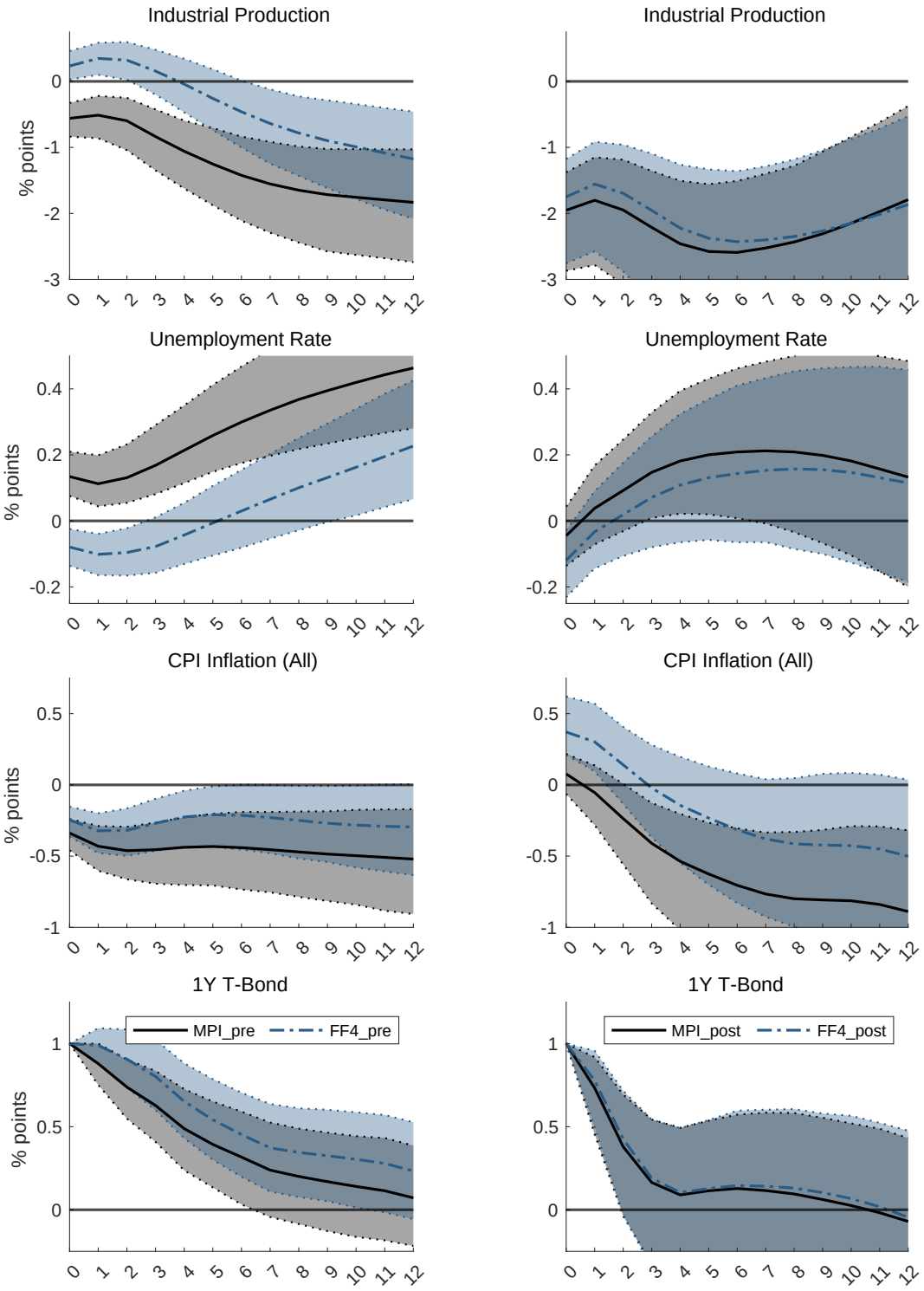


Figure 2.7: Responses to a monetary shock: Sub-samples

Notes: Impulse responses from Bayesian SVAR with standard macroeconomic priors and external instrument identification. VAR sample: January 1979 - December 2019. Instrument samples: February 1990 - July 2003 (left panel) and August 2003 - December 2015 (right panel). Shaded areas correspond to 90 percent credible intervals.

2.5 The impact of information effects on forecasters' expectations

Finally, we characterize the response of private forecasters to monetary policy announcements. As is customary in the literature, we regress monthly revisions in the BCEI consensus forecasts on a series of high-frequency market-based surprises in Federal Funds Futures Rates around FOMC meetings. Using a similar regression with monetary policy surprises that reflect both current and future short-term rates, Nakamura and Steinsson (2018) find that the way survey forecasters update their predictions is inconsistent with standard New Keynesian models. Their results support the existence of an information channel according to which professional forecasters believe that FOMC policy surprises contain useful and otherwise unavailable information to the public. Campbell et al. (2012) come to a similar conclusion by decomposing monetary policy surprises into orthogonal "target" and "path" components. Campbell et al. (2017) show that the puzzling signs can be explained by decomposing the surprises into a "Delphic" component, associated with information releases by the central bank, and an orthogonal ("Odyssean") component: the signs remain puzzling for the former while are consistent with economic theory for the latter. Finally, Paul (2020) shows that there is little evidence of information releases in short horizon Federal Funds Futures Rate when focusing on scheduled FOMC meetings, while including unscheduled meetings leads to significant forecast revisions.

In line with the existing literature, we consider the following regression to analyze whether forecasters revise their expectations after a monetary policy announcement:

$$\Delta x_{t+h|t}^{BCEI} = \alpha + \beta \, \mathbb{S}_t + \varepsilon_{t+h} \tag{2.8}$$

where $\Delta x_{t+h|t}^{BCEI}$ denotes the BCEI consensus forecasts revision at horizon h between FOMC meetings and $\mathbb{S}_t$ is the FF4 surprise. We also consider an alternative specification where $\mathbb{S}_t$ is the surprise in 30 day Federal Funds Futures Rates (MP1), used in Paul (2020) and Lunsford (2020). Furthermore, to account for potential information effects in monetary policy announcements, we decompose the surprises, $\mathbb{S}_t$, into the robust monetary policy shock and the information shock described in Section 2.3, and estimate the following regression:

$$\Delta x_{t+h|t}^{BCEI} = \alpha + \beta_1 \, \mathbb{S}_t^{MPI} + \beta_2 \, \mathbb{S}_t^{CBINFO} + \eta_{t+h} \tag{2.9}$$

where $\mathbb{S}_t^{MPI}$ is the surprise component "cleaned" from the Fed information and $\mathbb{S}_t^{CBINFO}$ is the information component (see eq. 2.7). We analyze this specification both in the full sample as well as in the two sub-samples identified in the previous sections.

DATA. To implement the regressions in eqs. (2.8) and (2.9), we focus on the same sample of monetary policy announcements that we used to analyze the information content

of monetary policy surprises in Section 2.3. To study the forecasters' response, we augment this dataset with a measure of revisions to the private sector forecasts which we calculate for each meeting, variable and forecast horizon as the difference between the BCEI forecasts bracketing each FOMC announcement.[21] As in Nakamura and Steinsson (2018) and Paul (2020), we drop those FOMC announcements in which the meeting date falls into the BCEI survey period (first three business days of each month until December 2000; first two business days of each month after December 2000) to ensure that the FOMC announcements are after the BCEI survey dates.[22] Further, as in Nakamura and Steinsson (2018), we focus on scheduled FOMC meetings only. This is consistent with our analysis of the information advantage from Section 2.2. Appendix D reports results for the sample which includes unscheduled FOMC meetings.

RESULTS. Table 2.2 shows the results based on the full sample (February 1990 - December 2015). When considering FF4 surprises (first column), the sign of the coefficient β in eq. (2.8) is inconsistent with the responses of real GDP growth forecasts in the New Keynesian model across all forecast horizons. In contrast, for inflation, unemployment and the interest rate, all coefficients except for the one-quarter-ahead responses of inflation and the unemployment rate exhibit signs consistent with economic theory. Across all regressions, none of the coefficients are statistically significant.[23] Hence, based on the full sample, there is no evidence that regularly scheduled monetary policy announcements lead to significant forecast revisions by private forecasters. This evidence is consistent with the results in Paul (2020), who considers the regression in eq. (2.8) on a similar sample. In fact, by analyzing the average of the current quarter to four-quarter-ahead responses of the BCEI to FF4 surprises, Paul (2020) also finds a positive sign for the real GDP response and negative signs for inflation and the unemployment rate and no evidence of a statistically significant reaction based on scheduled meetings (see his Appendix A.10). Our results are robust to using Paul (2020)'s shorter 30-day Federal Funds Futures surprises (MP1, see the last column of Table 2.2).

Next, we repeat the analysis using the surprises decomposed as in eq. (2.9). A comparison of the second and third columns of Table 2.2 with the first column highlights two important results. First, accounting for the Federal Reserve's information content corrects some of the puzzling signs: the effects of the robust (MPI) surprises, particularly at the one-quarter-

[21] Specifically, we use the same procedure as Nakamura and Steinsson (2018) and compute the forecast revision as the difference between the BCEI forecast from the month following the FOMC announcement and the forecast which falls in the same month as the FOMC announcement. Since BCEI forecasts are collected at the beginning of each month, the latter typically falls before the announcement. When the target dates of the two forecasts used to calculate the revision are different, we use the previous BCEI forecast and adjust the forecast horizon to keep the target date fixed.

[22] The strategy adopted in Nakamura and Steinsson (2018) and Paul (2020) is slightly different as they drop all FOMC meetings occuring in the first week of each month whereas we drop only those meetings occurring during the BCEI survey period, leading to a slightly larger sample. The results presented in this section are robust to adopting their strategy.

[23] Here we use robust standard errors since the left-hand side variable is a forecast revision, which is not correlated over time.

ahead horizon, are more consistent with economic theory than those of FF4 surprises and yet none are significant. Second, accounting for the information content of the surprises reveals that survey participants react to the information component of FOMC announcements (CBINFO) with the expected signs and the responses are large in magnitude and statistically significant for GDP and inflation at short horizons.

Finally, Table 2.3 reports the forecasters' response in the same sub-samples considered in Section 2.4: February 1990 - July 2003 and August 2003 - December 2015. By comparing the coefficients on the FF4 surprises in both sub-samples with the ones for the full sample (reported in Table 2.2), we note that the coefficient signs and magnitudes in each of the sub-samples are consistent with the results obtained for the full sample. Similarly, there is no evidence of a significant reaction in any of the sub-samples. These results continue to hold for the surprises cleaned from the Federal Reserve information (MPI). In fact, there is only mild evidence that survey participants react to the discretionary component of monetary policy shocks: apart from the GDP growth response in the first sub-sample, which is significant only at a 10% level, there is no significance in either sub-sample. Similarly, the MPI responses mostly have the expected signs across the two sub-samples.[24] A striking difference between the two sub-samples is the response of interest rates' survey forecasts to the central bank's information component in monetary surprises (CBINFO): the response is highly significant in the first sub-sample and insignificant in the second one. Again, this evidence suggests that the information channel was relevant prior to 2003 but weakened substantially after that.

ROBUSTNESS. We explore the robustness of our findings to including unscheduled FOMC meetings. Such meetings may be more likely to be associated with the release of a central bank's private information since they often take place as a reaction to important economic events. Tables D.5 and D.6 in the Appendix report results from including unscheduled meetings in the full sample and sub-sample regressions, respectively. Most of our results are robust to this change. The most important difference is that, in both the full sample as well as in the earlier sub-sample, the inclusion of unscheduled meetings leads to a significant reaction of interest rate forecasts to all surprises as well as a significant reaction of short horizon GDP growth forecasts in the earlier sub-sample. This is consistent with the findings in Paul (2020), according to whom the inclusion of unscheduled meetings leads to statistically significant coefficients while excluding unscheduled meetings leads to insignificant responses. Importantly, as Table D.6 shows, the significant reaction disappears in the later sub-sample, consistently with the results presented earlier in our paper.

[24] There is one exception: the unemployment response at short horizons in the second sub-sample.

Table 2.2: Forecasters' response - Full sample, scheduled meetings

Horizon	FF4	MPI	CBINFO	MP1
GDP Growth				
Nowcast	1.02	0.93	2.51***	0.40
	(0.90)	(0.74)	(0.83)	(0.60)
1 q ahead	0.70	0.45	1.27*	0.39
	(0.67)	(0.66)	(0.67)	(0.40)
Avg. 2-4 q ahead	0.07	-0.21	-0.02	0.16
	(0.31)	(0.27)	(0.63)	(0.29)
GDP Deflator Inflation				
Nowcast	0.10	-0.22	0.74*	-0.06
	(0.42)	(0.22)	(0.41)	(0.27)
1 q ahead	-0.01	-0.08	0.51***	-0.04
	(0.20)	(0.18)	(0.17)	(0.14)
Avg. 2-4 q ahead	-0.08	-0.12	0.21	-0.05
	(0.17)	(0.14)	(0.16)	(0.11)
Unemployment Rate				
Nowcast	0.03	0.02	-0.39	0.07
	(0.22)	(0.19)	(0.17)	(0.14)
1 q ahead	-0.13	0.05	-0.67	0.06
	(0.29)	(0.25)	(0.21)	(0.20)
Avg. 2-4 q ahead	0.06	0.04	-0.49	0.18
	(0.41)	(0.31)	(0.41)	(0.26)
Interest Rate				
Nowcast	0.47	0.60	1.43***	0.19
	(0.47)	(0.56)	(0.47)	(0.33)
1 q ahead	0.78	0.87	1.99***	0.22
	(0.55)	(0.70)	(0.49)	(0.41)
Avg. 2-4 q ahead	0.61	0.79	1.29**	0.15
	(0.67)	(0.74)	(0.60)	(0.48)

Notes: The results are based on scheduled FOMC meetings that do not fall into the BCEI survey period (first week of the month). Robust standard errors in parentheses.

Table 2.3: Forecasters' response - Sub-samples, scheduled meetings

	Feb 1990 - July 2003			Aug 2003 - Dec 2015		
Horizon	FF4	MPI	CBINFO	FF4	MPI	CBINFO
GDP Growth						
Nowcast	0.87	1.42*	2.58***	1.30	0.05	4.19**
	(0.86)	(0.82)	(0.85)	(1.86)	(1.29)	(1.96)
1 q ahead	0.70	0.34	1.34*	0.84	0.32	2.85
	(0.47)	(0.59)	(0.69)	(1.52)	(1.17)	(1.77)
Avg. 2-4 q ahead	0.12	-0.36	-0.03	0.06	-0.15	1.36
	(0.29)	(0.35)	(0.71)	(0.66)	(0.31)	(1.42)
GDP Deflator Inflation						
Nowcast	-0.18	-0.15	0.26	0.51	-0.46	2.66***
	(0.37)	(0.24)	(0.40)	(0.89)	(0.42)	(0.91)
1 q ahead	-0.02	-0.04	0.50***	-0.02	-0.16	0.63
	(0.22)	(0.21)	(0.19)	(0.39)	(0.31)	(0.43)
Avg. 2-4 q ahead	-0.08	-0.11	0.08	-0.11	-0.13	0.36
	(0.23)	(0.22)	(0.18)	(0.19)	(0.15)	(0.54)
Unemployment Rate						
Nowcast	0.07	0.12	-0.47	-0.07	-0.05	-0.54
	(0.16)	(0.15)	(0.12)	(0.50)	(0.35)	(0.53)
1 q ahead	-0.12	0.16	-0.73	-0.14	-0.06	-0.73
	(0.23)	(0.21)	(0.15)	(0.66)	(0.45)	(0.71)
Avg. 2-4 q ahead	-0.09	0.06	-0.69	0.31	0.05	-0.49
	(0.20)	(0.20)	(0.19)	(1.12)	(0.57)	(1.93)
Interest Rate						
Nowcast	0.15	0.40	1.64***	1.01	0.81	1.28
	(0.37)	(0.43)	(0.56)	(1.03)	(1.11)	(0.84)
1 q ahead	0.38	0.52	2.21***	1.43	1.25	1.88*
	(0.46)	(0.51)	(0.57)	(1.21)	(1.36)	(1.00)
Avg. 2-4 q ahead	0.30	0.45	1.29**	1.25	1.09	2.43
	(0.51)	(0.54)	(0.50)	(1.69)	(1.33)	(2.30)

Notes: The results are based on scheduled FOMC meetings that do not fall into the BCEI survey period (first week of the month). Robust standard errors in parentheses.

2.6 Discussion

This paper explores the empirical importance of the information channel of U.S. monetary policy, paying particular attention to how it changed over time. We find that the information channel of monetary policy weakened around the early to mid-2000s since: (i) impulse responses to monetary policy shocks have the expected sign only when using the information-robust measure of monetary policy shocks before 2003, while after that the responses have the expected sign no matter whether the shock is cleaned for information effects or not; (ii) monetary policy surprises are correlated with central bank's forecasts only before 2003 but not afterward. Furthermore, the information advantage of the central bank in forecasting the state of the economy disappeared at the same time as the information channel weakened. These changes are related to improvements in the Fed's communication and transparency. Our results are robust to different estimation procedures and break tests, no matter whether we focus on scheduled or unscheduled meetings.

Acknowledgements

We thank T. Clark, D. Giannone, M. McCracken, S. Miranda-Agrippino, S. Van Norden and participants of 2020 IAAE Webinar series and conference participants at the Federal Reserve Bank of New York, the 2021 Padova Macro Talks, the 2021 Summer Meetings of the Econometric Society, the 2020 Econometric Society World Congress for their comments and constructive suggestions. We thank Lydia Petersen, Krista Box, John Roberts, Matthias Paustian and Eric Till for their help with the Greenbook/Tealbook interest rate forecast data, Refet Gürkaynak for providing the high-frequency monetary surprise dataset. This research is supported by the ERC Grant 615608 and the Cerca Programme/Generalitat de Catalunya. The Barcelona GSE acknowledges financial support from the Spanish ministry of the Economy and Competitiveness through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0063). The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System.

Appendices

A Information advantage and forecast timing

This section provides a sensitivity analysis to the time-ordering of the Greenbook/Tealbook and Blue Chip Economic Indicator forecasts in Section 2.2. Recall that, in order to compare Greenbook/Tealbook and BCEI forecasts, each Greenbook/Tealbook forecast (which is specifically prepared prior to each scheduled FOMC meeting) needs to be assigned a corresponding BCEI forecast. While the BCEI forecasts are always published on the tenth of each month, the publication day of the Greenbook/Tealbook forecasts varies with the date of the FOMC meeting. To match BCEI forecasts to Greenbook/Tealbook forecasts, in Section 2.2 we chose the BCEI forecast which occurred just before each FOMC meeting. Note that while this ensures a fixed ordering between BCEI forecasts and the FOMC announcements, it does not fix the publication order of Greenbook/Tealbook forecasts. In fact, in our dataset, there are 210 meetings for which the Greenbook/Tealbook is published after the BCEI forecast, while for 46 meetings the Greenbook/Tealbook forecast is published either on the same day as the BCEI or before.

Given the variation in the timing of Greenbook/Tealbook and BCEI forecasts, one might be concerned that a systematic change in the ordering of the forecasts over time, resulting from variation in the publication date of the Greenbook/Tealbook forecasts, might bias our findings of the informational advantage. For example, if Greenbook/Tealbook forecasts are systematically published after BCEI forecasts in the first part of the sample while this is not the case in the later part of the sample, the loss of information advantage could simply arise from this change in timing over the sample. News arriving between the publication of the forecasts could then create systematic differences in the information sets of the private sector and the central bank or forecasters could simply have had more time to process available information in one part of the sample than in the other, incorrectly leading us to conclude that there is time-variation in the information advantage.

To assess the importance of delays between the publication of both forecasts and to inspect whether the timing undergoes a systematic change over the sample, we calculate the difference between the publication dates of the Greenbook/Tealbook and the BCEI forecasts for two alternative timing assumptions which are used to match BCEI forecasts to their Greenbook/Tealbook equivalents. Figure A.8 reports boxplots for the number of days between the publication of the Greenbook/Tealbook forecasts and the BCEI forecasts. Positive values of the difference in publication dates imply that the Greenbook/Tealbook was published after the BCEI while negative values imply the reverse ordering. The difference between publication dates is computed using two different strategies for matching the BCEI forecasts: The left panel of Figure A.8 reports the difference between the publication dates for the baseline timing of Section 2.2 where BCEI forecasts are matched to

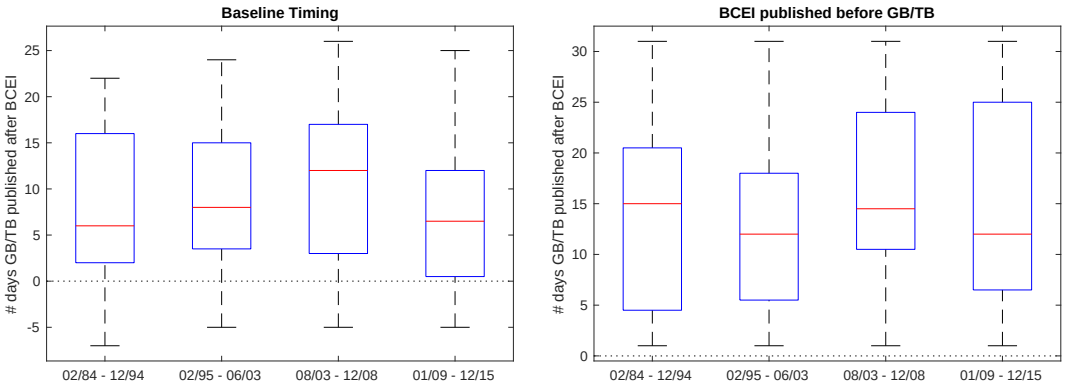


Figure A.8: Difference between forecast publication dates

Notes: Number of days between Greenbook/Tealbook and BCEI forecast publication for the baseline timing of Section 2.2 and an alternative timing where the BCEI is always published before the Greenbook/Tealbook forecasts.

FOMC announcements by ensuring that they are strictly ordered before the FOMC meeting, but without enforcing a particular ordering relative to the Greenbook/Tealbook forecasts. In contrast, the right panel of Figure A.8 reports the difference between publication dates when matching forecasts such that BCEI forecasts are published strictly before their Greenbook/Tealbook counterparts.[25]

By inspecting the boxplots for the baseline timing in the left panel of Figure A.8, we note several key points. First, for the majority of FOMC meetings, the Greenbook/Tealbook forecast is typically published within two weeks after the BCEI forecasts. There are some cases in which the Greenbook/Tealbook was published first. For those cases, the BCEI is typically published within a week of the Greenbook/Tealbook publication.[26] Second, the distribution of the delay in Greenbook/Tealbook forecasts relative to the BCEI forecasts has changed over time. In particular, the mean lag between BCEI and Greenbook/Tealbook publication dates increased from about 6-7 days to about 12 days for the mid-2000s sample period. However, in the most recent period, this lag decreases to 6 days. Note that this change in the timing could in principle affect the results from our information advantage tests. However, importantly, this timing change would bias our analysis towards finding an information advantage for the Fed forecasts in the later part of the sample rather than the earlier part of the sample. As we find the opposite, namely a disappearance of the information advantage in the most recent sample, removing such potential bias would further strengthen our conclusion that there is no information advantage in recent years.

Next, compare these results with the boxplots in the right panel of Figure A.8, which reports the Greenbook/Tealbook publication lag for the alternative matching strategy where BCEI

[25] In practice, this matching strategy implies that if a Greenbook/Tealbook publication date falls after the 10th of each month, the relevant BCEI forecasts is the one published in the same month as the Greenbook/Tealbook while for Greenbook/Tealbook forecasts which are published before the 10th of each month, the previous month's BCEI forecast is associated with the meeting.

[26] Note that at the time of "publication", the Greenbook/Tealbook forecasts are still not available to private forecasters as they are only released to the public with a five-year delay.

forecasts are always published before Greenbook/Tealbook forecasts. We note that using this matching strategy the mean publication lag of the Greenbook/Tealbook is generally higher by about 12-15 days and there is also no systematic change in timing which could explain a loss in information advantage. Thus, under the alternative matching scheme, the Greenbook/Tealbook should on average have more information advantage compared to the baseline scheme.

To verify that a difference in timing of the forecasts does not lead to a dramatically different conclusion regarding the disappearance of the information advantage, we repeat the Information-Advantage Fluctuation test from Section 2.2 using the alternative timing assumption where BCEI forecasts are always published before Greenbook/Tealbook forecasts. Figures A.9 and A.10 show the path of the $\tau_{GB,t}$ with the alternative forecast timing (blue-dashed line) compared to the baseline timing of Section 2.2 (black solid line). The figures clearly show that the disappearance of the information advantage in the most recent sample period remains robust to changing the timing of the forecasts. Specifically, the paths of $\tau_{GB,t}$ are very close to the original ones. There is more evidence of an information advantage in the recent sample period relative to the baseline only for the nowcast and one-quarter-ahead forecasts of GDP growth and the interest rate.

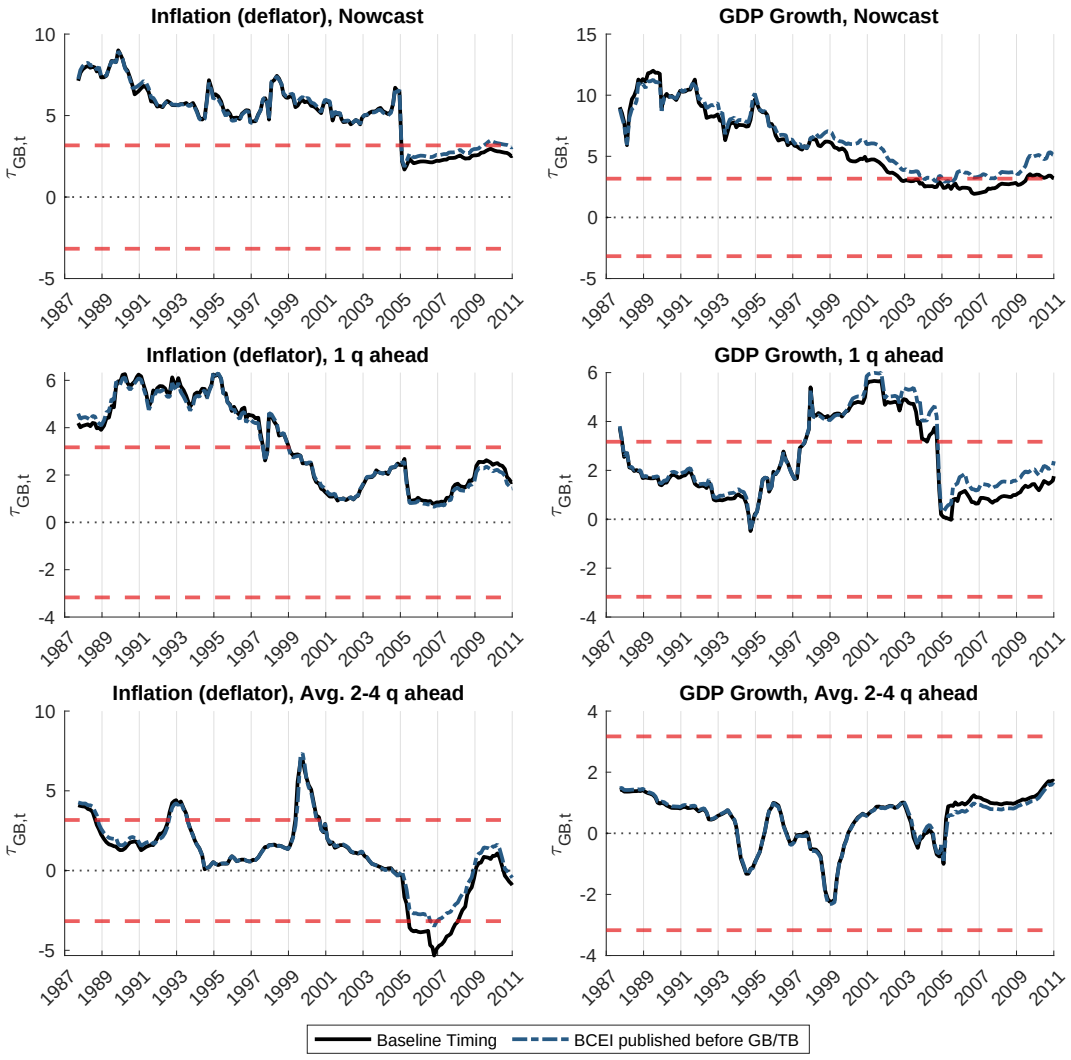


Figure A.9: Information advantage timing: GDP growth and inflation

Notes: The figure shows $\tau_{GB,t}$ from eq. (2.1) based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. Dashed (red) lines denote 5% critical value lines based on Rossi and Sekhposyan (2016)'s two-sided Fluctuation test.

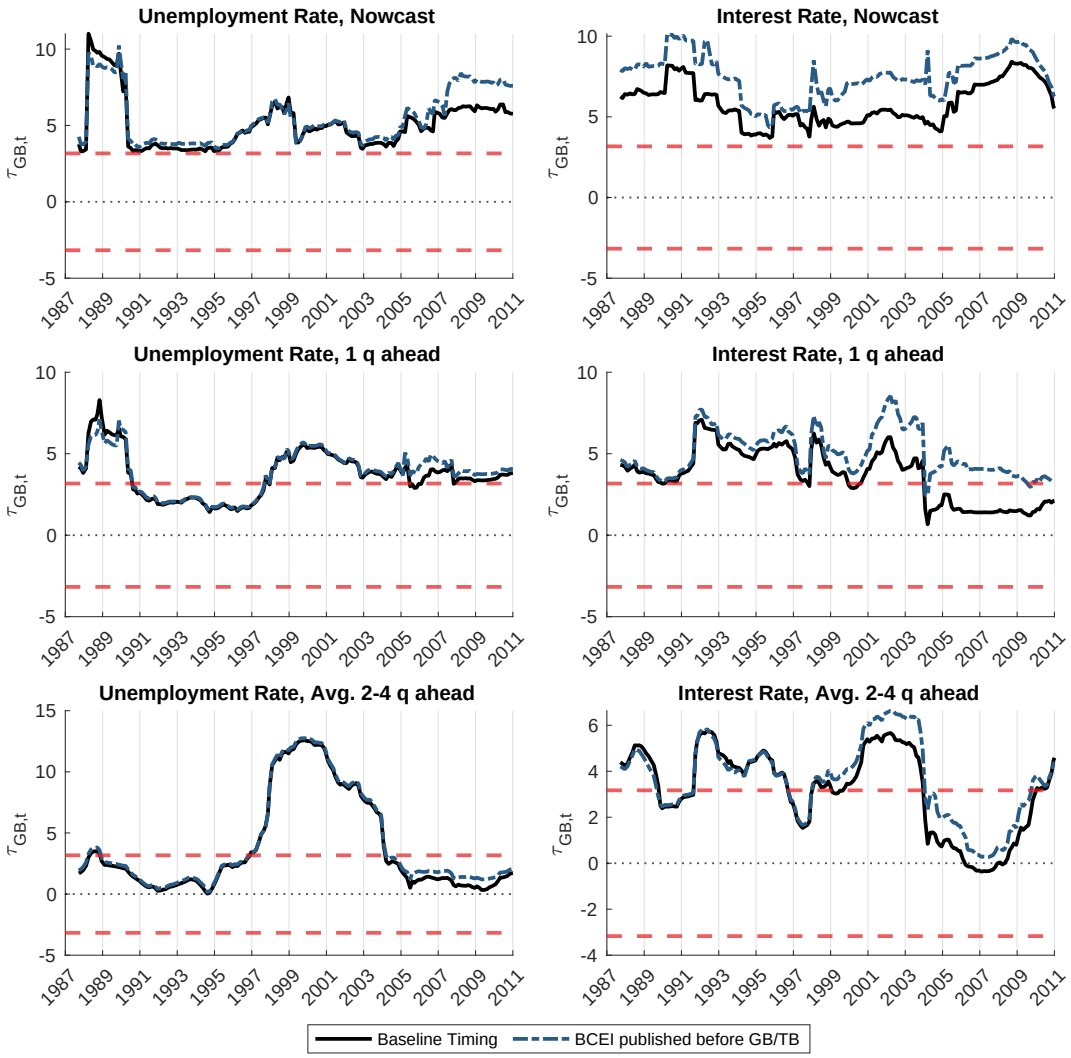


Figure A.10: Information advantage timing: Unemployment and interest rate

Notes: The figure shows $\tau_{GB,t}$ from eq. (2.1) based on $m = 60$ meetings rolling windows using a Newey-West covariance estimator with a truncation lag of $m^{1/4}$. Horizontal axes correspond to mid-window dates. Dashed (red) lines denote 5% critical value lines based on Rossi and Sekhposyan (2016)'s two-sided Fluctuation test.

B Additional evidence on the information content of high-frequency market-based surprises

In addition to the horizon-by-horizon projections reported in Section 2.3, we also consider a specification similar to Romer and Romer (2004) which jointly includes the forecasts and their revisions for all horizons. This specification is the same used to construct the information-robust instrument in Section 2.3.B. Miranda-Agrippino and Ricco (2020) also consider this regression and report that in their sample (1990-2009), an F-test rejects the null of joint significance of the coefficients (at the 5% level).

Table B.4 below reports the same F-test for our dataset. Column (1) shows that for the sample considered in Miranda-Agrippino and Ricco (2020), we replicate their F-statistic exactly. Column (2) shows that the coefficients continue to be insignificant at the 5% level even when extending the dataset to 2015. Finally, columns (3) and (4) show that our result from Section 2.3 continues to hold in this specification: High-frequency market-based surprises were significantly predictable by the Federal Reserve staff before the mid-2000s (the F-test rejects at 5% level), but that the predictability disappeared in the most recent period (the F-test does not reject at 5% level).

Table B.4: Projection on Fed information (all horizons)

	Feb 1990 - Dec 2009	Feb 1990 - Dec 2015	Feb 1990 - Jul 2003	Aug 2003 - Dec 2015
F	1.651	1.598	2.170	1.575
p	0.039	0.046	0.004	0.070
N	186	234	127	107

Notes: The table shows F-tests, p-values and number of observations from regressing the FF4 surprises on all the forecasts and at all horizons. F-statistics and p-values are based on heteroskedasticity-robust standard errors. Note that column (1) is the original sample of Miranda-Agrippino and Ricco (2020).

C Additional SVAR evidence

We assess the robustness of our SVAR conclusions in Section 2.4 by carrying out two additional exercises.

Figure C.11 compares impulse responses obtained using the information-robust instrument ($\mathbb{S}_t^{MPI}$) to the associated information component ($\mathbb{S}_t^{MPI}$) for the two sub-samples considered in Section 2.4. Our conclusion that information effects were important historically, but much less important in the most recent sample period is robust to this change: In the earlier sub-sample, the two sets of impulse responses have the opposite sign for real activity variables, and their differences are even more pronounced. In contrast, in the later sub-sample, both impulse responses become indistinguishable. In addition, the information component shows large estimation uncertainty in the later part of the sample. This is consistent with the result established in Section 2.3 that the information associated with the economic outlook of the Federal Reserve becomes less relevant in the most recent period.

Figure C.12 addresses potential misspecification concerns by repeating the analysis using a local projection approach rather than a BVAR. As the figure shows, our conclusion from Section 2.4 continues to hold, even though the confidence bands obtained from the local projections are much larger than the BVAR credible intervals.

D Additional evidence on the impact of information effects on forecasters' expectations

We explore the robustness of our findings in Section 2.5 to including unscheduled FOMC meetings. The latter are more likely to be associated with the release of central bank's information as they often take place as a reaction to important economic events.

Tables D.6 and D.6 report the results. Consistent with the findings in Paul (2020), the inclusion of unscheduled meetings leads to a larger number of statistically significant coefficients compared those based on scheduled meetings. Importantly, Table D.6 shows that most of the significance disappears in the second sub-sample. Overall, the coefficient estimates are similar to those based on scheduled FOMC meetings.

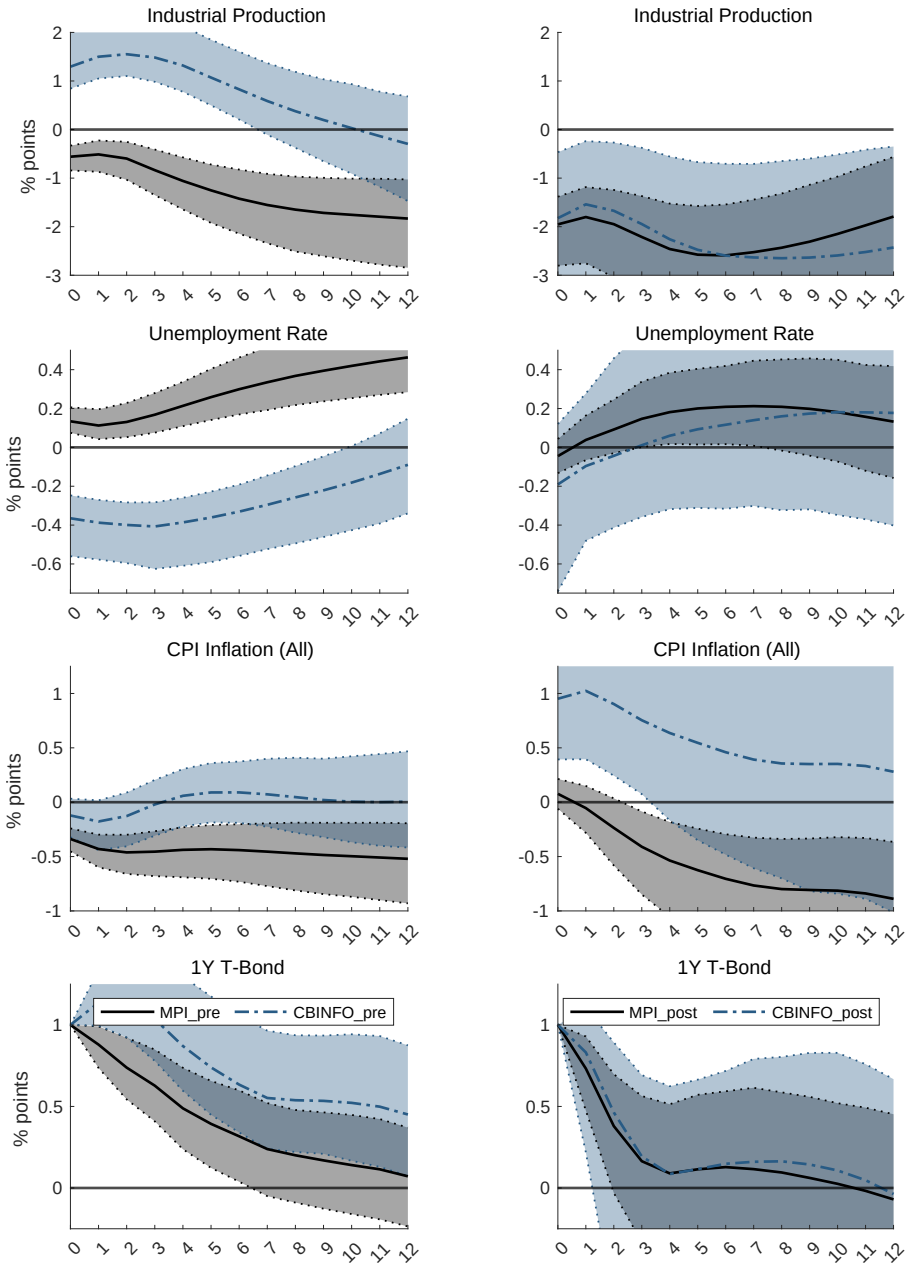


Figure C.11: Responses to a monetary shock: Decomposition

Notes: Impulse responses from Bayesian SVAR with standard macroeconomic priors and external instrument identification. VAR sample: January 1979 - December 2019. Instrument samples: February 1990 - July 2003 (left panel) and August 2003 - December 2015 (right panel). Shaded areas correspond to 90 percent credible intervals.

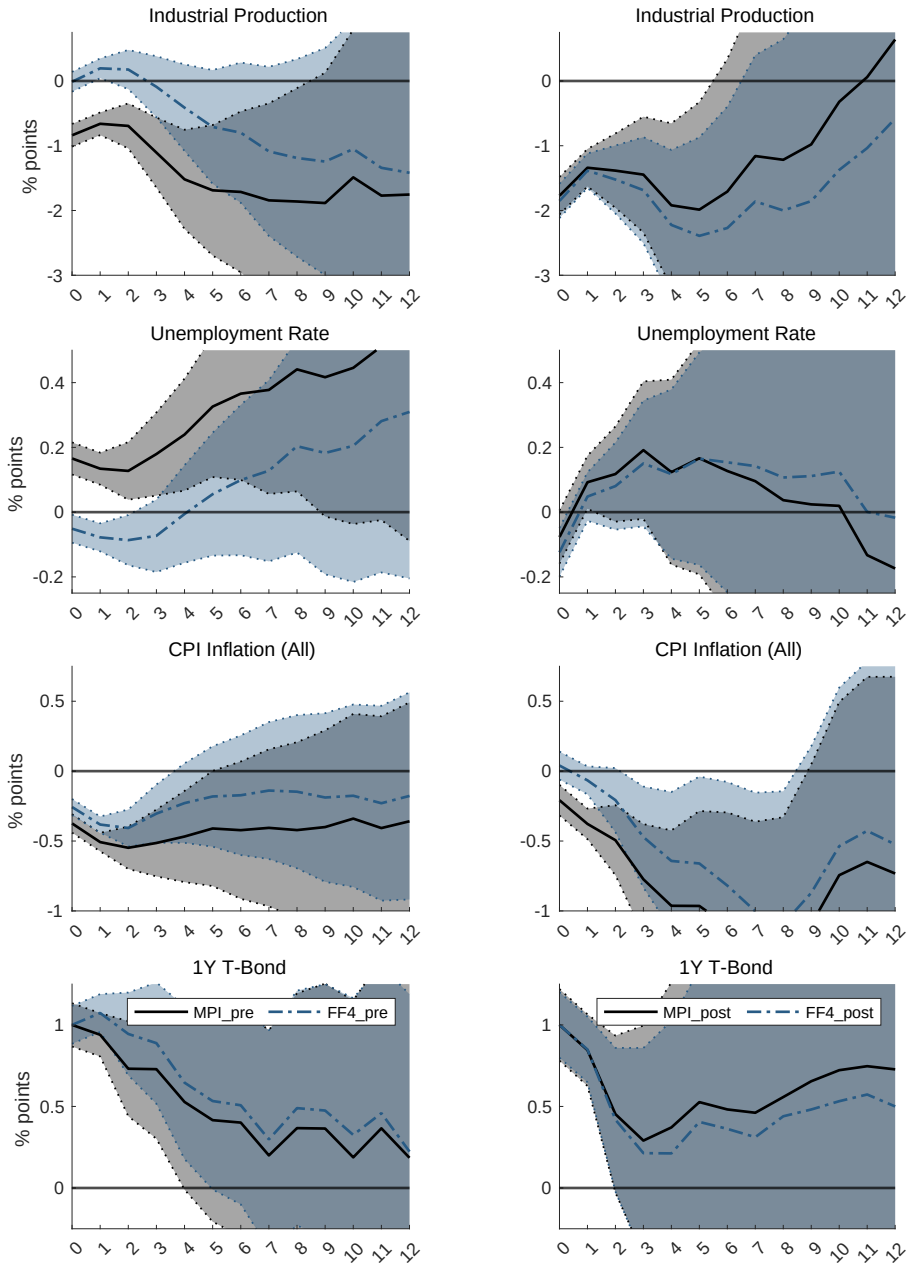


Figure C.12: Responses to a monetary shock: Local projection

Notes: Impulse responses from local projections (LP) with external instrument identification. LP sample: January 1979 - December 2019. Instrument samples: February 1990 - July 2003 (left panel) and August 2003 - December 2015 (right panel). Shaded areas correspond to 90 percent confidence intervals.

Table D.5: Forecasters' response - Full sample, all meetings

Horizon	FF4	MPI	CBINFO	MP1
GDP Growth				
Nowcast	1.20**	0.39	2.74***	1.57**
	(0.60)	(0.57)	(0.89)	(0.67)
1 q ahead	0.66	0.14	1.26**	0.89***
	(0.40)	(0.44)	(0.63)	(0.33)
Avg. 2-4 q ahead	-0.12	-0.17	-0.07	-0.02
	(0.18)	(0.20)	(0.58)	(0.13)
GDP Deflator Inflation				
Nowcast	0.15	-0.19	0.81**	0.10
	(0.23)	(0.19)	(0.38)	(0.13)
1 q ahead	0.12	-0.12	0.57***	0.11
	(0.14)	(0.14)	(0.17)	(0.10)
Avg. 2-4 q ahead	-0.04	-0.15	0.27*	0.02
	(0.09)	(0.10)	(0.15)	(0.07)
Unemployment Rate				
Nowcast	-0.20	-0.05	-0.47	-0.22
	(0.12)	(0.13)	(0.16)	(0.09)
1 q ahead	-0.22	0.04	-0.78	-0.24
	(0.15)	(0.16)	(0.21)	(0.14)
Avg. 2-4 q ahead	-0.15	0.02	-0.66	-0.31
	(0.20)	(0.18)	(0.39)	(0.19)
Interest Rate				
Nowcast	1.08***	0.74**	1.57***	0.98***
	(0.31)	(0.34)	(0.41)	(0.19)
1 q ahead	1.32***	0.84*	2.06***	1.15***
	(0.37)	(0.43)	(0.44)	(0.24)
Avg. 2-4 q ahead	0.99**	0.78*	1.40***	1.01***
	(0.40)	(0.44)	(0.51)	(0.27)

Notes: The results are based on all (scheduled and unscheduled) FOMC meetings that do not fall into the BCEI survey period (first week of the month). Robust standard errors in parentheses.

Table D.6: Forecasters' response - Sub-samples, all meetings

	Feb 1990 - July 2003			Aug 2003 - Dec 2015		
Horizon	FF4	MPI	CBINFO	FF4	MPI	CBINFO
GDP Growth						
Nowcast	1.26*	0.54	2.62***	1.13	-0.09	4.58**
	(0.65)	(0.61)	(0.91)	(1.31)	(1.28)	(1.93)
1 q ahead	0.65**	0.08	1.26*	0.97	0.23	3.11*
	(0.32)	(0.35)	(0.66)	(1.09)	(1.17)	(1.83)
Avg. 2-4 q ahead	-0.11	-0.18	-0.02	0.07	-0.19	1.59
	(0.19)	(0.24)	(0.67)	(0.40)	(0.31)	(1.41)
GDP Deflator Inflation						
Nowcast	0.06	-0.08	0.32	0.36	-0.52	2.85***
	(0.23)	(0.22)	(0.37)	(0.56)	(0.43)	(0.88)
1 q ahead	0.13	-0.09	0.52***	0.05	-0.20	0.77*
	(0.16)	(0.16)	(0.20)	(0.28)	(0.33)	(0.44)
Avg. 2-4 q ahead	-0.06	-0.15	0.13	-0.05	-0.16	0.58
	(0.11)	(0.12)	(0.16)	(0.13)	(0.16)	(0.57)
Unemployment Rate						
Nowcast	-0.22	-0.06	-0.55	-0.18	-0.02	-0.62
	(0.10)	(0.11)	(0.12)	(0.33)	(0.35)	(0.53)
1 q ahead	-0.21	0.07	-0.84	-0.24	-0.02	-0.84
	(0.12)	(0.13)	(0.18)	(0.44)	(0.45)	(0.70)
Avg. 2-4 q ahead	-0.21	-0.00	-0.80	-0.03	0.10	-0.80
	(0.12)	(0.11)	(0.26)	(0.68)	(0.57)	(1.93)
Interest Rate						
Nowcast	1.14***	0.72***	1.76***	0.93	0.78	1.37
	(0.26)	(0.21)	(0.48)	(0.92)	(1.11)	(0.84)
1 q ahead	1.30***	0.69***	2.26***	1.41	1.22	1.97*
	(0.31)	(0.25)	(0.51)	(1.11)	(1.35)	(1.01)
Avg. 2-4 q ahead	0.90***	0.66**	1.39***	1.29	1.07	2.60
	(0.26)	(0.28)	(0.49)	(1.31)	(1.32)	(2.29)

Notes: The results are based on all (scheduled and unscheduled) FOMC meetings that do not fall into the BCEI survey period (first week of the month). Robust standard errors in parentheses.

Chapter 3

Robust Inference in Structural VAR Models Identified by Non-Gaussianity

This paper is co-authored with Adam Lee and Geert Mesters.

3.1 Introduction

In this paper, we develop robust inference methods for structural vector autoregressive (SVAR) models that are identified using non-Gaussian error distributions. To outline our contribution, consider the K-dimensional SVAR model

$$Y_t = c + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + A^{-1}\epsilon_t \tag{3.1}$$

where Y_t is a $K \times 1$ vector of macroeconomic variables, c is an intercept, $B_1, \ldots, B_p$ are the $(K \times K)$ autoregressive matrices, A is the $(K \times K)$ invertible contemporaneous effect matrix and ϵ_t is the $K \times 1$ vector of independent structural shocks, whose components are assumed to have mean zero and unit variance.

It is well known that, without further restrictions, the first and second moments of the process $\{Y_t\}$ are insufficient to identify all parameters of the SVAR (e.g. Sims, 1980). Interestingly, however, when at least $K - 1$ components of ϵ_t admit a non-Gaussian distribution, all parameters of the SVAR can be recovered up to sign and permutation of the columns of A^{-1}, see e.g. Comon (1994) and the discussion in Gouriéroux et al. (2017). Several recent papers have exploited such a non-Gaussian identification strategy to conduct inference in SVAR models (e.g. Lanne and Lütkepohl, 2010; Moneta et al., 2013; Lanne et al., 2017; Maxand, 2018; Lanne and Luoto, 2019; Gouriéroux et al., 2017, 2019; Tank et al., 2019; Herwartz, 2019; Bekaert et al., 2019, 2020; Fiorentini and Sentana, 2020; Velasco, 2020; Guay, 2020; Braun, 2021; Sims, 2021).

Unfortunately, as we show in this paper, existing inference methods in non-Gaussian SVARs are not robust to situations in which the densities of the structural errors that generated the data are "close" to a Gaussian distribution. Intuitively, what matters for correctly sized inference is not non-Gaussianity per se, but a sufficient distance from the Gaussian distribution. When the true distributions of the structural errors are close to the Gaussian distribution, local identification deteriorates and coverage distortions occur.[1] The problem is somewhat analogue to the weak instruments problem where it is well known that a non-zero correlation between the instrument and the endogenous variable is not sufficient to conduct standard inference, but that the correlation must be sufficiently large in order to yield strong identification. Similarly, in our setting, non-Gaussianity alone is not sufficient for standard MLE or GMM methods to yield correct inference, if the distance to the Gaussian distribution is not sufficiently large.

In this paper, we propose a solution to this problem by combining insights from the econometric literature on weak identification robust hypothesis testing as well as the statistical literature on semi-parametric inference approaches. Specifically, we treat the SVAR model as a semi-parametric model where the densities of the structural errors form the non-parametric part. We use a semi-parametric equivalent of the Neyman-Rao score statistic in order to conduct inference on the possibly weakly identified (or not identified) parameters of the SVAR. More precisely, the semi-parametric score statistic that we employ is based on the quadratic form of the efficient score function, which projects out the infinite dimensional nuisance parameters (i.e. the densities of the structural errors) from the conventional score function. Similar as for the parametric score test, being able to fix the weakly identified parameters under the null hypothesis, circumvents the identification problem (e.g. Andrews and Mikusheva, 2016) and we show that the semi-parametric score test has a standard χ^2 limit under the null hypothesis. Additionally, if the normalizing matrix, i.e. the efficient information matrix, is non-singular, the test is asymptotically uniformly most powerful within the class of invariant tests (e.g. Choi et al., 1996). Implementing the score test is computationally easy as it only requires estimating the log density scores of the structural errors. For this, we use a B-spline methodology that was developed in Jin (1992) and also considered in Chen and Bickel (2006) for independent component analysis. This approach allows us to consider a wide variety of possible distributions for the structural errors without committing to a specific distribution. The efficient score test allows us to construct confidence regions for the possibly weakly identified (or not identified) parameters of the SVAR. These regions can then be used to construct bands for the structural impulse responses while exploiting efficient estimates for the well identified parameters (e.g. van der Vaart, 2002, Chapter 7).

We assess the finite-sample performance of our method in a large simulation study and find that the empirical rejection frequencies of the semi-parametric score test are always close

[1] Simulation studies in, among others, Gouriéroux et al. (2017, 2019) and Lanne and Luoto (2019) have previously highlighted such coverage distortions in the case of "weakly" non-Gaussian distributions.

to the nominal size, regardless of the true distribution of the errors. This is in contrast to existing methods that are not robust to weak non-Gaussianity, which show substantial size distortions for some of the non-Gaussian distributions considered. We also analyze power of the proposed procedure and find that the power of the semi-parametric score test comes close to the power of an "oracle"-approach where the test is conducted based on a parametric score statistic using the true densities of the structural errors (which are unknown in practice).

Finally, we employ the proposed approach in an empirical study that revisits identification of supply and demand elasticities in the U.S. labor market based on the model of Baumeister and Hamilton (2015), but using a frequentist procedure. This model was recently studied by Lanne and Luoto (2019) using a non-Gaussian identification strategy. They find evidence to reject some of the restrictions imposed in Baumeister and Hamilton (2015). Using our semi-parametric approach, we construct confidence regions for the labor supply and labor demand elasticities and show that the labor demand elasticity is only very weakly identified by non-Gaussianity of the structural shocks. In fact, our confidence regions for the demand and supply elasticities are consistent with the posterior distributions reported by Baumeister and Hamilton (2015). We also construct identification-robust confidence bands for the impulse responses for labor demand and labor supply shocks based on our semi-parametric approach. We find that a non-Gaussianity identification strategy is not sufficient to identify the dynamic response of the economy.

The remainder of this paper is organized as follows. Continuing the introduction, we briefly discuss the related literature. In Section 3.2, we illustrate how non-Gaussian distributions can help with identification and how the weak identification problem arises. Section 3.3 casts the SVAR model as a semi-parametric model and establishes a number of preliminary results. The semi-parametric score testing approach is derived in Section 3.4. Section 3.5 evaluates the finite-sample performance of the proposed method. Section 3.6 presents the empirical study. Finally, Section 3.7 concludes.

Related literature

This paper relates to three strands of the existing literature: (i) identification robust testing, (ii) semi-parametric modeling and (iii) structural VAR models.

First, the econometric literature has developed a great number of methods and approaches for conducting inference on parameters that are weakly identified or not identified. Both general approaches (e.g. Kleibergen, 2005; Andrews and Cheng, 2012) and model specific solutions have been developed, see for instance the solutions to the weak instruments problem in the linear IV model summarized in Andrews et al. (2019). The crucial difference in our setting is that the nuisance parameters are infinite dimensional. In fact, in our case, the distance between the distributions of the structural shocks and the Gaussian distribution determines the strength of identification. Despite this difference, conceptually our approach

is similar to the score testing approach developed for parametric models in Andrews and Mikusheva (2016).

Second, we consider the SVAR model as a semi-parametric model and we build on the general statistical theory discussed in Bickel et al. (1998) and van der Vaart (2002). While the majority of the statistical literature focuses on efficient estimation in semi-parametric models, a few papers have contributed to testing in well identified models (e.g. Choi et al., 1996; Bickel et al., 2006). The difference to our paper is that in our setting, a subset of the parameters of interest is possibly not identified.

Third, the SVAR literature has produced numerous identification strategies including imposing short run, long run or sign restrictions, relying on external instruments and exploiting statistical features such as heteroskedasticity and non-Gaussianity (see Ramey (2016) for a recent review). The last category is potentially attractive as it does not require taking a stand on any economic mechanism, nor does it rely on the availability of strong instruments. The non-Gaussian approach has been considered in an increasing number of papers mentioned in the introduction above. For this approach, an important contribution is Guay (2020) who provides a pre-test that is useful to determine whether identification is sufficiently strong. Our paper removes the need for such pre-tests, as it provides an efficient inference procedure that does not assume that the structural errors are sufficiently non-Gaussian.

Last, but not least, this paper builds on Lee and Mesters (2021) who consider a similar score testing approach. The crucial difference is that Lee and Mesters (2021) require that the observations are independent across time. Allowing for dependence is non-trivial in the semi-parametric setting as the majority of the theory is developed for random samples (e.g. Bickel et al., 1998; van der Vaart, 2002). For instance, here we rely on LAN results developed in Hallin and Saidi (2007), but numerous other differences in the approach exist.

3.2 Illustration of non-Gaussianity identification

In this section, we illustrate briefly how non-Gaussian distributions of the structural errors can help to identify the parameters of the SVAR model. Furthermore, we provide an intuitive explanation of the weak identification problem that is the main problem of interest in this paper.

For illustration, consider a bivariate SVAR model as defined in eq. (3.1) in the introduction to this paper, but assume for simplicity that (i) the number of lags is zero ($p = 0$) and (ii) that the contemporaneous effect matrix A is orthonormal.[2] Under these assumptions, the (2×2) matrix A can be parameterized by a scalar parameter α and the model can be written

[2]Note, that the assumption of an orthonormal A matrix can be asymptotically justified if the data Y_t is jointly re-scaled to have mean zero and identity variance matrix (pre-whitening). For details, see Gouriéroux et al. (2017).

as follows:

$$Y_t = A^{-1}\epsilon_t \qquad \text{where} \qquad A^{-1} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

In this model, the coefficient of interest is the scalar α that determines the angle of the rotation matrix A. If, for example, $\alpha = 0$, A equals the identity matrix and each of the structural shocks only impacts its respective component in Y_t. For $0 < \alpha < \pi$, or integer multiples thereof, the off-diagonal elements are non-zero so that the shocks affect all variables, with signs depending on the value of α.

To illustrate how non-Gaussian distributions for ϵ_t may help to identify α, we study the log-likelihood of Y_t in the model above using different choices for the distributions of the structural shocks $\epsilon_{k,t}$. Figure 3.1 shows contour plots of the log-likelihood of the bivariate SVAR(0) model for four choices for the distributions of the structural shocks: A Gaussian distribution, a student-t distribution with 5 degrees of freedom, a skewed distribution and a bimodal distribution.[3] In addition to the likelihood contours, each of the plots also shows a red and a blue line. These indicate the vector Y_t (i.e. a linear combination of the structural errors ϵ_t), corresponding to two different choices for α.

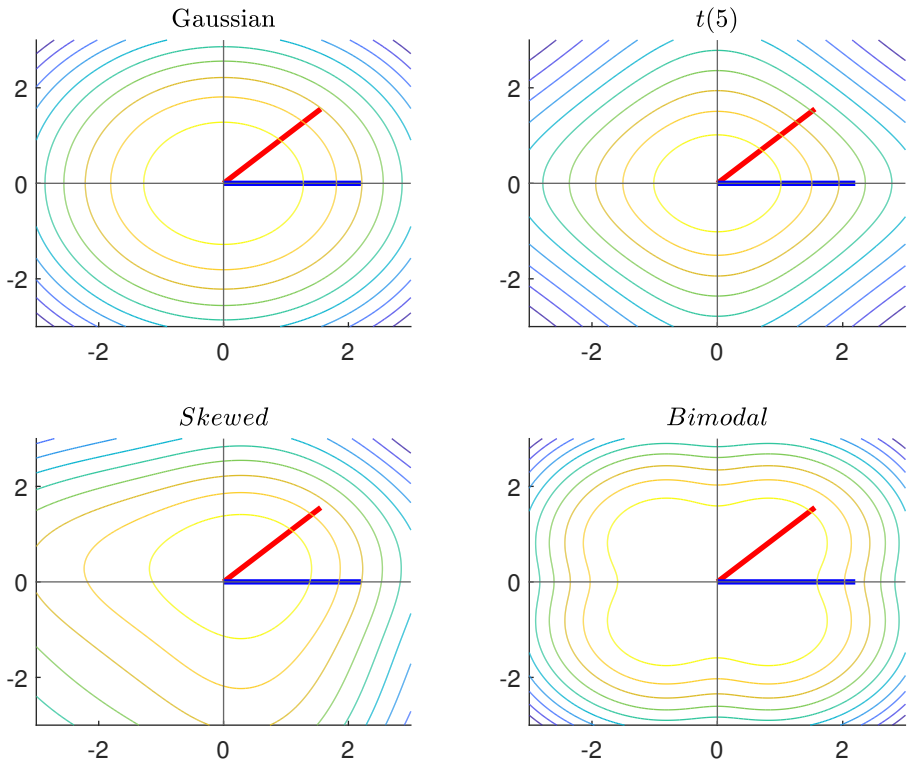
Focus first on the Gaussian case (top-left panel). When the structural errors have an exact Gaussian distribution, the value of the log-likelihood is identical for all choices of α and the red and blue vectors reach the same level curve of the likelihood. Therefore, α cannot be identified from the likelihood of the model. This illustration reflects the standard identification problem of an SVAR model where, without additional identification restrictions, the impact effects of the structural shocks are not identifiable.

In contrast, consider next the cases where the structural errors follow non-Gaussian distributions. These cases are depicted in the remaining panels of the figure. Notice that in each of the non-Gaussian cases, the blue and red vectors end on different contour lines. This means that different choices of α lead to different values of the log-likelihood and hence, α is identifiable. More precisely, when we change the value of α i.e. rotate the red and blue vectors, the only values that end up on the same contour lines are those values of α that lead to the same elements of A, up to permutation and sign changes of the columns (i.e. exchanging the four quadrants of each plot). These examples illustrate, how non-Gaussianity of the structural errors can help to identify parameters up to permutation and sign changes of the columns.

Finally, we turn to the problem of weak non-Gaussianity identification which arises, when the distance from the Gaussian distribution is not very large. In such scenarios, changes in α only imply minor changes in the level of the likelihood, so that the likelihood ends up being rather flat in the direction of α. Compare for instance, the panels corresponding to the $t(5)$ distribution and the bimodal distribution. In the case of the bimodal density, the

[3] For simplicity, both structural errors are assumed to have identical distributions.

Figure 3.1: Identification with non-Gaussian distributions



Notes: The figure shows the log-likelihood contours of Y_t in the SVAR(0) model with scalar parameter α for different distributions of the structural shocks, $\epsilon_{k,t}$. The red and blue lines in each plot denote the vector Y_t corresponding to two different choices for α.

red and blue vectors end on clearly distinguishable contour lines of the log-likelihood. In contrast, for the student-t density, the level difference is small and the red and blue vectors almost reach the same contour line. In the extreme case of weak identification, we end up in the Gaussian case (upper-left panel of Figure 3.1) where α is completely unidentified. More importantly, however, there are many empirically relevant scenarios, including the $t(5)$ density, where α always remains weakly identified.[4]

[4]In fact, the theoretical likelihood plots shown in Figure 3.1 correspond quite closely to the power results obtained in our simulation study below; see Figure 3.2 in Section 3.5.

3.3 Semi-parametric SVAR model

In this section, we cast the SVAR model as a semi-parametric model and state the hypotheses of interest. For convenience, we adopt the following notation

$$Y_t = c + BX_t + A(\alpha, \sigma)^{-1}\epsilon_t \,, \qquad t = 1, \ldots, n \,, \tag{3.2}$$

where $X_t \coloneqq (Y'_{t-1}, \ldots, Y'_{t-p})'$ and $B \coloneqq (B_1, \ldots, B_p)$. The contemporaneous effect matrix $A(\alpha, \sigma)$ is parameterized by the vectors α and σ, where we adopt the convention that α cannot be recovered from the second moments of $\{Y_t\}$ but σ can. We provide some examples below. Further, we let $\eta = (\eta_1, \ldots, \eta_K)$ correspond to the density functions of $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{K,t})'$ and summarize the parameters in

$$\theta = (\gamma, \eta) \,, \qquad \gamma = (\alpha, \beta) \,, \qquad \beta = (\sigma, c, b) \,, \tag{3.3}$$

where $b = \mathrm{vec}(B)$. Let $Y^n = (Y_1, \ldots, Y_n)'$ and let P^n_θ denote the distribution of Y^n conditional on the initial values $(Y_{1-p}, \ldots, Y_0)$. The semi-parametric SVAR model is defined by

$$\mathcal{P}^n_\Theta = \{P^n_\theta : \theta \in \Theta\} \qquad \Theta = \underbrace{\mathcal{A} \times \mathcal{B}}_{\Gamma} \times \mathcal{H} \,, \tag{3.4}$$

where $\Gamma \subset \mathbb{R}^\gamma$ and $\mathcal{H} = \prod_{k=1}^K \mathsf{H}$ with

$$\mathsf{H} \coloneqq \left\{ g \in L_1(\lambda) \cap \mathcal{C}^1(\lambda) : \; g(z) \geq 0, \int g(z)\,\mathrm{d}z = 1, \int z g(z) dz = 0, \int \kappa(z) g(z)\,\mathrm{d}z = 0 \right\} \,,$$

where λ denotes Lebesgue measure on $\mathbb{R}$, $\mathcal{C}^1(\lambda)$ is the class of real functions on $\mathbb{R}$ which are continuously differentiable λ-a.e. and $\kappa(z) = z^2 - 1$. The parameter space for the densities η_k is restricted such that that $\epsilon_{k,t}$ has mean zero and variance one. Further restrictions, required to estimate certain quantities consistently, will be placed on the parameter spaces below.

To allow for different parameterizations of the SVAR, we refrain from imposing additional structure on $A(\alpha, \sigma)$. The following examples illustrate various options.

EXAMPLE 3.3.1: *Consider $A(\alpha, \sigma)^{-1} = \Sigma^{1/2}(\sigma)R(\alpha)$ where $\Sigma^{1/2}(\sigma)$ is lower triangular with nonzero entries determined by σ and $R(\alpha)$ is a rotation matrix parameterized by α. Different choices for $\alpha \to R(\alpha)$ are possible including the Cayley and trigonometric transformations. We note that σ is identified by the second moments of $\{Y_t\}$ and the identification problem is isolated in the rotation matrix R as in Section 3.2.*

EXAMPLE 3.3.2: *Consider a bivariate SVAR(p) model for $Z_t = (Z_{1,t}, Z_{2,t})'$ which defines a demand and supply equation*

$$\begin{bmatrix} -\alpha^d & 1 \\ -\alpha^s & 1 \end{bmatrix} Z_t = \sum_{i=1}^{p} \Phi_i Z_{t-i} + \epsilon_t$$

where $\epsilon_t = (\epsilon_t^d, \epsilon_t^s)'$ contains a demand shock and a supply shock with standard deviations σ_d and σ_s, respectively. α^s captures a supply elasticity and α^d a demand elasticity. Let $\sigma = (\sigma_d, \sigma_s)$ and $\alpha = (\beta^d, \beta^s)$. Then $A^{-1}(\alpha, \sigma)$ can be parameterized as

$$A^{-1}(\alpha, \sigma) = \begin{bmatrix} \frac{\sigma_d}{\alpha^s - \alpha^d} & -\frac{\sigma_s}{\alpha^s - \alpha^d} \\ \frac{\sigma_d \alpha^s}{\alpha^s - \alpha^d} & -\frac{\sigma_s \alpha^d}{\alpha^s - \alpha^d} \end{bmatrix}$$

The standard deviations of the supply and demand shocks σ can be identified by the second moments of $\{Z_t\}$ and the identification problem consists in determining the demand and supply elasticities contained in α.

3.3.1 Assumptions

Before discussing the testing methodology, we discuss our assumptions. Broadly speaking, we have two types of assumptions: (i) the main assumptions that are required for deriving and estimating the efficient score function and (ii) an additional assumption that imposes that the density scores can be consistently estimated. We separate the two as different density score estimators can be considered which then only need to satisfy the second assumption. In the Appendix, we show that the density score estimator that we adopt in this paper satisfies this assumption under mild conditions on the densities, η_k. Our main assumption is stated as follows.

ASSUMPTION 3.3.1: *For model* (3.2)*, we assume that*

1. $|I_K - \sum_{j=1}^{p} B_j z^j| \neq 0$ *for all* $|z| \leq 1$ *with* $z \in \mathbb{C}$
2. $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{K,t})'$ *is independently and identically distributed across t, with independent components $\epsilon_{k,t}$ that have a nowhere vanishing density η_k with respect to the Lebesgue measure on $\mathbb{R}$. The densities η_k are continuously differentiable with log density scores denoted by $\phi_k(z) := \partial \log \eta_k(z)/\partial z$, and for all $k = 1, \ldots, K$*
   - *(a)* $\mathbb{E}\epsilon_{k,t} = 0$, $\mathbb{E}\epsilon_{k,t}^2 = 1$, $\mathbb{E}\epsilon_{k,t}^{4+\delta} < \infty$, $\mathbb{E}(\epsilon_{k,t}^4) - 1 > \mathbb{E}(\epsilon_{k,t}^3)^2$, *and* $\mathbb{E}\phi_k^{4+\delta}(\epsilon_{k,t}) < \infty$ *(for some $\delta > 0$);*
   - *(b)* $\mathbb{E}\,\phi_k(\epsilon_{k,t}) = 0$, $\mathbb{E}\,\phi_k^2(\epsilon_{k,t}) > 0$, $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t} = -1$, $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^2 = 0$ *and* $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^3 = -3$*;*

Part 1 imposes that the SVAR is stationary and causal. Part 2 imposes that the densities of

the errors are continuously differentiable and certain moment conditions hold. Specifically, part (a) normalizes the errors to have mean zero, variance one and finite four+δ moments.[5] Additionally, we require the log density scores $\phi_k(x) = \partial \log \eta_k(x)/\partial x$ evaluated at the errors to have finite four+δ moments. Part (b) simplifies the construction of the efficient score functions. Whilst this may at first glance appear a strong condition, Lemma S8 in Lee and Mesters (2021) shows that if the first part holds, then a simple sufficient condition is that the tails of the densities η_k converge to zero at a polynomial rate. We define $\mathcal{H}_0$ as the subset of $\mathcal{H}$ that satisfies Assumption 3.3.1 part 2. Similarly, $\mathcal{B}_0 \subset \mathcal{B}$ satisfies Assumption 3.3.1 part 1.

Together, these assumptions ensure that (i) the LAN result of Hallin and Saidi (2007) holds (see Theorem 2.1 in their paper) and (ii) the efficient score function can be derived analytically.

Next, we assume that the density scores can be consistently estimated

ASSUMPTION 3.3.2: *Let β_n be a deterministic sequence with $\sqrt{n}(\beta_n - \beta) = O(1)$ and let $\theta_n = (\alpha_0, \beta_n, \eta)$ for some $(\beta, \eta) \in \mathcal{B}_0 \times \mathcal{H}_0$ and suppose we have an array of estimates $\{\hat{\phi}_{k,n}(A_{n,k\bullet}(Y_t - c_n - B_n X_t))\}_{n\geq 1, t\leq n}$ for $k = 1, \ldots, K$ where $A_n = A(\alpha_0, \sigma_n)$ such that for each $k \neq j$:*

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}(Y_t - c_n - B_n X_t)) - \phi_k(A_{n,k\bullet}(Y_t - c_n - B_n X_t))\right] W_{t,n} = o_{P_{\theta_n}}(n^{-1/2}), \tag{3.5}$$

and for $\nu_n = \nu_{n,p}^2$ with $p := \min\{1 + \delta/4, 2\}$ and $\nu_{n,p} = n^{(1-p)/p}$ if $p \in (1,2)$ or $\nu_{n,p} = n^{-1/2}\log(n)^{1/2+\rho}$, for some $\rho > 0$, if $p = 2$, we have

$$\frac{1}{n}\sum_{t=1}^{n}\left(\left[\hat{\phi}_{k,n}(A_{n,k\bullet}(Y_t - c_n - B_n X_t)) - \phi_k(A_{n,k\bullet}(Y_t - c_n - B_n X_t))\right] W_{t,n}\right)^2 = o_{P_{\theta_n}}(\nu_n). \tag{3.6}$$

where $W_{t,n}$ is some arbitrary random variable that is independent from $A_{n,k\bullet}(Y_t - c_n - B_n X_t)$ and such that $\frac{1}{n}\sum_{t=1}^{n} W_{t,n}^2 \xrightarrow{P_{\theta_n}} \mathbb{E}_{\theta_n} W_{t,n}^2 < \infty$.

The assumption imposes that certain functions of the log density scores can be estimated sufficiently well. The assumption is stated for deterministic sequences β_n which simplifies verification for specific choices of $\hat{\phi}_{k,n}$.[6] In the Appendix, we show that the B-spline based estimator of Jin (1992) and Chen and Bickel (2006) satisfies this assumption under mild regularity conditions on the densities η_k.

[5] $\mathbb{E}(\epsilon_{k,t}^4) - 1 \geq \mathbb{E}(\epsilon_{k,t}^3)^2$ always holds; this is known as Pearson's inequality. See e.g. Result 1 in Sen (2012). Assuming that $\mathbb{E}(\epsilon_{k,t}^4) - 1 > \mathbb{E}(\epsilon_{k,t}^3)^2$ rules out (only) cases where $1, \epsilon_{k,t}$ and $\epsilon_{k,t}^2$ are linearly dependent when considered as elements of L_2. See e.g. Theorem 7.2.10 in Horn and Johnson (2013).

[6] More specifically, it implies that under P_{θ_n} we have that $A_{n,k\bullet}(Y_t - c_n - B_n X_t) \approx \epsilon_{k,t}$ and terms involving $\epsilon_{k,t}$ are easy to handle using Assumption 3.3.1

3.3.2 Efficient score function

One of the key ingredients in our framework is the efficient score function for $\gamma = (\alpha, \beta)$ with respect to η. Loosely speaking, the efficient score function is defined as the projection of the scores for γ on the space orthogonal to the tangent space of η (e.g. Definition 2.15 in van der Vaart, 2002).

When η is finite dimensional, the tangent space is merely the linear span of the scores of η, but when η is infinite dimensional, constructing the tangent space and the subsequent projection requires more care. A formal derivation is provided in Lee and Mesters (2021); here we merely state our main result.

LEMMA 3.3.1: *Given Assumption 3.3.1 we have that the efficient score function for γ in the semi-parametric SVAR model $\mathcal{P}_\Theta^n$ at any $\theta = (\gamma, \eta)$ with $\gamma = (\alpha, \beta)$, $\alpha \in \mathcal{A}_0$, $\beta = (\sigma, b, c) \in \mathcal{B}_0$ and $\eta \in \mathcal{H}_0$ is given by $\tilde{\ell}_{n,\theta}(y^n) = \sum_{t=1}^n \tilde{\ell}_\theta(y_t, x_t)$, where*

$$\tilde{\ell}_\theta(y_t, x_t) = \left(\left(\tilde{\ell}_{\theta,\alpha_l}(y_t, x_t) \right)_{l=1}^{K_\alpha}, \left(\tilde{\ell}_{\theta,\sigma_l}(y_t, x_t) \right)_{l=1}^{K_\sigma}, \left(\tilde{\ell}_{\theta,c_l}(y_t, x_t) \right)_{l=1}^{K_c}, \left(\tilde{\ell}_{\theta,b_l}(y_t, x_t) \right)_{l=1}^{K_b} \right)$$

with components

$$\begin{aligned}
\tilde{\ell}_{\theta,\alpha_l}(y_t, x_t) &= \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta_{l,k,j}^\alpha \phi_k(A_{k\bullet} v_t) A_{j\bullet} v_t + \sum_{k=1}^K \zeta_{l,k,k}^\alpha \left[\tau_{k,1} A_{k\bullet} v_t + \tau_{k,2} \kappa(A_{k\bullet} v_t) \right] \\
\tilde{\ell}_{\theta,\sigma_l}(y_t, x_t) &= \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta_{l,k,j}^\sigma \phi_k(A_{k\bullet} v_t) A_{j\bullet} v_t + \sum_{k=1}^K \zeta_{l,k,k}^\sigma \left[\tau_{k,1} A_{k\bullet} v_t + \tau_{k,2} \kappa(A_{k\bullet} v_t) \right] \\
\tilde{\ell}_{\theta,c_l}(y_t, x_t) &= \sum_{k=1}^K \phi_k(A_{k\bullet} v_t) \times [-A_{k\bullet} D_{c_l}] \\
\tilde{\ell}_{\theta,b_l}(y_t, x_t) &= \sum_{k=1}^K \phi_k(A_{k\bullet} v_t) \times [-A_{k\bullet} D_{b_l} (x_t - (\iota_p \otimes \mu))]
\end{aligned}$$

where $v_t = y_t - c - Bx_t$, $\zeta_{l,k,j}^\alpha := [D_{\alpha_l}(\alpha, \sigma)]_{k\bullet} A_{\bullet j}^{-1}$ with $D_{\alpha_l}(\alpha, \sigma) = \partial A(\alpha, \sigma)/\partial \alpha_l$, $\zeta_{l,k,j}^\sigma := [D_{\sigma_l}(\alpha, \sigma)]_{k\bullet} A_{\bullet j}^{-1}$ with $D_{\sigma_l}(\alpha, \sigma) = \partial A(\alpha, \sigma)/\partial \sigma_l$, $D_{c_l} = \partial c/\partial c_l$, $D_{b_l} = \partial B/\partial b_l$, $\mu = (I - B_1 - \ldots - B_p)^{-1} c$ and $\tau_k := (\tau_{1,k}, \tau_{2,k})'$ is defined as

$$\tau_k := M_k^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \text{where } M_k := \begin{pmatrix} 1 & \mathbb{E}_\theta(A_{k\bullet} v_t)^3 \\ \mathbb{E}_\theta(A_{k\bullet} v_t)^3 & \mathbb{E}_\theta(A_{k\bullet} v_t)^4 - 1 \end{pmatrix}.$$

The derivation of the efficient scores follows by combining results from Amari and Cardoso (1997) and Bickel et al. (1998). For future reference, we partition $\tilde{\ell}_\theta(y_t, x_t) = (\tilde{\ell}_{\theta,\alpha}(y_t, x_t), \tilde{\ell}_{\theta,\beta}(y_t, x_t))$ where $\tilde{\ell}_{\theta,\alpha}(y_t, x_t) = (\tilde{\ell}_{\theta,\alpha_l}(y_t, x_t))_{l=1}^{K_\alpha}$ and also we have $\tilde{\ell}_{\theta,\beta}(y_t, x_t) = \left((\tilde{\ell}_{\theta,\sigma_l}(y_t, x_t))_{l=1}^{K_\sigma}, (\tilde{\ell}_{\theta,c_l}(y_t, x_t))_{l=1}^{K_c}, (\tilde{\ell}_{\theta,b_l}(y_t, x_t))_{l=1}^{K_b} \right)$.

Based on the efficient scores, we define the efficient information matrix for γ by

$$\tilde{I}_\theta \coloneqq \mathbb{E}\,\tilde{\ell}_\theta \tilde{\ell}_\theta' \qquad \text{with partitioning} \qquad \tilde{I}_\theta = \begin{pmatrix} \tilde{I}_{\theta,\alpha\alpha} & \tilde{I}_{\theta,\alpha\beta} \\ \tilde{I}_{\theta,\beta\alpha} & \tilde{I}_{\theta,\beta\beta} \end{pmatrix} . \tag{3.7}$$

With Lemma 3.3.1 and the efficient information matrix in place, we can define the efficient score function for α with respect to β and η. In particular, from Bickel et al. (1998), it follows that this score can be computed by the second projection

$$\tilde{\kappa}_\theta(y_t, x_t) \coloneqq \tilde{\ell}_{\theta,\alpha}(y_t, x_t) - \tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta}\tilde{\ell}_{\theta,\beta}(y_t, x_t) . \tag{3.8}$$

The efficient score $\tilde{\kappa}_\theta$ can be easily computed once the efficient scores for γ (i.e. Lemma 3.3.1) are estimated. The corresponding efficient information matrix is given by

$$\tilde{\mathcal{I}}_\theta \coloneqq \tilde{I}_{\theta,\alpha\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta}\tilde{I}_{\theta,\beta\alpha} . \tag{3.9}$$

Building tests or estimators based on the efficient score function $\tilde{\kappa}_\theta$ is attractive as efficiency results are well established, see Choi et al. (1996), Bickel et al. (1998) and van der Vaart (2002). A crucial difference in our setting is that the efficient information matrix might be singular. For instance, if more than one component of ϵ_t follows an exact Gaussian distribution, $\tilde{\mathcal{I}}_\theta$ is singular, see Lee and Mesters (2021). The singularity plays an important role in the construction of the test statistic below.

3.4 Hypothesis testing in the semi-parametric SVAR

To conduct inference on α without a priori assuming that α is identified, i.e. assuming that at most one component of ϵ_t has a non-Gaussian distribution, we consider testing hypotheses of the following form.

$$H_0 : \alpha = \alpha_0 \,,\ \beta \in \mathcal{B} \,,\ \eta \in \mathcal{H} \qquad \text{against} \qquad H_1 : \alpha \neq \alpha_0 \,,\ \beta \in \mathcal{B} \,,\ \eta \in \mathcal{H} \,. \tag{3.10}$$

The main idea is to consider test statistics whose computation does not require evaluation under the alternative H_1, thus avoiding the need to estimate α. Clearly, based on the trinity of classical tests, the score test is the only viable candidate and we will proceed by constructing score tests in the spirit of Neyman-Rao, but adapted for the semi-parametric setting (e.g. Choi et al., 1996). Such test statistics can then be inverted to yield confidence intervals for α with correct coverage. In our setting, we rely on the efficient score functions for the SVAR model to construct test statistics.

Using Assumption 3.3.2, which implies the existence of a suitable density score estimator,

we can estimate the efficient scores for γ, as defined in Lemma 3.3.1, by

$$\hat{\ell}_\theta(Y_t) = \left(\left(\hat{\ell}_{\theta,\alpha_l}(Y_t)\right)_{l=1}^{K_\alpha}, \left(\hat{\ell}_{\theta,\sigma_l}(Y_t)\right)_{l=1}^{K_\sigma}, \left(\hat{\ell}_{\theta,c_l}(Y_t)\right)_{l=1}^{K_c}, \left(\hat{\ell}_{\theta,b_l}(Y_t)\right)_{l=1}^{K_b} \right)$$

with components

$$\begin{aligned}
\hat{\ell}_{\theta,\alpha_l}(Y_t) &= \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j} \hat{\phi}_k(A_{k\bullet} v_t) A_{j\bullet} V_t + \sum_{k=1}^{K} \zeta^{\alpha}_{l,k,k} \left[\hat{\tau}_{k,1} A_{k\bullet} V_t + \hat{\tau}_{k,2} \kappa(A_{k\bullet} V_t)\right] \\
\hat{\ell}_{\theta,\sigma_l}(Y_t) &= \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{l,k,j} \hat{\phi}_k(A_{k\bullet} V_t) A_{j\bullet} V_t + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k} \left[\hat{\tau}_{k,1} A_{k\bullet} V_t + \hat{\tau}_{k,2} \kappa(A_{k\bullet} V_t)\right] \\
\hat{\ell}_{\theta,c_l}(Y_t) &= \sum_{k=1}^{K} \hat{\phi}_k(A_{k\bullet} V_t) \times [-A_{k\bullet} D_{c,l}] \\
\hat{\ell}_{\theta,b_l}(Y_t) &= \sum_{k=1}^{K} \hat{\phi}_k(A_{k\bullet} V_t) \times [-A_{k\bullet} D_{b,l} (X_t - \bar{X}_n)]
\end{aligned}$$

where $V_t = Y_t - c - BX_t$ and $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$. The estimates for the τ_k's are obtained by replacing the population moments by their sample counterparts in the definition of τ_k:

$$\hat{\tau}_k \coloneqq \hat{M}_k^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \text{where } \hat{M}_k \coloneqq \begin{pmatrix} 1 & \frac{1}{n}\sum_{t=1}^{n} (A_{k\bullet} V_t)^3 \\ \frac{1}{n}\sum_{t=1}^{n} (A_{k\bullet} V_t)^3 & \frac{1}{n}\sum_{t=1}^{n} (A_{k\bullet} V_t)^4 - 1 \end{pmatrix}. \tag{3.11}$$

We obtain the estimates for the efficient scores and information

$$\hat{\ell}_{n,\theta}(Y^n) = \sum_{t=1}^{n} \hat{\ell}_\theta(Y_t, X_t) \qquad \text{and} \qquad \hat{I}_\theta = \frac{1}{n} \sum_{t=1}^{n} \hat{\ell}_\theta(Y_t, X_t) \hat{\ell}'_\theta(Y_t, X_t)\,. \tag{3.12}$$

With the estimates for the efficient scores for γ, we can estimate the efficient score and information for α. We have that

$$\hat{\kappa}_\theta(Y_t, X_t) = \hat{\ell}_{\theta,\alpha}(Y_t, X_t) - \hat{I}_{\theta,\alpha\beta} \hat{I}^{-1}_{\theta,\beta\beta} \hat{\ell}_{\theta,\beta}(Y_t, X_t) \tag{3.13}$$

and

$$\hat{\mathcal{I}}_\theta = \hat{I}_{\theta,\alpha\alpha} - \hat{I}_{\theta,\alpha\beta} \hat{I}^{-1}_{\theta,\beta\beta} \hat{I}_{\theta,\beta\alpha}\,. \tag{3.14}$$

Since the information matrix estimate may be singular, we need to make an adjustment. Specifically, given the truncation rate ν_n defined in Assumption 3.3.2, we define a truncated eigenvalue version of the information matrix estimate as

$$\hat{\mathcal{I}}^t_\theta = \hat{U}_n \hat{\Lambda}_n(\nu_n) \hat{U}'_n\,, \tag{3.15}$$

where $\hat{\Lambda}_n(\nu_n)$ is a diagonal matrix with the ν_n-truncated eigenvalues of $\hat{\mathcal{I}}_\theta$ on the main diagonal and $\hat{U}_n$ is the matrix of corresponding orthonormal eigenvectors. To be specific, let $\{\hat{\lambda}_{n,i}\}_{i=1}^{L}$ denote the non-increasing eigenvalues of $\hat{\mathcal{I}}_\theta$, then the (i,i)th element of $\hat{\Lambda}_n(\nu_n)$

is given by $\hat{\lambda}_{n,i}\mathbf{1}(\hat{\lambda}_{n,i} > \nu_n)$.

Based on this, we define the singularity and identification robust score statistic as follows.

$$\hat{S}_\theta^{SR} := \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{\kappa}_\theta(Y_t, X_t)\right)' \hat{\mathcal{I}}_\theta^{t,\dagger}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{\kappa}_\theta(Y_t, X_t)\right), \tag{3.16}$$

where $\hat{\mathcal{I}}_\theta^{t,\dagger}$ is the Moore-Penrose pseudo-inverse of $\hat{\mathcal{I}}_\theta^t$. The limiting distribution of $\hat{S}_\theta^{SR}$ is characterized in the following theorem, which implies that we can use the estimated rank of $\hat{\mathcal{I}}_\theta^t$ to compute the critical value for $\hat{S}_\theta^{SR}$.

THEOREM 3.4.1: *Let $\hat{\beta}_n$ be a $\sqrt{n}-$consistent estimator of β and let $S_n = n^{-1/2}C\,\mathcal{Z}^{L_2}$ for some $C > 0$ and let $\bar{\beta}_n$ be a discretized version of $\hat{\beta}_n$ which replaces its value with the closest point in S_n. Define $\bar{\theta}_n = (\alpha_0, \bar{\beta}_n, \eta)$, suppose that Assumptions 3.3.1 and 3.3.2 hold, that $\tilde{I}_{\theta_0,\beta\beta}$ is nonsingular and the maps $(\alpha,\sigma) \to [D_{a_l}(\alpha,\sigma)]_{k\bullet}A(\alpha,\sigma)^{-1}_{\bullet j}$ and $(\alpha,\sigma) \to [D_{\sigma_l}(\alpha,\sigma)]_{k\bullet}A(\alpha,\sigma)^{-1}_{\bullet j}$ are Lipschitz continuous. Let $r_n = \operatorname{rank}(\hat{\mathcal{I}}^t_{\bar{\theta}_n})$ and denote by c_n the $1-a$ quantile of the $\chi^2_{r_n}$ distribution, for any $a \in (0,1)$. Then, under H_0*

$$\lim_{n\to\infty} P_{\theta_0}(\hat{S}^{SR}_{\bar{\theta}_n} > c_n) \leq a,$$

with inequality only if $\operatorname{rank}(\tilde{\mathcal{I}}_{\theta_0}) = 0$.

Note the following comments on this theorem.

First, the theorem imposes that β is well identified as $\tilde{I}_{\theta_0,\beta\beta}$ is required to be nonsingular. We do not impose which estimator $\hat{\beta}_n$ should be adopted as the theorem holds for any $\sqrt{n}$-consistent estimator. However, given that the efficient scores of γ need to be computed anyway, it is attractive to rely on one-step efficient estimates for β as discussed in van der Vaart (1998), as this improves the power of the test.

Second, the score statistic is evaluated at the discretized estimator $\bar{\beta}_n$. This is merely a technical device due to Le Cam that simplifies the proof, see Le Cam and Yang (2000, p. 125) or van der Vaart (1998, p. 72-73) for more discussion. Since the discretization can be arbitrarily fine, this has no practical implications.

Third, the eigenvalue truncation rate has little effect on the results and can be viewed as a technical device as well. In practice, we always truncate at machine precision which implies that $\hat{\mathcal{I}}_\theta^{t,\dagger}$ is similar to $\hat{\mathcal{I}}_\theta^{\dagger}$, the generalized inverse of $\hat{\mathcal{I}}_\theta$. Experimenting with different, but small, truncation rates shows that the differences matter little in practice.

Fourth, if $\tilde{\mathcal{I}}_{\theta_0}$ has full rank, the singularity adjusted score statistic is asymptotically equivalent to its non-singular version that is computed with $\hat{\mathcal{I}}^{-1}_{\bar{\theta}_n}$ instead of $\hat{\mathcal{I}}^{t,\dagger}_{\bar{\theta}_n}$, see Lee and Mesters (2021) for a proof. Given this equivalence, it follows from Choi et al. (1996) that non-singularity tests based on $\hat{S}^{SR}_{\bar{\theta}_n}$ are asymptotically uniformly most powerful within

the class of rotation invariant tests (when $L = 1$, the rotational invariance can be dropped for one-sided tests and replaced with unbiasedness for two-sided tests). This implies that asymptotically when testing the hypothesis (3.10), the power of the test is the greatest possible in the class of rotationally invariant tests. This makes tests based on $\hat{S}_{\bar{\theta}_n}$ attractive for scenarios where there is no explicit direction in which one wants to maximize power. When such directions are given, alternative test statistics, also based on the efficient score function, can be considered (e.g. Bickel et al., 2006).

3.5 Finite sample performance

This section presents a collection of simulation studies to evaluate the size and power of the proposed hypothesis testing procedure. We also compare the performance of our method to existing approaches available in the literature.

3.5.1 Size

We start by evaluating the finite sample size of the proposed hypothesis test in the semi-parametric SVAR model. We consider SVAR(p) specifications with $p = 1, 2, 4, 8, 12$ lags, $K = 2, 3$ variables and $T = 200, 500, 1000$. We simulate the SVAR(p) model for ten different choices of the distributions of the structural errors $\epsilon_{k,t}$ with $k = 1, \ldots, K$. The density function of each distribution and their abbreviated names are reported in Appendix D of the paper. For the size simulations, we assume that all components of the error term are distributed identically.

We parameterize the contemporaneous effect matrix as in Example 3.3.1, choosing $A^{-1} = \Sigma^{1/2}(\sigma)R(\alpha)$ where $\Sigma^{1/2}$ is lower triangular and the rotation matrix $R(\alpha)$ is parameterized using a trigonometric transform. In the bivariate case, $K_\alpha = 1$ and we choose $\alpha_0 = \pi/5$ for the data-generating process. In the trivariate SVAR, $K_\alpha = 3$ and we use $\alpha_0 = (3\pi/5, 2\pi/5, -\pi/5)'$. Furthermore, we choose Σ such that the diagonal elements are equal to one, $\sigma_{ii} = 1$ for $i = 1, \ldots, K$, and we set the off-diagonal elements to $\sigma_{ij} = 0.2$ for $i \neq j$. The SVAR coefficient matrices, $A_1, \ldots, A_p$ are generated as diagonal matrices with diagonal elements drawn from a $\mathcal{N}(0, 1)$ distribution.[7] Importantly, even though the data-generating process assumes diagonal coefficient matrices, the test is carried out treating the coefficient matrices as full $K \times K$ matrices. We use 250 burn-in periods to simulate the SVAR(p) model and use $M = 5,000$ Monte Carlo replications to compute the finite-sample rejection rates of the test procedure.

Tables 3.1-3.2 report the empirical rejection frequencies of the semi-parametric score test defined in Section 3.4 for testing the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$. In order to estimate the log density scores, we use the B-spline methodology based on four cubic B-splines. Table 3.1 reports results when estimating the nuisance parameters β using OLS

[7] To ensure stationarity of the SVAR(p) model, the coefficient matrices are sampled until inspection of the roots of the corresponding SVAR(1) companion matrix indicates that the SVAR passes the stationarity condition.

while Table 3.2 reports results from using the one-step efficient estimates for β which update the OLS estimates using one Gauss-Newton iteration. All tests are conducted assuming 5% nominal size.

Table 3.1 shows the empirical rejection frequencies of the proposed testing procedure when using OLS estimates for the nuisance parameters. Focus first on the case of a small sample ($T = 200$) reported in the first panel of the table. For the SVAR(p) with $K = 2$ variables, the size of the test is very close to the nominal size of 5%. Importantly, this holds regardless of the distribution of the structural errors – even when the shocks are normally distributed and hence α is not identified, the test is correctly sized. As the number of parameters in the SVAR increases with the lag size p or the number of variables K, the rejection rates increase and the test starts to over-reject in small samples. For an increase in the number of lags, rejection rates only increase slightly, but when the number of variables increases, the number of parameters grows quadratically and hence rejection rates show a more substantial increase. For the trivariate SVAR with twelve lags, the test shows pronounced over-rejections in small samples. However, importantly, as a comparison with the panels reporting rejection rates for a medium sample size ($T = 500$) and large sample size ($T = 1,000$) shows, size distortions quickly disappear as the sample grows and the rejection frequencies converge to the nominal size of the test. Thus, even in the case of an SVAR with a larger lag length, the testing procedure gives correct inference, as long as the sample is not too small.

Table 3.2 reports the empirical rejection frequencies for the same simulations when one-step efficient estimates are used for the nuisance parameters. The one-step efficient estimates of β are obtained by updating the OLS estimates of the nuisance parameters β towards the efficient estimates by one Gauss-Newton iteration. The procedure is computationally efficient as it does not require full numerical optimization routines to obtain estimates of the nuisance parameters. Comparing the rejection rates in Table 3.2 with those reported in the case of OLS estimates of the nuisance parameters in Table 3.1, shows that using the one-step estimates yields substantial improvements in the size of the test in small samples, especially when the number of lags is large. For example, for the case of an SVAR with three variables and 12 lags, the size of the rejection rates are very close to the nominal size of 5%. As the sample size grows, the difference between the two approaches becomes less pronounced and the procedures yield comparable rejection rates. For medium and large samples, either of the procedures can result in rejection rates closer to the nominal size, depending on the number of lags, the number of variables and the distribution of the structural errors that generated the data.

Table 3.1: Empirical rejection frequencies using OLS estimates

K	p	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$T = 200$											
2	1	0.049	0.055	0.048	0.045	0.043	0.041	0.046	0.045	0.039	0.042
2	2	0.055	0.051	0.049	0.049	0.044	0.037	0.048	0.047	0.045	0.050
2	4	0.057	0.059	0.056	0.052	0.046	0.029	0.052	0.056	0.050	0.047
2	8	0.086	0.069	0.066	0.060	0.060	0.039	0.065	0.074	0.062	0.062
2	12	0.086	0.081	0.074	0.075	0.071	0.041	0.082	0.092	0.073	0.084
3	1	0.072	0.066	0.071	0.085	0.057	0.048	0.047	0.052	0.060	0.045
3	2	0.074	0.084	0.079	0.082	0.061	0.042	0.049	0.055	0.059	0.047
3	4	0.100	0.102	0.091	0.109	0.080	0.039	0.069	0.080	0.081	0.064
3	8	0.166	0.151	0.145	0.135	0.122	0.060	0.135	0.151	0.132	0.124
3	12	0.228	0.225	0.216	0.188	0.178	0.092	0.232	0.249	0.179	0.218
$T = 500$											
2	1	0.052	0.046	0.043	0.040	0.036	0.035	0.042	0.043	0.042	0.041
2	2	0.046	0.049	0.047	0.045	0.037	0.035	0.039	0.040	0.041	0.043
2	4	0.045	0.044	0.047	0.044	0.039	0.031	0.042	0.044	0.043	0.046
2	8	0.055	0.053	0.046	0.043	0.041	0.029	0.049	0.052	0.047	0.049
2	12	0.061	0.059	0.054	0.044	0.046	0.031	0.055	0.060	0.051	0.056
3	1	0.046	0.052	0.049	0.055	0.039	0.043	0.044	0.038	0.041	0.037
3	2	0.053	0.055	0.048	0.056	0.040	0.039	0.046	0.043	0.045	0.042
3	4	0.065	0.058	0.061	0.054	0.048	0.031	0.043	0.049	0.046	0.043
3	8	0.080	0.078	0.073	0.071	0.055	0.032	0.061	0.069	0.063	0.055
3	12	0.106	0.097	0.098	0.087	0.070	0.028	0.081	0.100	0.077	0.067
$T = 1,000$											
2	1	0.045	0.040	0.035	0.030	0.039	0.032	0.042	0.042	0.039	0.037
2	2	0.048	0.040	0.037	0.035	0.038	0.037	0.043	0.044	0.040	0.037
2	4	0.042	0.042	0.045	0.036	0.038	0.030	0.045	0.047	0.043	0.047
2	8	0.048	0.053	0.045	0.037	0.042	0.030	0.049	0.048	0.042	0.042
2	12	0.048	0.049	0.048	0.037	0.040	0.025	0.050	0.051	0.044	0.046
3	1	0.046	0.043	0.038	0.039	0.036	0.036	0.037	0.037	0.039	0.041
3	2	0.045	0.042	0.042	0.043	0.036	0.034	0.036	0.039	0.038	0.042
3	4	0.051	0.048	0.049	0.042	0.038	0.034	0.042	0.040	0.039	0.043
3	8	0.059	0.050	0.056	0.046	0.044	0.029	0.048	0.050	0.047	0.044
3	12	0.060	0.058	0.064	0.059	0.039	0.027	0.050	0.052	0.053	0.048

Notes: The table reports empirical rejection frequencies for the semi-parametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ in the K-variable SVAR(p) model with nominal size 5%. The nuisance parameters β are estimated by OLS. The columns correspond to different choices for the distributions of the structural shocks, $\epsilon_{k,t}$ for $k = 1, \ldots, K$. The distributions are reported in Appendix D. Rejection rates are computed based on $M = 5,000$ Monte Carlo replications.

Table 3.2: Empirical rejection frequencies using one-step efficient estimates

K	p	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$T = 200$											
2	1	0.071	0.074	0.066	0.059	0.064	0.051	0.052	0.053	0.048	0.051
2	2	0.068	0.067	0.058	0.052	0.060	0.055	0.050	0.050	0.051	0.056
2	4	0.066	0.062	0.058	0.058	0.061	0.053	0.047	0.049	0.053	0.048
2	8	0.057	0.057	0.055	0.050	0.062	0.044	0.045	0.048	0.052	0.047
2	12	0.058	0.054	0.059	0.066	0.083	0.050	0.052	0.051	0.055	0.049
3	1	0.095	0.096	0.099	0.089	0.081	0.062	0.059	0.063	0.078	0.059
3	2	0.093	0.097	0.100	0.092	0.080	0.079	0.057	0.067	0.069	0.055
3	4	0.091	0.099	0.102	0.117	0.091	0.072	0.057	0.069	0.078	0.048
3	8	0.057	0.055	0.063	0.073	0.057	0.055	0.031	0.032	0.051	0.028
3	12	0.043	0.042	0.043	0.059	0.049	0.042	0.038	0.040	0.042	0.039
$T = 500$											
2	1	0.068	0.069	0.061	0.058	0.057	0.051	0.050	0.055	0.052	0.051
2	2	0.065	0.063	0.065	0.055	0.059	0.047	0.047	0.047	0.052	0.049
2	4	0.060	0.061	0.058	0.051	0.054	0.049	0.045	0.046	0.052	0.050
2	8	0.062	0.060	0.054	0.050	0.057	0.047	0.047	0.046	0.049	0.049
2	12	0.065	0.062	0.056	0.052	0.061	0.044	0.050	0.055	0.052	0.052
3	1	0.085	0.081	0.082	0.071	0.069	0.067	0.057	0.055	0.061	0.053
3	2	0.092	0.085	0.074	0.072	0.065	0.066	0.053	0.057	0.062	0.060
3	4	0.093	0.087	0.084	0.074	0.072	0.060	0.058	0.066	0.065	0.054
3	8	0.093	0.095	0.094	0.085	0.069	0.062	0.053	0.059	0.072	0.052
3	12	0.100	0.094	0.086	0.092	0.087	0.060	0.051	0.057	0.077	0.044
$T = 1,000$											
2	1	0.064	0.056	0.057	0.050	0.063	0.048	0.050	0.052	0.051	0.044
2	2	0.068	0.058	0.053	0.052	0.059	0.052	0.051	0.055	0.052	0.047
2	4	0.063	0.058	0.062	0.053	0.057	0.044	0.051	0.052	0.055	0.055
2	8	0.062	0.068	0.061	0.049	0.059	0.048	0.051	0.052	0.049	0.048
2	12	0.060	0.063	0.061	0.051	0.057	0.041	0.049	0.049	0.055	0.054
3	1	0.078	0.076	0.067	0.058	0.064	0.053	0.051	0.058	0.055	0.055
3	2	0.078	0.076	0.068	0.064	0.065	0.051	0.055	0.058	0.057	0.057
3	4	0.089	0.079	0.077	0.063	0.060	0.053	0.049	0.053	0.055	0.057
3	8	0.086	0.076	0.078	0.068	0.070	0.061	0.058	0.064	0.062	0.053
3	12	0.091	0.084	0.088	0.074	0.067	0.063	0.060	0.061	0.069	0.051

Notes: The table reports empirical rejection frequencies for the semi-parametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ in the K-variable SVAR(p) model with nominal size 5%. The nuisance parameters β are estimated by the one-step efficient procedure. The columns correspond to different choices for the distributions of the structural shocks, $\epsilon_{k,t}$ for $k = 1, \ldots, K$. The distributions are reported in Appendix D. Rejection rates are computed based on $M = 5,000$ Monte Carlo replications.

3.5.2 Comparison to alternative approaches

Next, we compare the size of the proposed testing procedure to a variety of different approaches based on (pseudo-) maximum likelihood and the generalized method of moments. In particular, we consider standard maximum likelihood Lagrange-Multiplier (LM) and Likelihood Ratio (LR) tests based on a standardized student-t density for ϵ_k[8]. Note that when the structural errors that generated the data are distributed according to a t-distribution, these tests correspond to exact maximum likelihood tests. For all remaining distributions, the standard maximum likelihood tests are misspecified. In addition to the parametric MLE tests, we consider the pseudo-maximum-likelihood procedure of Gouriéroux et al. (2017) and the GMM-based approach of Lanne and Luoto (2020). These procedures have been employed in various empirical studies to test hypotheses in non-Gaussian SVAR models. Our aim in this section is to study how such methods perform in the case of "strong" and "weak" non-Gaussian distributions and to compare their finite-sample size to that of our proposed method. It is important to note that, unlike our approach, none of the alternative methods listed above are designed to be robust against cases where the true densities are "close" to Gaussian. In fact, previous simulation studies have documented size distortions in those cases (see e.g. Gouriéroux et al., 2017; Lanne and Luoto, 2020).

To evaluate the finite-sample performance, we focus on an SVAR(1) model with $K = 2$ variables and a sample size of $T = 500$. We use the same parameterization and parameter values as described in the previous subsection to generate the data. However, different to the previous simulation study evaluating the size of the score test, we report results both for the case where the structural errors $\epsilon_{1,t}, \epsilon_{2,t}$ are identically distributed, but also for the case where the first error is fixed to have a Gaussian distribution while the distribution of the second structural error varies. Note that in the latter case, theoretically non-Gaussianity can still be used to identify the parameters of the SVAR if the second structural error does not follow a Gaussian distribution. However, we suspect identification to be weaker in this case.

We implement the tests as follows. For the semi-parametric score test, we use one-step estimates of the nuisance parameters, as in Table 3.2. The density scores are approximated using the same procedure described in the previous subsection. For the alternative methods, we do not estimate the nuisance parameters, but rather use the true nuisance parameter values that were used to generate the data. We do so to have a strong benchmark against which we evaluate the performance of our method and to ensure that any differences in the finite-sample performance of the tests does not originate in the procedure that is used to estimate the nuisance parameters. Importantly, note that since the true nuisance parameters are always unknown in practice and hence need to be estimated, this constitutes an "oracle" benchmark which is expected to perform better than when nuisance parameters

[8]Note that the standardized student-t density requires the degrees of freedom parameter, ν, to be larger than 4 as otherwise the density has infinite kurtosis.

are estimated.[9]

Since the nuisance parameters are fixed at their true values, implementation of the maximum-likelihood LM test amounts to estimating the restricted model where α is fixed under the null hypothesis and only requires estimating the degrees of freedom of the standardized student-t distributions, ν_1, ν_2. To compute the LR test, we also need to estimate the unrestricted model in which α is an additional parameter to be estimated. For the numerical optimization, we use a Quasi-Newton algorithm with initial values of $\tilde{v} = 8$ and $\tilde{\alpha} = \pi/4$.

To implement the procedure of Gouriéroux et al. (2017), we first obtain standardized residuals of the reduced form VAR errors using the true values for the nuisance parameters. We then compute the pseudo maximum likelihood LR test using a standardized student-t distribution with 5 degrees of freedom. For the unrestricted model, we estimate α using a Quasi-Newton optimization algorithm with initial value of $\tilde{\alpha} = \pi/4$.

Finally, to implement the GMM approach of Lanne and Luoto (2020), we use the two-step GMM estimator based on the following moment conditions: (i) $\mathbb{E}[\epsilon_t \otimes x_{t-1}] = 0$ where $x_{t-1} \equiv (1, Y'_{t-1})'$, (ii) $\mathbb{E}[\epsilon^2_{k,t}] = 1$ for $k = 1, 2$, (iii) $\mathbb{E}[\epsilon_{1,t}\epsilon_{2,t}] = 0$ and (iv) the asymmetric co-kurtosis condition $\mathbb{E}[\epsilon^3_{1,t}\epsilon_{2,t}] = 0$. The same set of moment conditions was used in Lanne and Luoto (2019). In our SVAR(1) model, this yields ten moment conditions which just-identify the ten parameters of the SVAR model.[10]

Table 3.3 reports the results of the simulation study and compares the size of the alternative testing procedures to the size of the score test ($\hat{S}$). The first panel reports the case where the two structural errors, $\epsilon_{1,t}, \epsilon_{2,t}$ are drawn from the same (non-Gaussian) distribution while the second panel reports the results where $\epsilon_{1,t}$ is fixed to have a Gaussian distribution. First, note that the rejection rates for the score test ($\hat{S}$) in the case of identically distributed shocks are the same rejection rates reported in Table 3.2 and are close to the nominal size of 5%, regardless of the distribution of the structural shocks. Inspecting the second panel of the table, we note that the performance of the score test does not deteriorate when the first structural error is Gaussian; the rejection rates continue to be close to the nominal size of 5% regardless of the distribution of the second error.

Second, we inspect the performance of the parametric maximum-likelihood tests (LM and LR). We find that the tests perform well, as long as the densities are not "heavily" misspecified. In particular, for the student-t densities and the unimodal densities, either the LR or the LM test have a size close to the nominal level. This makes sense as the parameter α is fixed under the null and thus no identification problem arises. However, when the density is strongly misspecified (e.g. bimodal, separated bimodal or trimodal), the

[9] In a simulation not reported in this paper, we verify that this is indeed the case and that performance of the alternative procedures deteriorates when estimating nuisance parameters e.g. by OLS.

[10] We also ran simulations using an additional over-identifying symmetric co-kurtosis condition; however this gave worse finite-sample performance than the just-identified case.

Table 3.3: Empirical rejection frequencies for alternative tests

Test	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$\epsilon_{1,t}, \epsilon_{2,t}$ *identically distributed*										
$\hat{S}$	0.068	0.069	0.061	0.058	0.057	0.051	0.050	0.055	0.052	0.051
LM	0.166	0.117	0.100	0.088	0.078	0.083	0.239	0.261	0.115	0.209
LR	0.054	0.082	0.069	0.053	0.064	0.001	0.000	0.000	0.083	0.000
LR^{GM}	0.590	0.282	0.166	0.057	0.147	0.007	1.000	1.000	0.388	1.000
LR^{LL}	0.115	0.123	0.126	0.174	0.126	0.101	0.056	0.062	0.123	0.062
$\epsilon_{1,t}$ *Gaussian*										
$\hat{S}$	0.068	0.067	0.068	0.059	0.063	0.048	0.049	0.052	0.054	0.052
LM	0.166	0.132	0.117	0.103	0.103	0.083	0.254	0.233	0.143	0.276
LR	0.054	0.078	0.075	0.042	0.073	0.003	0.018	0.018	0.077	0.012
LR^{GM}	0.590	0.408	0.325	0.142	0.291	0.061	0.999	0.990	0.497	0.999
LR^{LL}	0.115	0.110	0.100	0.092	0.101	0.092	0.081	0.088	0.119	0.071

Notes: The table reports empirical rejection frequencies for tests of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$. $\hat{S}$ denotes the semi-parametric score test using one-step efficient estimates for β, LM denotes the parametric score test based on a standardized student-t density, LR denotes the associated parametric likelihood-ratio test, LR^{GM} denotes the pseudo maximum-likelihood-ratio test of Gouriéroux et al. (2017) and LR^{LL} denotes the GMM-based test of Lanne and Luoto (2020). The columns correspond to different choices for the distributions of the structural shocks, $\epsilon_{k,t}$ for $k = 1, \ldots, K$. The distributions are reported in Appendix D. Rejection rates are computed based on $M = 5,000$ Monte Carlo replications.

LM test substantially over-rejects and the LR test is strongly conservative. Further, when we move to the case where the first shock is Gaussian, most rejection rates of the LM and LR tests increase further.

Next, we inspect the performance of the pseudo maximum likelihood approach of Gouriéroux et al. (2017) (LR^{GM}). We find that the test shows substantial over-rejections, unless the distribution is close to the pseudo distribution. In fact, only in the case of the $t(5)$ distribution, where the approach is an exact maximum likelihood test, does the test yield a rejection rate close to the nominal level. Interestingly, for the kurtotic unimodal distribution (KU), the test is strongly conservative. Further, when we move to the case where the first component is fixed at a Gaussian distribution, the performance of the test significantly deteriorates for most of the densities. Finally, we evaluate the GMM-based approach of Lanne and Luoto (2020). Compared to the pseudo-maximum likelihood procedure, the GMM approach performs much better and yields rejection rates closer to the nominal level. Compared to the maximum likelihood tests, the procedure yields more accurate inference for the bimodal or trimodal densities, but leads to over-rejections for the other distributions.

To summarize, most of the alternative procedures lead to incorrect inference if the distribution of the structural shocks is not "sufficiently" non-Gaussian. Furthermore, the identity of the best-performing alternative procedure crucially depends on which non-Gaussian distribution generated the data. In contrast, the semi-parametric score test proposed in this paper gives correct inference regardless of the distribution of the structural

errors and whether one or both errors are non-Gaussian.

3.5.3 Power

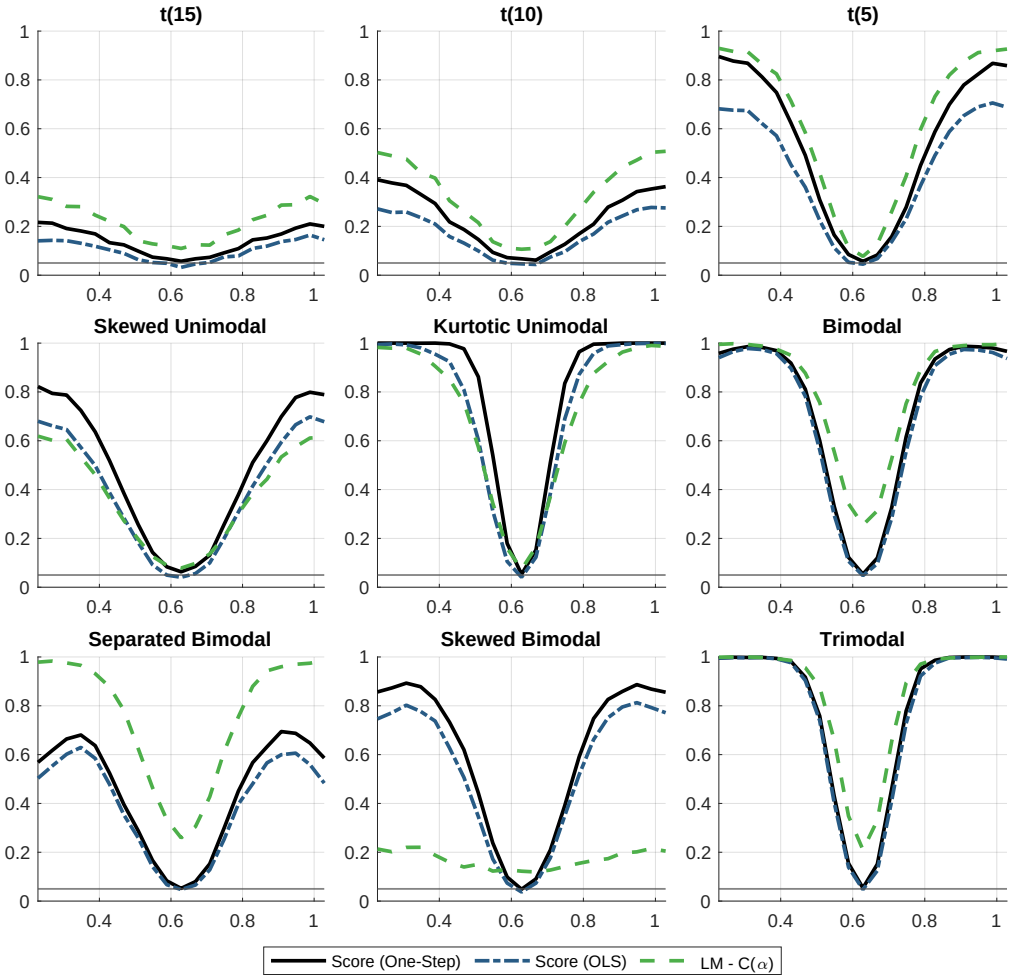
Finally, we study the power of the proposed procedure and compare it to the power of the parametric maximum likelihood LM test described in the previous subsection. The LM test is the natural counterpart for the first four densities considered in the simulations and, together with the LR test, is the only procedure that controlled size comparably in some of the size simulations reported in the previous subsection.

To evaluate the finite-sample power of the tests, we focus on an SVAR(1) model with $K = 2$ variables and a sample size of $T = 500$. We use the same parameterization and parameter values as used in the other simulations to generate the data. As in the size simulations for the score test, we report results for the case where the structural errors $\epsilon_{1,t}, \epsilon_{2,t}$ are identically distributed. We compare three tests of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ with 5% nominal size: (i) the semi-parametric score test using OLS estimates for the nuisance parameters, (ii) the semi-parametric score test using one-step efficient estimates for the nuisance parameters and (iii) the parametric maximum-likelihood LM test. All tests are implemented as described in the previous simulation studies.

Figure 3.2 reports the raw (i.e not size-adjusted) power for the semi-parametric score test using one-step nuisance parameter estimates (black solid line), the semi-parametric score test using OLS nuisance parameter estimates (dot dashed blue line) and the parametric maximum-likelihood LM test (green dashed line). Each panel of the figure corresponds to a different choice for the distributions of the structural shocks. We use the same distributions as in the simulation studies described earlier. The x-axis of each plot corresponds to different alternatives for α around $\alpha_0 = \pi/5$ that were used to generate the data. Each point on the curve is based on $M = 2,500$ simulations. We first inspect the power of the semi-parametric score tests. Comparing the power curves for the different distributions of the structural shocks, we find that power of the test substantially increases with the distance to the Gaussian distribution. In particular, for the kurtotic unimodal density, the bimodal density or the trimodal density, the rejection rates quickly increase with α departing from its value under the null hypothesis. For the densities which are very similar to a Gaussian distribution ($t(15)$ and $t(10)$), power is much smaller, but still non-negligible. Comparing the two variants of the semi-parametric score test, we note that using the one-step efficient estimates yields generally larger power than using the OLS nuisance parameter estimates.

Finally, comparing the power of the semi-parametric test to the parametric LM test, we note that for the distributions where the tests are similarly sized, the power of the parametric test is larger when the density is correctly specified ($t(15)$, $t(10)$ and $t(5)$). For all other density choices, the semi-parametric test convincingly outperforms its parametric counterpart or the parametric test shows substantial size distortions so that a comparison is infeasible. Finally, it is important to note that even under non-Gaussianity, α is only identified up to scale and

Figure 3.2: Power in the SVAR(1) model



Notes: The figure reports unadjusted empirical power curves for tests of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$. The x-axis corresponds to different alternatives for α around $\alpha_0 = \pi/5$. The solid black line reports the empirical rejection frequencies for the semi-parametric score test using one-step nuisance parameter estimates, the blue dot dashed line corresponds to the semi-parametric score test using OLS nuisance parameter estimates and the green dashed line denotes the parametric LM test based on standardized student-t densities. Rejection frequencies are computed using $M = 2,500$ Monte Carlo replications.

permutation of the columns. Hence for $\alpha \in [0, 2\pi]$, there are multiple optimal points. This is the reason why the figures show a decrease in power for alternatives that are far away from α_0, but close to the next value of α under which the null hypothesis holds. In summary, we conclude that the semi-parametric score test has adequate power even when compared to correctly specified parametric tests.

3.6 Application

In this section, we revisit identification of labor supply and labor demand elasticities in a frequentist version of the model of the U.S. labor market from Baumeister and Hamilton (2015). We employ our semi-parametric approach for SVAR inference to construct identified regions for the labor supply and labor demand elasticities. We also construct identification-robust confidence bands for the impulse responses based on our inference approach.

Baumeister and Hamilton (2015) study a simple bivariate SVAR model of the U.S. labor market using a Bayesian approach that includes carefully motivated priors on the short-run labor supply and demand elasticities based on estimates in the micro-econometric and macroeconomic literature. In addition, they impose a long-run restriction on the effect of labor demand shocks on employment. Their posterior densities provide evidence in favor of a negative demand elasticity and favor estimates of the short-run supply elasticity that fall into the lower range of the micro-econometric literature (below 0.5) rather than values around or above 1 that are often used in macroeconomic studies. Recently, Lanne and Luoto (2019) revisited this specification and criticized the use of the long-run restriction. Instead, they proposed to exploit non-Gaussianity of the data to identify the labor supply and demand elasticities by means of the GMM approach of Lanne and Luoto (2020) which relies on moment conditions that are informative under non-Gaussian structural errors. Using non-Gaussianity identification, they find a significant and persistent effect of the labor demand shock on employment and reject the long-run restriction of Baumeister and Hamilton (2015).

Similar to Lanne and Luoto (2019), in this section, we revisit the bivariate SVAR relaxing the prior specifications in Baumeister and Hamilton (2015) through the use of non-Gaussianity identification. However, we use the semi-parametric approach proposed in this paper which is robust to weak non-Gaussianity. This is particularly important, since both our and Lanne and Luoto (2020)'s simulation studies show that the adopted GMM approach can lead to invalid inference when the structural shocks only have weak non-Gaussian distributions.

The specification of Baumeister and Hamilton (2015) is a bi-variate SVAR(p) model of the U.S. labor market for $Y_t = (\Delta w_t, \Delta \eta_t)'$, where Δw_t is the growth rate of real compensation per hour and $\Delta \eta_t$ is the growth rate of total U.S. employment:

$$B_0 Y_t = \mu + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + D^{1/2}\epsilon_t, \qquad B_0 \equiv \begin{bmatrix} -\beta^d & 1 \\ -\beta^s & 1 \end{bmatrix}, \quad D \equiv \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

In the model, β^d is the short-run wage elasticity of demand, and β^S is the short-run wage elasticity of supply. The number of lags used in the SVAR is $p = 8$ and sign restrictions imposed on the supply and demand elasticities require that $\beta^s > 0$ and $\beta^d < 0$.

Rewriting their model using our notation, we get

$$Y_t = c + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + A^{-1}(\alpha, \sigma)\, \epsilon_t, \qquad A^{-1} \equiv \begin{bmatrix} \frac{\sigma_1}{\beta^s - \beta^d} & \frac{\sigma_2}{\beta^s - \beta^d} \\ \frac{\sigma_1 \beta^s}{\beta^s - \beta^d} & \frac{\sigma_2 \beta^d}{\beta^s - \beta^d} \end{bmatrix},$$

where the structural shocks, $\epsilon_{1,t}, \epsilon_{2,t}$ are independent and have unit variance. Note that the impact multiplier matrix, A^{-1} depends on four parameters: The standard deviations of the structural shocks, $\sigma = (\sigma_1, \sigma_2)'$ with $\sigma_k > 0$, that can be identified from second moments of $\{Y_t\}$, and the elasticity parameters $\alpha = (\beta^s, \beta^d)'$ that are only identified if the demand and supply shocks are (sufficiently) non-Gaussian.

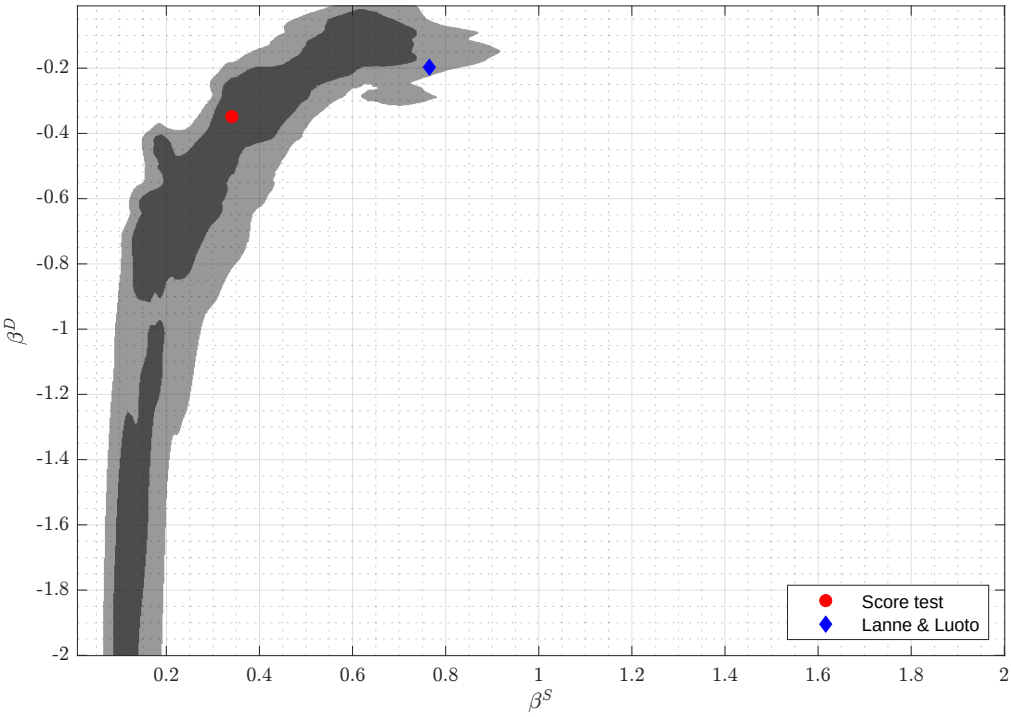
3.6.1 Confidence regions for labor demand and supply elasticities

We start our analysis by constructing confidence regions for the implied elasticity parameters of the model. Our approach relies on Theorem 3.4.1 as it effectively computes the bands for each vector α_0 that is (i) not rejected by the score test and (ii) satisfies the sign restrictions that pin down the permutation and sign of A. Specifically, we construct a grid of parameter values that satisfy the sign restrictions $\beta^d < 0$ and $\beta^s > 0$ and record the region of parameter values (β^d, β^s) that satisfies $\hat{S}_\theta^{SR} > c_n$ where c_n is the critical value for the $\chi^2_{r_n}$ distribution associated with the desired significance level. To construct the grid, we use 1,000,000 equally spaced combinations of (β^s, β^d) in the interval $(0, 2] \times [-2, 0)$.

Figure 3.3 reports the joint confidence regions for the labor demand (β^d) and labor supply (β^s) parameters of the model for a 95% significance level (light gray) and a 67% significance level (dark gray). The red circle indicates the values associated with the minimum score statistic while the blue diamond marks the estimates of labor demand and supply elasticities reported in Lanne and Luoto (2019). As the figure shows, the short-run supply elasticity is reasonably well identified from non-Gaussianity: The hypothesis $\beta^s = 0$ can be rejected at the 95% level and the 67% confidence region contains elasticities $\beta^S \in (.1, .8)$, depending on the value of β^d. Furthermore, for sufficiently negative demand elasticities, the supply elasticity is precisely identified and falls into the range supported by the posterior densities in Baumeister and Hamilton (2015). In contrast, the demand elasticity is poorly identified by non-Gaussianity: The confidence region is wide in the direction of β^d and spans almost all values in the inspected grid.

Our results differ significantly from Lanne and Luoto (2019). While the labor demand elasticity implied by the minimum score statistic is reasonably close to their estimate, the corresponding labor supply elasticity is much smaller. In addition, they report substantially narrower confidence intervals for the elasticity parameters. In fact, our identification-robust confidence regions are consistent with the posterior distributions reported in Baumeister and Hamilton (2015) which place most posterior weight on a supply parameter between zero and 0.5, while supporting a large range of negative demand elasticities.

Figure 3.3: Confidence regions for labor demand and supply elasticities



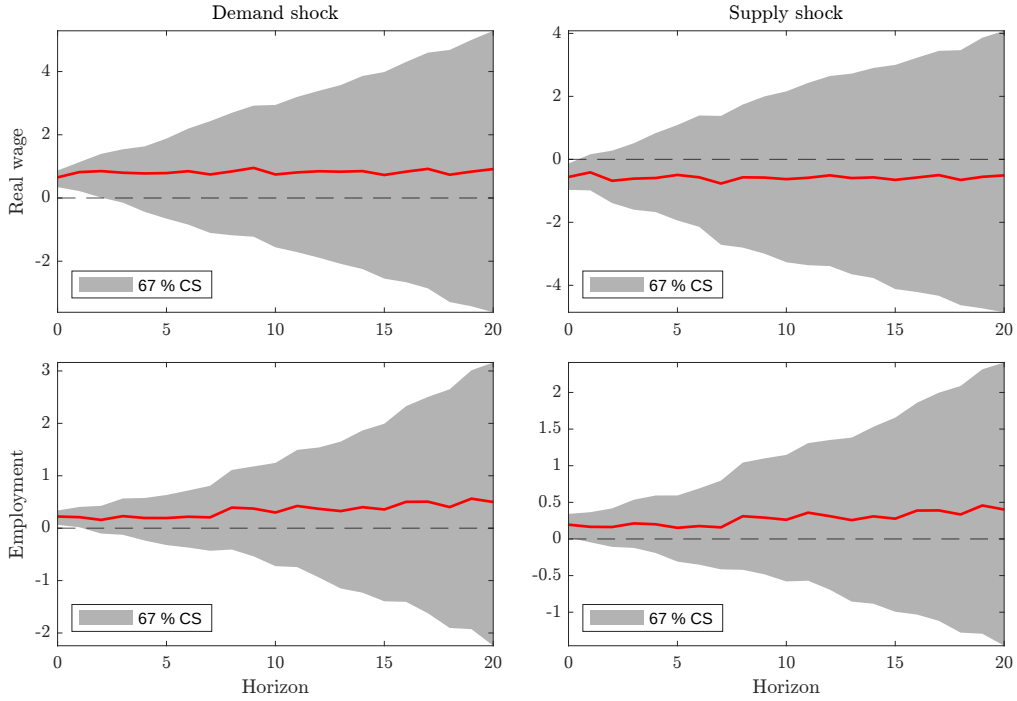
Notes: Confidence regions for labor demand and supply elasticities (95% in light gray, 67% in dark gray) obtained using 1,000,000 equally-spaced grid points for $(\beta^s, \beta^d) \in (0, 2] \times [-2, 0)$. The red circle denotes the elasticities associated with the min. score statistic, the blue diamond denotes the estimates of Lanne and Luoto (2019).

3.6.2 Impulse responses for labor demand and supply shocks

After having established that non-Gaussianity is not sufficient to precisely identify both labor supply and demand elasticities, we turn to evaluating the consequences for identifying the impulse responses associated with labor demand and supply shocks.

The semi-parametric inference approach proposed in this paper can be used to construct identification-robust confidence bands for the structural impulse responses of the SVAR. Similar to our approach to constructing confidence regions discussed in the previous subsection, we rely on Theorem 3.4.1 to compute the bands for each vector α_0 that is (i) not rejected by the score test for a given significance level and (ii) satisfies the sign restrictions that pin down the permutation and sign of A. For each vector α_0 satisfying these conditions, we then compute and store the upper and lower pointwise asymptotic confidence intervals of the impulse responses based on the one-step efficient estimate for the nuisance parameter of the test, β. The final confidence bands are then simply taken as the widest intervals within the set of admissible bands for any level of confidence. Appendix C provides details on the algorithm that is used to compute confidence bands for the impulse responses and reports the necessary formulae to implement the method. We use the approach to construct robust confidence intervals for the impulse responses for labor supply and labor demand shocks. As in the previous subsection, we construct a grid of 1,000,000 equally spaced grid points

Figure 3.4: IRF confidence bands for labor demand and supply shocks



Notes: Impulse responses for labor supply and labor demand shocks with 67% identification-robust confidence bands based on 1,000,000 grid points for $(\beta^s, \beta^d,) \in (0, 2] \times [-2, 0)$.

(β^s, β^d) in the interval $(0, 2] \times [-2, 0)$.

Figure 3.4 shows the 67% confidence intervals (gray shaded area) for the impulse responses together with the impulse response associated with the minimum score statistic (red solid line). Comparing the impulse response implied by the minimum score statistic to the IRF estimated by Lanne and Luoto (2019), we note that the impulse responses are overall very similar and show long and persistent responses to the supply and demand shocks; only the response of employment to a demand shock has a more pronounced hump shape in Lanne and Luoto (2019) than what is implied by the minimum score IRF estimate. Further, focusing on the impact effect, at a 67% level, a demand shock has significant and positive effects on both real wages and employment, while a supply shock has a significant negative effect on the wage and a significantly positive effect on employment. These results are in accordance with the posterior distributions reported by Baumeister and Hamilton (2015) and are similar to the ones reported by Lanne and Luoto (2019). However, inspecting the dynamic responses to the labor supply and demand shocks, there are substantial differences to the results of Lanne and Luoto (2019). Specifically, they find a strong and significant dynamic response of both the real wage and employment to the labor demand shock, inconsistent with Baumeister and Hamilton (2015)'s prior on the restriction that the long-run labor demand elasticity is zero. In contrast to their findings, our confidence intervals, using an approach robust to weak non-Gaussianity, show that the data cannot rule out that the long-run response of either variable to the demand shock is zero and hence imply that

the long-run restriction cannot be rejected solely on the basis of non-Gaussianity.

3.7 Conclusion

This paper develops robust inference methods for structural vector autoregressive (SVAR) models that are identified using non-Gaussian error distributions. We treat the SVAR model as a semi-parametric model where the densities of the structural errors form the non-parametric part and use hypothesis testing to conduct inference on the possibly weakly identified or non identified parameters of the SVAR, using a semi-parametric equivalent of the Neyman-Rao score statistic. We assess the finite-sample performance of our method in a large simulation study and find that the empirical rejection frequencies of the semi-parametric score test are always close to the nominal size, regardless of the true distribution of the errors. For the power of the score test, we find that it comes close to the power of a correctly specified maximum likelihood test using the true (unknown) distributions of the structural errors. Finally, we employ the proposed approach in an empirical study that revisits the identification of supply and demand elasticities in the U.S. labor market as in Baumeister and Hamilton (2015), but using a frequentist procedure. We show how our approach can be used to construct confidence regions for the labor supply and demand elasticities as well as identification-robust confidence bands for the impulse responses. Our confidence regions for the demand and supply elasticities are consistent with the posterior distributions reported in Baumeister and Hamilton (2015). Further, we find that a non-Gaussianity identification strategy is insufficient to precisely identify the economy's response to labor supply and labor demand shocks.

Appendices

A Proofs

The proof of Theorem 3.4.1 is structured as follows. We first state an intermediate lemma with convergence results based on which the theorem can be proven. The proof for the lemma is included below.

LEMMA A.1: *Let $\gamma_0 = (\alpha_0, \beta)$ and $\theta_0 = (\alpha_0, \beta, \eta)$ for any $(\beta, \eta) \in \mathcal{B} \times \mathcal{H}$. Additionally, let $\gamma_n = \{(\alpha_0, \beta_n)\}_{n\in\mathbb{N}}$ be a deterministic sequence such that $\sqrt{n}(\gamma_n - \gamma_0) = O(1)$ and define $\theta_n = (\gamma_n, \eta)$ for each $n \in \mathbb{N}$. Then, under the conditions of Theorem 3.4.1, we have that*

1. $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{\ell}_{\theta_0}(Y_t, X_t) \rightsquigarrow Z \sim \mathcal{N}(0, \tilde{I}_{\theta_0})$ *under* P_{θ_0}
2. *We have that*
$$\frac{1}{n} \sum_{t=1}^{n} \left(\hat{\ell}_{\theta_n}(Y_t, X_t) - \tilde{\ell}_{\theta_n}(Y_t, X_t) \right) = o_{P_{\theta_n}}(n^{-1/2})$$
3. $P_{\theta_n}\left(\|\hat{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_2 < \nu_n \right) \to 1$ *where ν_n is defined in Assumption 3.3.2 .*
4. *We have that*
$$\int \left\| \tilde{\ell}_{\theta_n} \mathrm{d}P_{\theta_n}^{1/2} - \tilde{\ell}_{\theta_0} \mathrm{d}P_{\theta_0}^{1/2} \right\|^2 \mathrm{d}\mu \to 0.$$

Note: The following proof of Theorem 3.4.1 uses the parameterization $A^{-1} = \Sigma^{1/2} R(\alpha)$ (see Example 3.3.1) and assumes that Lemma 7.3 of van der Vaart (2002) holds, given Lemma A.1 part 4.

Proof of Theorem 3.4.1. We break the proof into steps. First, we show that under Lemma A.1 we have
$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{\kappa}_{\theta_n}(Y_t, X_t) \rightsquigarrow Z \sim \mathcal{N}(0, \tilde{\mathcal{I}}_{\theta_0}) \tag{17}$$
under $P_0 \equiv P_{\theta_0}$, where θ_n is defined in Lemma A.1, and
$$\sqrt{n}\mathbb{P}_n \left[\hat{\ell}_{\theta_n} - \tilde{\ell}_{\theta_n} \right] \xrightarrow{P_0} 0 \quad \text{and} \quad P_0 \left(\left\| \hat{I}_{\theta_n} - \tilde{I}_{\theta_0} \right\| < \nu_n \right) \to 1. \tag{18}$$

Define $h_n := \sqrt{n}(\bar{\beta}_n - \beta)$ and let $(n_m)_{m\geq 1}$ be an arbitrary subsequence of $(n)_{n\geq 1}$. It is sufficient for equations (17) and (18) that we can demonstrate that there is a further subsequence $(n_{m(k)})_{k\geq 1}$ along which the claimed convergence holds. There exists a sub-

subsequence such that $h_{n_{m(k)}} \to h$ for some $h \in \mathbb{R}^{K_\beta}$.[11] Taking such a subsequence will suffice since as we will now demonstrate, the claimed convergence holds for an arbitrary convergent sequence $h_n \to h$.

Let Q_n^n denote the law of Y^n corresponding to θ_n and P_0^n that corresponding to θ_0. Let $\Lambda_{n,\theta_n/\theta_0}(Y^n)$ be the corresponding log-likelihood ratio. Given Assumption 3.3.1 we have that Theorem A.1 implies that

$$\Lambda_{n,\theta_n/\theta_0}(Y^n) = \sqrt{n}\mathbb{P}_n h_n' \dot{\ell}_{\theta_n,\beta} - \frac{1}{2} h_n' \dot{I}_{\theta_0,\beta\beta} h_n + R_n,$$

where $R_n \to 0$ in probability under P_0^n, $\dot{I}_{\theta_0,\beta\beta} = \mathbb{V}\mathrm{ar}(\dot{\ell}_{\theta_0})$, and

$$\Lambda_{n,\theta_n/\theta_0}(Y^n) \rightsquigarrow \mathcal{N}\left(-\frac{1}{2} h' \dot{I}_{\theta_0,\beta\beta} h, h' \dot{I}_{\theta_0,\beta\beta} h\right),$$

under P_0^n, from which we can conclude that $P_0^n \triangleleft\triangleright Q_n^n$ (e.g. van der Vaart and Wellner, 1996, example 3.10.6). This mutual contiguity and Le Cam's first lemma (e.g. van der Vaart, 1998, Lemma 6.4) ensure that (18) holds. Next, by Lemma 7.3 in van der Vaart (2002) we have that

$$\sqrt{n}\mathbb{P}_n\left[\tilde{\ell}_{\theta_n} - \tilde{\ell}_{\theta_0}\right] + \tilde{I}_{\theta_0}(0, b')' \xrightarrow{P_0} 0, \tag{19}$$

It follows that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{\kappa}_{\theta_n}(Y_t, X_t) - \tilde{\kappa}_{\theta_0}(Y_t, X_t) &= [I \;\; -\tilde{I}_{\theta_0,\alpha\beta}\tilde{I}_{\theta_0,\beta\beta}^{-1}] \frac{1}{\sqrt{n}}\sum_{t=1}^n (\tilde{\ell}_{\theta_n}(Y_t, X_t) - \tilde{\ell}_{\theta_0}(Y_t, X_t)) \\
&= -[I \;\; -\tilde{I}_{\theta_0,\alpha\beta}\tilde{I}_{\theta_0,\beta\beta}^{-1}] \begin{bmatrix} \tilde{I}_{\theta_0,\alpha\alpha} & \tilde{I}_{\theta_0,\alpha\beta} \\ \tilde{I}_{\theta_0,\beta\alpha} & \tilde{I}_{\theta_0,\beta\beta} \end{bmatrix} \begin{bmatrix} 0 \\ h \end{bmatrix} \\
&\quad + o_{P_0}(1)
\end{aligned}$$

as well as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{\kappa}_{\theta_n}(Y_t, X_t) &= \frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{\kappa}_{\theta_0}(Y_t, X_t) + \frac{1}{\sqrt{n}}\sum_{t=1}^n [\tilde{\kappa}_{\theta_n}(Y_t, X_t) - \tilde{\kappa}_{\theta_0}(Y_t, X_t)] \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{\kappa}_{\theta_0}(Y_t, X_t) + o_{P_0}(1) \rightsquigarrow \mathcal{N}(0, \mathcal{I}_{\theta_0}),
\end{aligned}$$

under P_0 by Lemma A.1-part 1, establishing (17).

Next we show that (17) and (18) continue to hold if θ_n is replaced by $\bar{\theta}_n$ as defined in the theorem. We will start with the first expression in (18).[12] Since $\bar{\beta}_n$ remains $\sqrt{n}$-consistent there is an $M > 0$ such that $P_0\left(\sqrt{n}\|\bar{\beta}_n - \beta_0\| > M\right) < \varepsilon$. If $\sqrt{n}\|\bar{\beta}_n - \beta_0\| \leq M$ then β_n is equal to one of the values in the finite set $S_n^c = \{\beta \in n^{-1/2}C\,\mathcal{Z}^{L_2} : \|\beta - \beta_0\| \leq n^{-1/2}M\}$.

[11] Such a subsequence and h exist by the Bolzano-Weierstrass theorem.

[12] The proof that (18) continues to hold is adapted from the proof of Theorem 5.48 in van der Vaart (1998).

For each M this set has finite number of elements bounded independently of n, call this upper bound $\overline{B}$. Let

$$R_n(\beta) \coloneqq \sqrt{n}\mathbb{P}_n\left[\hat{\ell}_\theta - \tilde{\ell}_\theta\right]$$

where $\theta = (\alpha_0, \beta, \eta)$. We have that for any $\upsilon > 0$

$$\begin{aligned} P_0\left(\|R_n(\bar{\beta}_n)\| > \upsilon\right) &\leq \varepsilon + \sum_{\beta_n \in S_n^c} P_0\left(\|R_n(\beta_n)\| > \upsilon \wedge \bar{\beta}_n = \beta_n\right) \\ &\leq \varepsilon + \sum_{\beta_n \in S_n^c} P_0\left(\|R_n(\beta_n)\| > \upsilon\right) \\ &\leq \varepsilon + \overline{B} P_0(\|R_n(\beta_n^*)\| > \upsilon), \end{aligned}$$

where β_n^* is chosen from S_n to maximize $P_0\left(\|R_n(\beta_n)\| > \upsilon\right)$. Since the sequence $(\beta_n^*)_{n\geq 1}$ is a deterministic $\sqrt{n}$-consistent sequence for β_n we have that $P_0(\|R_n(\beta_n^*)\| > \upsilon) \to 0$ by equation (18).

The second term in (18) follows from the same set-up. But we now let

$$R_n(\beta) \coloneqq \left\{\left\|\hat{I}_\theta - \tilde{I}_{\theta_0}\right\| < \nu_n\right\}, \quad \text{with } \theta = (\alpha_0, \beta, \eta_0).$$

We have

$$\begin{aligned} P_0\left(R_n(\bar{\beta}_n)\right) &\leq \varepsilon + \sum_{\beta_n \in S_n^c} P_0\left(R_n(\beta_n) \wedge \bar{\beta}_n = \beta_n\right) \\ &\leq \varepsilon + \sum_{\beta_n \in S_n^c} P_0\left(R_n(\beta_n)\right) \\ &\leq \varepsilon + \overline{B} P_0(R_n(\beta_n^*)), \end{aligned}$$

where β_n^* is chosen from S_n^c to maximize $P_0(R_n(\beta_n))$. Since the sequence $(\beta_n^*)_{n\geq 1}$ is a deterministic $\sqrt{n}$-consistent sequence for β_n we have that $P_0(R_n(\beta_n^*)) \to 0$ by equation (18).

Also, for (17) we can argue somewhat similarly. For any β let $Z_n(\beta) \coloneqq \sqrt{n}\mathbb{P}_n\kappa_{\theta_n}$ where $\theta_n \coloneqq (\alpha_0, \beta, \eta)$ and let $Z \sim \mathcal{N}(0, \mathcal{I}_{\theta_0})$. Let f be an arbitrary bounded continuous function: $|f| \leq F$, say, and fix $\varepsilon > 0$. Since $\sqrt{n}\|\bar{\beta}_n - \beta\| = O_{P_0}(1)$ there is a $M > 0$ such that $\sqrt{n}\|\bar{\beta}_n - \beta\| \leq M$ with probability at least $1 - \varepsilon/(4F)$ for all sufficiently large $n \in \mathbb{N}$. Note that by the construction of S_n^c (a) $P_0(\bar{\beta}_n \notin S_n^c) \leq \varepsilon/(4F)$ for all sufficiently large $n \in \mathbb{N}$ and (b) as previously mentioned S_n^c is finite and the number of elements it contains, $\overline{B}$, is bounded independently of n.

For each n, we can partition the sample space $\mathcal{W}^n$ as follows:

$$\mathcal{W}^n = \left\{\bar{\beta}_n \notin S_n^c\right\} \cup \bigcup_{\beta_n \in S_n^c} \left\{\bar{\beta}_n = \beta_n\right\}.$$

Therefore, we can write

$$\begin{aligned}
|P_0 f(Z_n(\bar{\beta}_n)) - P_0 f(Z)| &\leq P_0 \left| f\left(Z_n(\bar{\beta}_n)\right) - f(Z) \right| \mathbf{1}\left\{\bar{\beta}_n \notin S_n^c\right\} \\
&\quad + \sum_{\beta_n \in S_n} P_0 \left| f\left(Z_n(\bar{\beta}_n)\right) - f(Z) \right| \mathbf{1}\left\{\bar{\beta}_n = \beta_n\right\} \\
&\leq \varepsilon/2 + \sum_{\beta_n \in S_n^c} P_0 \left| f\left(Z_n(\beta_n)\right) - f(Z) \right| \mathbf{1}\left\{\bar{\beta}_n = \beta_n\right\} \\
&\leq \varepsilon/2 + \sum_{\beta_n \in S_n^c} P_0 \left| f\left(Z_n(\beta_n)\right) - f(Z) \right| \\
&\leq \varepsilon/2 + \overline{B} P_0 \left| f\left(Z_n(\beta_n^*)\right) - f(Z) \right|,
\end{aligned}$$

where each $\beta_n^* \in S_n^c$ is chosen to maximize $P_0 \left| f\left(Z_n(\beta)\right) - f(Z) \right|$. The corresponding sequence $(\beta_n^*)_{n\in\mathbb{N}}$ is $\sqrt{n}$-consistent by construction and deterministic. Therefore, for the sequence $(\theta_n^*)_{n\in\mathbb{N}}$ with $\theta_n^* = (\alpha_0, \beta_n^*, \eta)$, by (17) we have $Z_n^* \coloneqq Z_n(\beta_n^*) \rightsquigarrow Z$ under P_0. By Skorohod's representation theorem there is a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$ on which we can define random vectors $(X_n^*)_{n\geq 1}$ and X such that the law of each X_n^* is the law of Z_n^* and the law of X is the law of Z and $X_n^*(\omega) \to X(\omega)$ for each $\omega \in \tilde{\Omega}$.[13] Then, since f is bounded, the sequence $(f(X_n^*))_{n\geq 1}$ is uniformly integrable and so by e.g. the Corollary to Theorem 16.14 of Billingsley (1995, pp. 217 – 218) we have that

$$P_0 \left| f(Z_n^*) - f(Z) \right| = \tilde{\mathrm{P}} \left| f(X_n^*) - f(X) \right| \to 0 \text{ as } n \to \infty.$$

By combining the last two displays we have that for sufficiently large $n \in \mathbb{N}$,

$$|P_0 f(Z_n(\bar{\beta}_n)) - P_0 f(Z)| \leq \varepsilon.$$

We conclude that $P_0 f(Z_n(\bar{\beta}_n)) \to P_0 f(Z)$. Since f was an arbitrary bounded continuous function, we have that $Z_n(\bar{\beta}_n) \rightsquigarrow Z$ under P_0 as required.

Next, let $Z_n \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\kappa}_{\theta_n}(Y_i)$ and re-write it as

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\kappa}_{\bar{\theta}_n}(Y_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\kappa}_{\theta_n}(Y_i) - \tilde{\kappa}_{\bar{\theta}_n}(Y_i)),$$

to conclude that $Z_n \rightsquigarrow Z \sim N(0, \tilde{\mathcal{I}}_{\theta_0})$. We have

$$\left\| \hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0} \right\|_2 \leq \left\| \hat{I}_{\bar{\theta}_n,\alpha\alpha} - \tilde{I}_{\theta_0,\alpha\alpha} \right\|_2 + \left\| \hat{I}_{\bar{\theta}_n,\alpha\beta} \hat{I}_{\bar{\theta}_n,\beta\beta}^{-1} \hat{I}_{\bar{\theta}_n,\beta\alpha} - \tilde{I}_{\theta_0,\alpha\beta} \tilde{I}_{\theta_0,\beta\beta}^{-1} \tilde{I}_{\theta_0,\beta\alpha} \right\|_2.$$

By repeated addition and subtraction along with the observations that any submatrix has a smaller operator norm than the original matrix we obtain and the matrix inverse is Lipschitz

[13] See e.g. Billingsley, 1999, Theorem 6.7.

continuous at a non-singular matrix we obtain

$$\left\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\right\|_2 \lesssim \left\|\hat{I}_{\bar{\theta}_n} - \tilde{I}_{\theta_0}\right\|_2.$$

Hence by equation (18) with $\bar{\theta}_n$ replacing θ_n we have $P_0\left(\left\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\right\|_2 < \check{\nu}_n\right) \to 1$ where $\check{\nu}_n = C\nu_n$ for some positive constant $C \geq 1$.

Now to complete the proof we first consider the case where $\operatorname{rank}(\tilde{\mathcal{I}}_{\theta_0}) = r > 0$. We first show that $\hat{\mathcal{I}}^t_{\theta_0} \xrightarrow{P_0} \tilde{\mathcal{I}}_{\theta_0}$ and the rank estimate $r_n = \operatorname{rank}(\hat{\mathcal{I}}^t_{\theta_0})$ satisfies $P_0(\{r_n = r\}) \to 1$.

Let λ_l denote the lth largest eigenvalue of $\tilde{\mathcal{I}}_{\theta_0}$, similarly define $\hat{\lambda}_{l,n}$ for $\hat{\mathcal{I}}_{\theta_0}$ and $\hat{\lambda}^t_{l,n}$ for $\hat{\mathcal{I}}^t_{\theta_0}$. Define the set $R_n \coloneqq \{r_n = r\}$, let $\underline{\nu} \coloneqq \lambda_r/2 > 0$ and note that — $P(\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 < \check{\nu}_n) \to 1$ — implies that $\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 = o_{P_0}(1)$.

By Weyl's perturbation theorem[14] we have $\max_{l=1,\ldots,L} |\hat{\lambda}_{l,n} - \lambda_l| \leq \|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 = o_{P_0}(1)$. Hence, if we define $E_n \coloneqq \{\hat{\lambda}_{r,n} \geq \check{\nu}_n\}$, for n large enough such that $\check{\nu}_n < \underline{\nu}$, we have

$$P(E_n) = P\left(\hat{\lambda}_{r,n} \geq \check{\nu}_n\right) \geq P\left(\hat{\lambda}_{r,n} \geq \underline{\nu}\right) \geq P\left(|\hat{\lambda}_{r,n} - \lambda_r| < \underline{\nu}\right) \to 1.$$

If $r = L$ we have that $R_n \supset E_n$ and therefore $P(R_n) \to 1$. Additionally, if $\hat{\lambda}_{L,n} \geq \nu_n$ then $\hat{\lambda}^t_{l,n} = \hat{\lambda}_{l,n}$ for each $l \in [L]$ and hence $\hat{\mathcal{I}}^t_{\theta_0} = \hat{\mathcal{I}}_{\theta_0}$. Thus, $E_n \cap \{\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\| \leq \upsilon\} \subset \{\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\| \leq \upsilon\}$, from which it follows that $\hat{\mathcal{I}}^t_{\theta_0} \xrightarrow{P_0} \tilde{\mathcal{I}}_{\theta_0}$.

Now suppose instead that $r < L$ and define $F_n \coloneqq \{\hat{\lambda}_{r+1,n} < \nu_n\}$. It follows by Weyl's perturbation theorem and the fact that $\lambda_l = 0$ for $l > r$ that as $n \to \infty$

$$P(F_n) = P(\hat{\lambda}_{r+1,n} < \check{\nu}_n) \geq P(\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 < \check{\nu}_n) \to 1.$$

Since $R_n \supset E_n \cap F_n$, this implies that $P(R_n) \to 1$ as $n \to \infty$. Additionally, if $\hat{\lambda}_{r,n} \geq \nu_n$, $\hat{\lambda}_{r+1,n} < \nu_n$ and $\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 \leq \upsilon$, we have that $\hat{\lambda}^t_{k,n} = \hat{\lambda}_{k,n}$ for $k \leq r$ and $\hat{\lambda}^t_{l,n} = 0 = \lambda_l$ for $l > r$ and so

$$\|\hat{\Lambda}_n(\nu_n) - \Lambda\|_2 = \max_{l=1,\ldots,r} |\hat{\lambda}^t_{l,n} - \lambda_l| = \max_{l=1,\ldots,r} |\hat{\lambda}_{l,n} - \lambda_l| \leq \|\hat{\Lambda}_n - \Lambda\|_2 \leq \|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 \leq \upsilon,$$

and hence $\{\|\hat{\Lambda}_n(\nu_n) - \Lambda\|_2 \leq \upsilon\} \cap E_n \cap F_n \subset \{\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 \leq \upsilon\}$, from which it follows that $\hat{\Lambda}_n(\nu_n) \xrightarrow{P_0} \Lambda$.

To complete this part of the proof, suppose that $(\lambda_1, \ldots, \lambda_r)$ consists of s distinct eigenvalues with values $\lambda^1 > \lambda^2 > \cdots > \lambda^s$ and multiplicities $\mathfrak{m}_1, \ldots, \mathfrak{m}_s$ (each at least one), where the superscripts on the λs are indices, not exponents. $\lambda^{s+1} = 0$ is an eigenvalue with multiplicity $\mathfrak{m}_{s+1} = L - r$. Let l^k_i for $k = 1, \ldots, s+1$ and $i = 1, \ldots, \mathfrak{m}_k$

[14] E.g. Corollary III.2.6 in Bhatia (1997).

denote the column indices of the eigenvectors in U corresponding to each λ^k. For each λ^k, the total eigenprojection is $\Pi_k \coloneqq \sum_{i=1}^{\mathfrak{m}_k} u_{l_i^k} u'_{l_i^k}$.[15] Total eigenprojections are continuous.[16] Therefore, if we construct $\hat{\Pi}_{k,n}$ in in an analogous fashion to Π_k but replace columns of U with columns of $\hat{U}_n$, we have $\hat{\Pi}_{k,n} \xrightarrow{P} \Pi_k$ for each $k \in [s+1]$ since $\hat{I}_{\theta_0,n} \xrightarrow{P} \tilde{I}_{\theta_0}$. Spectrally decompose $\tilde{I}_{\theta_0}$ as $\tilde{I}_{\theta_0} = \sum_{k=1}^{s} \lambda^k \Pi_k$, where the sum runs to s rather than $s+1$ since $\lambda^{s+1} = 0$. Then,

$$\hat{\mathcal{I}}^t_{\theta_0} = \sum_{k=1}^{s+1} \sum_{i=1}^{\mathfrak{m}_k} \hat{\lambda}^t_{l_i^k,n} \hat{u}_{l_i^k,n} \hat{u}'_{l_i^k,n} = \sum_{k=1}^{s+1} \sum_{i=1}^{\mathfrak{m}_k} (\hat{\lambda}^t_{l_i^k,n} - \lambda^k) \hat{u}_{l_i^k,n} \hat{u}'_{l_i^k,n} + \sum_{k=1}^{s} \lambda^k \hat{\Pi}_{k,n},$$

and so

$$\|\hat{\mathcal{I}}^t_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 \leq \sum_{k=1}^{s+1} \sum_{i=1}^{\mathfrak{m}_k} |\hat{\lambda}^t_{l_i^k,n} - \lambda^k| \|\hat{u}_{l_i^k,n} \hat{u}'_{l_i^k,n}\|_2 + \sum_{k=1}^{s} |\lambda^k| \|\hat{\Pi}_{k,n} - \Pi_k\|_2 \xrightarrow{P_0} 0,$$

by $\hat{\Pi}_{k,n} \xrightarrow{P} \Pi_k$, $\hat{\Lambda}_n(\nu_n) \xrightarrow{P_0} \Lambda$ and since we have $\|u_{l_i^k,n} u'_{l_i^k,n}\|_2 = 1$ for any i, k, n.

Hence, we have that $\hat{\mathcal{I}}^t_{\theta_0} \xrightarrow{P_0} \tilde{\mathcal{I}}_{\theta_0}$ and $P_0(\{r_n = r\}) \to 1$. This implies that $\hat{\mathcal{I}}^{t,\dagger}_{\theta_0} \xrightarrow{P_0} \tilde{\mathcal{I}}^{\dagger}_{\theta_0}$ where $\tilde{\mathcal{I}}^{\dagger}_{\theta_0}$ is the Moore-Penrose inverse of $\tilde{\mathcal{I}}_{\theta_0}$.[17]

Now consider the score statistic $\hat{S}^{SR}_{\bar{\theta}_n}$, by Slutsky's lemma and the continuous mapping theorem we have that

$$\hat{S}^{SR}_{\bar{\theta}_n} = Z'_n \hat{\mathcal{I}}^{t,\dagger}_{\theta_0} Z_n \rightsquigarrow Z' \tilde{\mathcal{I}}^{\dagger}_{\theta_0} Z \sim \chi^2_r$$

where the distributional result $X \coloneqq Z' \tilde{\mathcal{I}}^{\dagger}_{\theta_0} Z \sim \chi^2_r$, follows from e.g. Theorem 9.2.2 in Rao and Mitra (1971).

Finally, recall that $R_n = \{r_n = r\}$. On these sets c_n is the $1-a$ quantile of the χ^2_r distribution, which we will call c. Hence, we have $c_n \xrightarrow{P_0} c$ as $P_0(R_n) \to 1$. As a result, we obtain $\hat{S}^{SR}_{\bar{\theta}_n} - c_n \rightsquigarrow X - c$ where $X \sim \chi^2_r$. Since the χ^2_r distribution is continuous, we have by the Portmanteau theorem

$$\begin{aligned} P_0\left(\hat{S}^{SR}_{\bar{\theta}_n} > c_n\right) = 1 - P_0\left(\hat{S}^{SR}_{\bar{\theta}_n} - c_n \leq 0\right) \to 1 - P_0\left(X - c \leq 0\right) = 1 - P_0\left(X \leq c\right) \\ = 1 - (1-a) = a\,, \end{aligned}$$

which completes the proof in the case that $r > 0$.

It remains to handle the case with $r = 0$. We first note that $Z_n \rightsquigarrow Z \sim \mathcal{N}(0, \tilde{\mathcal{I}}_{\theta_0})$ continues to hold by our assumptions, though in this case $\tilde{\mathcal{I}}_{\theta_0}$ is the zero matrix and hence the limiting distribution is degenerate: $Z = 0$. Let $E_n = \{r_n = 0\}$. Part 3 of Lemma A.1 and Weyl's

[15] See e.g Chapter 8.8 of Magnus and Neudecker (2019).

[16] E.g. Theorem 8.7 of Magnus and Neudecker (2019).

[17] A necessary and sufficient condition for $(M + E_n)^\dagger \to M^\dagger$ as $E_n \to 0$ is that for all sufficiently large n, $\operatorname{rank}(M + E_n) = \operatorname{rank}(M)$; see, for example, chapter 6.6 of Ben-Israel and Greville (2003).

perturbation theorem imply that

$$P_0(E_n) = P_0\left(r_n = 0\right) = P_0\left(\max_{l=1,\ldots,L} |\hat{\lambda}_{n,l}| < \check{\nu}_n\right) \geq P_0\left(\|\hat{\mathcal{I}}_{\theta_0} - \tilde{\mathcal{I}}_{\theta_0}\|_2 < \check{\nu}_n\right) \to 1.$$

On the sets E_n we have that $\hat{\mathcal{I}}^t_{\theta_0}$ is the zero matrix, whose Moore-Penrose inverse is also the zero matrix. Hence on the sets E_n we have $\hat{S}^{SR}_{\tilde{\theta}_n} = 0$ and $c_n = 0$ and therefore do not reject, implying

$$P(\hat{S}^{SR}_{\tilde{\theta}_n} > c_n) \leq 1 - P_0(E_n) \to 0.$$

It follows that $P(\hat{S}^{SR}_{\tilde{\theta}_n} > c_n) \to 0$.

□

Auxiliary results

THEOREM A.1: *Suppose that Assumption 3.3.1 is satisfied. Let $\beta \in \mathcal{B}_0$ and $\beta_n = \beta + n^{-1/2}h_n$, $\tilde{\beta}_n = \beta + n^{-1/2}\tilde{h}_n$ such that $\|h_n\|$ and $\|\tilde{h}_n\|$ remain bounded as $n \to \infty$. Define $\theta_n = (\alpha_0, \beta_n, \eta)$, $\tilde{\theta}_n = (\alpha_0, \tilde{\beta}_n, \eta)$, $\theta_0 = (\alpha_0, \beta, \eta)$ for any $\alpha_0 \in \mathcal{A}_0$ and $\eta \in \mathcal{H}_0$ and*

$$\Lambda_{n,\tilde{\theta}_n/\theta_n}(Y^n) \coloneqq \log\left(\frac{\mathrm{d}P^n_{\tilde{\theta}_n}}{\mathrm{d}P^n_{\theta_n}}\right) = \sum_{t=1}^{n} \log\left(|\tilde{A}_n| \prod_{k=1}^{K} \eta_k(\tilde{\epsilon}_{k,t,n})/|A_n| \prod_{k=1}^{K} \eta_k(\epsilon_{k,t,n})\right) \quad (20)$$

where $\tilde{A}_n = A(\alpha_0, \tilde{\sigma}_n)$, $A_n = A(\alpha_0, \sigma_n)$, $\tilde{\epsilon}_{k,t,n} = \tilde{A}_{n,k\bullet}(Y_t - \tilde{c}_n - \tilde{B}_n X_t)$ and $\epsilon_{k,t,n} = A_{n,k\bullet}(Y_t - c_n - B_n X_t)$. Then, we have that

$$\Lambda_{n,\tilde{\theta}_n/\theta_n}(Y^n) = \tilde{h}_n' \sqrt{n}\bar{\ell}_{n,\theta_n,\beta} - \frac{1}{2}\tilde{h}_n' \dot{I}_{\theta_0,\beta\beta}\tilde{h}_n + o_P(1) \quad (21)$$

under $P^n_{\theta_n}$ as $n \to \infty$, where

$$\bar{\ell}_{n,\theta_n,\beta} \coloneqq \frac{1}{n}\sum_{t=1}^{n} \dot{\ell}_{\theta_n,\beta}(Y_t, X_t) \qquad \textit{with} \qquad \dot{\ell}_{\theta_n,\beta}(Y_t, X_t) = \frac{\mathrm{d}\log p_{\theta_n}}{\mathrm{d}\beta}(Y_t, X_t) \quad (22)$$

and

$$\dot{I}_{\theta_0,\beta\beta} \coloneqq \mathbb{E}\dot{\ell}_{\theta_n,\beta}\dot{\ell}'_{\theta_n,\beta}\,. \quad (23)$$

Moreover, we have that

$$\sqrt{n}\bar{\ell}_{n,\theta_n,\beta} \rightsquigarrow N(0, \dot{I}_{\theta_0,\beta\beta})\,, \quad (24)$$

under $P^n_{\theta_n}$ as $n \to \infty$.

Proof. See Hallin & Saidi Theorem 2.1. □

Proof of Lemma A.1. For part 1 we note that Lemma 3.1 implies that under P_{θ_0} we have that $\tilde{\ell}_{n,\theta_0}(Y^n) = \sum_{t=1}^{n} \tilde{\ell}_{n,\theta_0}(Y_t, X_t)$ is the sum of n independent components as $A_{k\bullet}(Y_t -$

$c - BX_t) = \epsilon_{k,t}$ and $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{K,t})'$ is independent across t by Assumption 3.3.1 part 2. Further, the efficient score lies in $L_2(P_0)$ by construction, hence the convergence follows from the CLT.

Next, let $\theta_n = (\alpha_0, \beta_n, \eta)$ and note that under P_{θ_n}, each $A_{n,k}(Y_t - c_n - B_n X_t) \approx \epsilon_{k,t} \sim \eta_k$. Hence we can compute certain properties of the efficient score using the equality in distribution:

$$\tilde{\ell}_{\theta,\alpha_l}(Y_t, X_t) \approx \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j} \phi_k(\epsilon_{k,t}) \epsilon_{j,t} + \sum_{k=1}^{K} \zeta^{\alpha}_{n,l,k,k} \left[\tau_{k,1} \epsilon_{k,t} + \tau_{k,2} \kappa(\epsilon_{k,t})\right] \tag{25}$$

$$\tilde{\ell}_{\theta,\sigma_l}(Y_t, X_t) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{n,l,k,j} \phi_k(\epsilon_{k,t}) \epsilon_{j,t} + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k} \left[\tau_{k,1} \epsilon_{k,t} + \tau_{k,2} \kappa(\epsilon_{k,t})\right] \tag{26}$$

$$\tilde{\ell}_{\theta,c_l}(Y_t, X_t) = \sum_{k=1}^{K} \phi_k(\epsilon_{k,t}) \times [-A_{k\bullet} D_{c,l}] \tag{27}$$

$$\tilde{\ell}_{\theta,b_l}(Y_t, X_t) = \sum_{k=1}^{K} \phi_k(\epsilon_{k,t}) \times [-A_{k\bullet} D_{b,l} (X_t - \mathbb{E} X_t)] \tag{28}$$

where we note that the same observation implies that $\tau_{k,n} = \tau_k$ for each n.[18] By our assumptions on the map $(\alpha, \sigma) \mapsto A(\alpha, \sigma)$, we have $\zeta^{\alpha}_{n,l,k,j} \to \zeta^{\alpha}_{\infty,l,k,j} := [D_{\alpha_l}(\alpha_0, \sigma)]_{k\bullet} A(\alpha_0, \sigma)^{-1}_{\bullet j}$ and $\zeta^{\sigma}_{n,l,k,j} \to \zeta^{\sigma}_{\infty,l,k,j} := [D_{\sigma_l}(\alpha_0, \sigma)]_{k\bullet} A(\alpha_0, \sigma)^{-1}_{\bullet j}$. Note that the entries of $D_{c,l}$ and $D_{b,l}$ are all zero except for entry l (corresponding to c_l or b_l) which is equal to one.

We verify part 2 for each component of the efficient score (25)-(28). Components (25) and (26) follows similarly and we focus on (25). We define

$$\varphi_{1,n}(y_t, x_t) := \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j,n} \phi_k(A_{n,k\bullet} v_{t,n}) A_{n,j\bullet} v_{t,n} \ ,$$

and

$$\hat{\varphi}_{1,n}(y_t, x_t) := \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j,n} \hat{\phi}_{k,n}(A_{n,k\bullet} v_{t,n}) A_{n,j\bullet} v_{t,n} \ ,$$

with $v_{t,n} = y_t - c_n - B_n x_t$, and let $\overline{\zeta}_n := \max_{l\in[L], j\in[K], k\in[K]} |\zeta^{\alpha}_{l,j,k,n}|$ which converges to $\overline{\zeta} := \max_{l\in[L], j\in[K], k\in[K]} |\zeta^{\alpha}_{l,j,k,\infty}| < \infty$. We have that

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n(\hat{\varphi}_{1,n} - \varphi_{1,n}) \leq \\
&\sqrt{n} \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \overline{\zeta}_n \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\phi}_{k,n}(A_{n,k\bullet} V_{t,n}) A_{n,j\bullet} V_{t,n} - \phi_k(A_{n,k\bullet} V_{t,n}) A_{n,j\bullet} V_{t,n} \right| \ ,
\end{aligned}$$

[18] In the preceding display we have written $\zeta^{\alpha}_{n,l,k,k}$ and $\zeta^{\sigma}_{n,l,k,k}$ rather than $\zeta^{\alpha}_{l,k,k}$ and $\zeta^{\sigma}_{l,k,k}$ to indicate their dependence on β_n. E.g. $\zeta^{\alpha}_{\infty,l,k,j}$ corresponds to evaluation at the point (α_0, β).

with $V_{t,n} = Y_t - c_n - B_n X_t$. Since each

$$\left| \frac{1}{n} \sum_{t=1}^{n} \hat{\phi}_{k,n}(A_{n,k\bullet} V_{t,n}) A_{n,j\bullet} V_{t,n} - \phi_k(A_{n,k\bullet} V_{t,n}) A_{n,j\bullet} V_{t,n} \right| = o_{P_{\theta_n}}(n^{-1/2})$$

by applying Assumption 3.3.2 with $W_{t,n} = A_{n,j\bullet} V_{t,n}$ (noting that under P_{θ_n}, $A_{n,k\bullet} V_{t,n} \simeq \epsilon_{k,t}$ and $A_{n,j\bullet} V_{t,n} \simeq \epsilon_{j,t}$ are independent with $\mathbb{E}_{\theta_n}(A_{n,j\bullet} V_{t,n})^2 = 1$ by Assumption 3.3.1-part 2, hence the LLN implies the required convergence under P_{θ_n}) and the outside summations are finite, it follows that

$$\sqrt{n}\mathbb{P}_n(\hat{\varphi}_{1,n} - \varphi_{1,n}) = o_{P_{\theta_n}}(1) \,. \tag{29}$$

Next, we show that $\hat{\tau}_{k,n} - \tau_k \to 0$ in P_{θ_n}-probability where $\hat{\tau}_{k,n}$ is defined in (3.11). First, note $A_{n,k\bullet} V_{t,n} \eqsim \epsilon_{k,t} \sim \eta_k$ under P_{θ_n} and the sequences $((\epsilon_{k,t})^3)_{t\geq 1}$ and $((\epsilon_{k,t})^4)_{t\geq 1}$ are i.i.d. and have finite mean by Assumption 3.3.1. Hence by the law of large numbers we have that $\hat{M}_k - M_k \xrightarrow{P_{\theta_n}} 0$. Since M_k is nonsingular by Assumption 3.3.1, the continuous mapping theorem then yields $\hat{\tau}_{k,n} \xrightarrow{P_{\theta_n}} \tau_k$.

Now, consider $\varphi_{2,\tau,n}(y_t, x_t)$ defined by

$$\varphi_{2,\tau,n}(y_t, x_t) \coloneqq \sum_{k=1}^{K} \zeta^{\alpha}_{n,l,k,k} \left[\tau_{k,1} A_{n,k\bullet} v_{t,n} + \tau_{k,2} \kappa(A_{n,k\bullet} v_{t,n}) \right].$$

Since sum is finite and each $|\zeta^{\alpha}_{n,l,k,k}| \to |\zeta^{\alpha}_{\infty,l,k,k}| < \infty$ it is sufficient to consider the convergence of the summands. In particular we have that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\hat{\tau}_{k,n,1} - \tau_{k,1} \right] A_{n,k\bullet} V_{t,n} &= \left[\hat{\tau}_{k,n,1} - \tau_{k,1} \right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{n,k\bullet} V_{t,n} \\ &= o_{P_{\theta_n}}(1) \times O_{P_{\theta_n}}(1) \\ &= o_{P_{\theta_n}}(1), \end{aligned}$$

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\hat{\tau}_{k,n,2} - \tau_{k,2} \right] \kappa(A_{n,k\bullet} V_{t,n}) &= \left[\hat{\tau}_{k,n,2} - \tau_{k,2} \right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \kappa(A_{n,k\bullet} V_{t,n}) \\ &= o_{P_{\theta_n}}(1) \times O_{P_{\theta_n}}(1) \\ &= o_{P_{\theta_n}}(1). \end{aligned}$$

since $A_{n,k\bullet} V_{t,n} \eqsim \epsilon_{k,t} \sim \eta_k$ under P_{θ_n} and $(\epsilon_{k,t})_{t\geq 1}$ and $(\kappa(\epsilon_{k,t}))_{t\geq 1}$ are i.i.d. mean-zero sequences with finite second moments such that the CLT holds.

Together these yield that

$$\sqrt{n}\mathbb{P}_n(\varphi_{2,\hat{\tau}_n,n} - \varphi_{2,\tau,n}) = o_{P_{\theta_n}}(1). \tag{30}$$

Putting (29) and (30) together yields the required convergence for components of the type (25) which follows identically), since $\tilde{\ell}_{\theta_n,\alpha_l} = \varphi_{1,n} + \varphi_{2,\tau,n}$ and $\hat{\ell}_{\theta_n,\alpha_l} = \hat{\varphi}_{1,n} + \varphi_{2,\hat{\tau}_n,n}$. We note that convergence for components of type (26) follows using identical steps.

Next, we consider components (27).

$$\sqrt{n}\mathbb{P}_n(\hat{\ell}_{\theta_n,c_l} - \tilde{\ell}_{\theta_n,c_l}) \leq$$
$$\sqrt{n}\sum_{k=1}^{K}\left|\frac{1}{n}\sum_{t=1}^{n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{c,l} - \phi_k(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{c,l}\right| = o_{P_{\theta_n}}(1)\,.$$

Since each $\left|\frac{1}{n}\sum_{t=1}^{n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{c,l} - \phi_k(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{c,l}\right| = o_{P_{\theta_n}}(n^{-1/2})$ by applying Assumption 3.3.2 with the constant $A_{n,k\bullet}D_{c,l}$ and we note that the outside sum is finite.

For components (28) let $\mu_n = (I - B_{n,1} - \ldots - B_{n,p})^{-1}c_n$ and note that $\mathbb{E}_{\theta_n}X_t = \iota_p \otimes \mu_n$. We have that

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\hat{\ell}_{\theta_n,b_l} - \tilde{\ell}_{\theta_n,b_l}) \leq \sqrt{n}\sum_{k=1}^{K}\Bigg|&\frac{1}{n}\sum_{t=1}^{n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)\\ &- \phi_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)\Bigg|\\ &= o_{P_{\theta_n}}(1)\end{aligned}$$

since each

$$\left|\frac{1}{n}\sum_{t=1}^{n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t) - \phi_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)\right|$$
$$= o_{P_{\theta_n}}(n^{-1/2})$$

by Assumption 3.3.2 applied with $W_{t,n} = A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)$, noting that $X_t - \mathbb{E}_{\theta_n}X_t = ((Y_{t-1} - \mu_n)', \ldots, (Y_{t-p} - \mu_n)')'$ which is independent of $A_{n,k\bullet}V_{t,n} \simeq \epsilon_{k,t}$ under P_{θ_n} by Assumption 3.3.1. Further note that, X_t is stationary-ergodic by Assumption 3.3.1 and $\mathbb{E}_{\theta_n}(X_t - \mathbb{E}_{\theta_n}X_t)(X_t - \mathbb{E}_{\theta_n}X_t)'$ is finite, hence the required convergence of $\frac{1}{n}\sum_{t=1}^{n}W_{t,n}^2 = \frac{1}{n}\sum_{t=1}^{n}A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)(X_t - \mathbb{E}_{\theta_n}X_t)'D_{b,l}'A_{n,k\bullet}'$ follows from the ergodic theorem and the continuous mapping theorem.

To verify part 3 we will show that

$$\left\|\hat{I}_{\theta_n} - \tilde{I}_{\theta_0}\right\|_2 \leq \left\|\hat{I}_{\theta_n} - \tilde{I}_{\theta_n}\right\|_2 + \left\|\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}\right\|_2 = o_{P_{\theta_n}}(\nu_n^{1/2}). \tag{31}$$

where $\tilde{I}_{\theta_n} \coloneqq \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}(Y_t, X_t)\tilde{\ell}_{\theta_n}(Y_t, X_t)'$. To obtain the rates we start with $\|\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_2$, for which we show that each component satisfies the required rate. To set this up, let $Q_{l,m,t,n}^{r,s} = \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)\tilde{\ell}_{\theta_n,s_m}(Y_t, X_t) - \tilde{\ell}_{\theta_0,r_l}(Y_t, X_t)\tilde{\ell}_{\theta_0,s_m}(Y_t, X_t)$, where $r, s \in$

$\{\alpha, \sigma, c, b\}$ and l, m denote the indices of the components of the efficient scores. Fix any r, s and l, m and note that under P_{θ_n} we have $A_{n,k}(Y_t - c_n - B_n X_t) \approx \epsilon_{k,t} \sim \eta_k$ and therefore to show $[\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}]_{l,m} = o_{P_{\theta_n}}(\nu_n^{1/2})$ it suffices to show

$$\frac{1}{n}\sum_{t=1}^{n} Q^{r,s}_{l,m,t,n} - GQ^{r,s}_{l,m,t,n} + \frac{1}{n}\sum_{t=1}^{n} G[Q^{r,s}_{l,m,t,n} - Q^{r,s}_{l,m,t,\infty}] = o_G(\nu_n^{1/2}),$$

where G denotes the law corresponding to η and each $Q^{r,s}_{l,m,t,n}$ is shown to satisfy $\|Q^{r,s}_{l,m,t,n}\|_{G,p} < \infty$ in Lemma A.4 given below. The convergence of the second term follows from the assumed Lipschitz continuity of the map defining the ζ's and the $\sqrt{n}$-consistency of β_n for β, since $n^{-1/2} = o(\nu_n^{1/2})$. For the first term, if $p = 2$ in Lemma A.4 and Theorem 2.5.11 in Durrett (2019), we have that for all $\iota > 0$

$$\frac{1}{n}\sum_{i=1}^{n} Q^{r,s}_{l,m,i,n} - GQ^{r,s}_{l,m,i,n} = o_G\left(n^{-1/2}\log(n)^{1/2+\iota}\right).$$

It follows that

$$\|\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_2 \leq \|\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_F = o_{P_{\theta_n}}\left(n^{-1/2}\log(n)^{1/2+\iota}\right).$$

If, instead, $p = 1 + \nu/4 < 2$ in Lemma A.4, then by the Marcinkiewicz & Zygmund SLLN (e.g. Theorem 2.5.12 in Durrett, 2019)

$$\frac{1}{n}\sum_{i=1}^{n} Q^{r,s}_{l,m,i,n} - GQ^{r,s}_{l,m,i,n} = o_G\left(n^{\frac{1-p}{p}}\right),$$

and similarly

$$\|\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_2 \leq \|\tilde{I}_{\theta_n,n} - \tilde{I}_{\theta_0}\|_F = o_{P_{\theta_n}}\left(n^{\frac{1-p}{p}}\right).$$

That is, for any $p \in (1, 2]$ we have $\|\tilde{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_2 = o_{P_{\theta_n}}(\nu_{n,p}) = o_{P_{\theta_n}}(\nu_n^{1/2})$.

For the other component of the sum, let $r \in \{\alpha, \sigma, c, b\}$ and let l denote an index, we write $\hat{U}_{n,t,r_l} := \hat{\ell}_{\theta_n,r_l}(Y_t, X_t)$, $\tilde{U}_{t,r_l} := \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)$ and $D_{n,t,r_l} := \hat{\ell}_{\theta_n,r_l}(Y_t, X_t) - \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)$.

Since it is the absolute value of the $(r, l) - (s, m)$ component of $\hat{I}_{\theta_n,n} - \tilde{I}_{\theta_0,n}$, it is sufficient to show that $\left|\frac{1}{n}\sum_{t=1}^{n} \hat{U}_{n,t,r_l} D_{n,t,s_m} + \frac{1}{n}\sum_{t=1}^{n} D_{n,t,r_l}\tilde{U}_{t,s_m}\right| = o_{P_{\theta_n}}(\nu_n^{1/2})$ as $n \to \infty$ for any $r, s \in \{\alpha, \sigma, c, b\}$ and l, m. By Cauchy-Schwarz and Lemma A.6

$$\begin{aligned}\left|\frac{1}{n}\sum_{t=1}^{n} D_{n,t,r_l}\tilde{U}_{t,s_m}\right| &\leq \left(\frac{1}{n}\sum_{t=1}^{n}\tilde{U}^2_{t,s_m}\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n} D^2_{n,t,r_l}\right)^{1/2} \\ &= O_{P_{\theta_n}}(1) \times o_{P_{\theta_n}}(\nu_n^{1/2}) = o_{P_{\theta_n}}(\nu_n^{1/2}),\end{aligned}$$

$$\left|\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m}\right| \leq \left(\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}D_{n,t,s_m}^2\right)^{1/2}$$
$$= O_{P_{\theta_n}}(1) \times o_{P_{\theta_n}}(\nu_n^{1/2}) = o_{P_{\theta_n}}(\nu_n^{1/2}),$$

for any $(r,l)-(s,m)$. It follows that

$$\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m} + D_{n,t,r_l}\tilde{U}_{t,s_m}\right]^2 \leq 2\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m}\right]^2$$
$$+ 2\left[\frac{1}{n}\sum_{t=1}^{n}D_{n,t,r_l}\tilde{U}_{t,s_m}\right]^2$$
$$= o_{P_{\theta_n}}(\nu_n)$$

and hence $\|\hat{I}_{\theta_n,n}-\tilde{I}_{\theta_0,n}\|_2 \leq \|\hat{I}_{\theta_n,n}-\tilde{I}_{\theta_0,n}\|_F = o_{P_{\theta_n}}(\nu_n^{1/2})$. We can combine these results to obtain:

$$\|\hat{I}_{\theta_n,n}-\tilde{I}_{\theta_0}\|_2 \leq \|\hat{I}_{\theta_n,n}-\tilde{I}_{\theta_n,n}\|_2+\|\tilde{I}_{\theta_n,n}-\tilde{I}_{\theta_0}\|_2 = o_{P_{\theta_n}}(\nu_n^{1/2})+o_{P_{\theta_n}}(\nu_n^{1/2}) = o_{P_{\theta_n}}(\nu_n^{1/2}).$$

It remains to show that part 4 holds. Recall that the dominating measure here is λ and re-write the integral in question as

$$\int \left\|\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2} - \tilde{\ell}_{\theta_0}p_{\theta_0}^{1/2}\right\|^2 \mathrm{d}\lambda = \sum_{l=1}^{\dim(\tilde{\ell}_{\theta_n})}\int\left[\tilde{\ell}_{\theta_n,l}p_{\theta_n}^{1/2} - \tilde{\ell}_{\theta_0,l}p_{\theta_0}^{1/2}\right]^2 \mathrm{d}\lambda. \tag{32}$$

It is evidently sufficient to show that each of the integrals in the sum on the rhs converges to zero. To this end, let $f_{r,n} \coloneqq \tilde{\ell}_{\theta_n,r_l}p_{\theta_n}^{1/2}$ and $f_r \coloneqq \tilde{\ell}_{\theta_0,r_l}p_{\theta_0}^{1/2}$ for $r \in \{\alpha,\sigma,c,b\}$ corresponding to (25)-(28) for some arbitrary l. By the expressions for $\tilde{\ell}_{\theta_0,r_l}$ given in Lemma 3.3.1 along with the continuity of A, D_{σ_l}, D_{α_l}, D_{c_l}, D_{b_l} and each η_k and ϕ_k (each of which follows from our assumptions), we have that $f_{r,n} \to f_r$ λ-a.e. for all r. Moreover, using the representation in (25) we have

$$\int f_{\alpha,n}^2 \mathrm{d}\lambda = \int\left(\sum_{k=1}^{K}\left[\zeta_{l,k,k,n}^{\alpha}\left[\tau_{k,1}\epsilon_{k,t}+\tau_{k,2}\kappa(\epsilon_{k,t})\right] + \sum_{j=1,j\neq k}^{K}\zeta_{l,k,j,n}\phi_k(\epsilon_{k,t})\epsilon_{j,t}\right]\right)^2 \mathrm{d}G$$
$$= \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\sum_{b=1}^{K}\sum_{m=1,m\neq b}^{K}\zeta_{l,k,j,n}^{\alpha}\zeta_{l,b,m,n}^{\alpha}\int\phi_k(\epsilon_{k,t})\epsilon_{j,t}\phi_b(\epsilon_{b,t})\epsilon_{m,t}\,\mathrm{d}G$$
$$+\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\sum_{b=1}^{K}\zeta_{l,k,j,n}^{\alpha}\zeta_{l,b,b,n}^{\alpha}\int\phi_k(\epsilon_{k,t})\epsilon_{j,i}\left[\tau_{b,1}\epsilon_{b,t}+\tau_{b,2}\kappa(\epsilon_{b,t})\right]\mathrm{d}G$$
$$+\sum_{k=1}^{K}\sum_{b=1}^{K}\zeta_{l,k,k,n}^{\alpha}\zeta_{l,b,b,n}^{\alpha}\int\left[\tau_{b,1}\epsilon_{b,t}+\tau_{b,2}\kappa(\epsilon_{b,t})\right]\left[\tau_{k,1}\epsilon_{k,t}+\tau_{k,2}\kappa(\epsilon_{k,t})\right]\mathrm{d}G$$

where G is the law of ϵ and each of the integrals are finite by Assumption 3.3.1. By the

continuity of A and D_{α_l}, this converges to

$$\begin{aligned}\int f_{(\alpha,\beta_1)}^2 \,\mathrm{d}\lambda &= \int \left(\sum_{k=1}^{K} \left[\zeta_{l,k,k,\infty}^{\alpha} \left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t}) \right] + \sum_{j=1,j\neq k}^{K} \zeta_{l,k,j,\infty}^{\alpha} \phi_k(\epsilon_{k,t})\epsilon_{j,t} \right] \right)^2 \mathrm{d}G \\ &= \sum_{k=1}^{K} \sum_{j=1,j\neq k}^{K} \sum_{b=1}^{K} \sum_{m=1,m\neq b}^{K} \zeta_{l,k,j,\infty}^{\alpha} \zeta_{l,b,m,\infty}^{\alpha} \int \phi_k(\epsilon_{k,t})\epsilon_{j,t}\phi_b(\epsilon_{b,t})\epsilon_{m,t} \,\mathrm{d}G \\ &\quad + \sum_{k=1}^{K} \sum_{j=1,j\neq k}^{K} \sum_{b=1}^{K} \zeta_{l,k,j,\infty}^{\alpha} \zeta_{l,b,b,\infty}^{\alpha} \int \phi_k(\epsilon_{k,t})\epsilon_{j,t} \left[\tau_{b,1}\epsilon_{b,t} + \tau_{b,2}\kappa(\epsilon_{b,t}) \right] \mathrm{d}G \\ &\quad + \sum_{k=1}^{K} \sum_{b=1}^{K} \zeta_{l,k,k,\infty} \zeta_{l,b,b,\infty} \int \left[\tau_{b,1}\epsilon_{b,t} + \tau_{b,2}\kappa(\epsilon_{b,t}) \right] \left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t}) \right] \mathrm{d}G,\end{aligned}$$

which is finite by Assumption 3.3.1. By Proposition 2.29 in van der Vaart (1998) we conclude that $\int (f_{\alpha,n} - f_\alpha)^2 \,\mathrm{d}\lambda \to 0$. Exactly, the same arguments holds for $r = \sigma, c, b$ and we omit the details. The convergence of each $\int (f_{r,n} - f_r)^2 \,\mathrm{d}\lambda \to 0$ in conjunction with equation (32) is sufficient for part 4. □

LEMMA A.2: *Suppose that Assumption 3.3.1 holds and let $k, j, s, b \in [K]$ with $j \neq k$ and $s \neq b$. Then, for any $p \in [1, 2]$ we have that*

(i) $\|\phi_k(\epsilon_k)\epsilon_j\phi_s(\epsilon_s)\epsilon_b\|_{G,p} < \infty$,

(ii) $\|\phi_k(\epsilon_k)\epsilon_j\epsilon_s\|_{G,p} < \infty$,

(iii) $\|\epsilon_k\epsilon_s\|_{G,p} < \infty$.

Proof. By Cauchy-Schwarz, independence and our moment conditions we have

$$\begin{aligned}\|\phi_k(\epsilon_k)\epsilon_j\phi_s(\epsilon_s)\epsilon_b\|_{G,p} &\leq \left[G[\phi_k(\epsilon_k)]^{2p} G[\epsilon_j]^{2p} G[\phi_s(\epsilon_s)]^{2p} G[\epsilon_b]^{2p} \right]^{\frac{1}{2p}} < \infty, \\ \|\phi_k(\epsilon_k)\epsilon_j\epsilon_s\|_{G,p} &\leq \left[G[\phi_k(\epsilon_k)]^{2p} G[\epsilon_j]^{2p} G[\epsilon_s]^{2p} \right]^{1/(2p)} < \infty, \\ \|\epsilon_k\epsilon_s\|_{G,p} = \|(\epsilon_k)^p(\epsilon_s)^p\|_{G,1}^{1/p} &\leq \|(\epsilon_k)^p\|_{G,2}^{1/p} \|(\epsilon_s)^p\|_{G,2}^{1/p} < \infty.\end{aligned}$$

□

LEMMA A.3: *Suppose that Assumption 3.3.1 holds and let $k, j, s \in [K]$ with $j \neq k$. Then, for $1 \leq p \leq \min(1 + \delta/4, 2)$, we have*

(i) $\|\phi_k(\epsilon_k)\epsilon_j\kappa(\epsilon_s)\|_{G,p} < \infty$,

(ii) $\|\epsilon_k\kappa(A_sY)\|_{G,p} < \infty$,

(iii) $\|\kappa(\epsilon_k)\kappa(A_sY)\|_{G,p} < \infty$.

Proof. By Cauchy-Schwarz, independence and our assumed moment conditions we have

$$\begin{aligned}
\|\phi_k(\epsilon_k)\epsilon_j\kappa(\epsilon_s)\|_{G,p} &\leq \left[\left[G[\phi_k(\epsilon_k)]^{2p}G[\epsilon_s]^{4p}\right]^{1/(2p)} + \|\phi_k(\epsilon_k)\|_{G,p}\right]\|\epsilon_j\|_{G,p} < \infty, \\
\|\epsilon_k\kappa(A_sY)\|_{G,p} &\leq \|(\epsilon_k)^p\|_{G,2}^{1/p}\|(\epsilon_s)^{2p}\|_{G,2}^{1/p} + \|\epsilon_k\|_{G,p} < \infty, \\
\|\kappa(\epsilon_k)\kappa(A_sY)\|_{G,p} &\leq \|(\epsilon_k)^{2p}\|_{G,2}^{1/p}\|(\epsilon_s)^{2p}\|_{G,2}^{1/p} + \|(\epsilon_k)^2\|_{G,p} + \|(\epsilon_s)^2\|_{G,p} + 1 < \infty.
\end{aligned}$$

□

LEMMA A.4: *Define*

$$\begin{aligned}
q^{\alpha}_{l,t,n} &\coloneqq \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\zeta^{\alpha}_{n,l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t} + \sum_{k=1}^{K}\zeta^{\alpha}_{n,l,k,k}\left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\right] \\
q^{\sigma}_{l,t,n} &\coloneqq \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\zeta^{\sigma}_{n,l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t} + \sum_{k=1}^{K}\zeta^{\sigma}_{n,l,k,k}\left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\right] \\
q^{c}_{l,t,n} &\coloneqq -\sum_{k=1}^{K}\phi_k(\epsilon_{k,t})A_{n,k\bullet}D_{c,l} \\
q^{b}_{l,t,n} &\coloneqq -\sum_{k=1}^{K}\phi_k(\epsilon_{k,t})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)
\end{aligned}$$

where the dependence of e.g. $\zeta_{n,l,k,j}$ on n is as in Footnote 18. Let $Q^{r,s}_{l,m,t,n} \coloneqq q^{r}_{l,t,n}q^{s}_{m,t,n}$. Suppose that Assumption 3.3.1 holds. Then, for $1 \leq p \leq \min(1+\delta/4, 2)$ we have $\|Q^{r,s}_{l,m,t,n}\|_{G,p} < \infty$.

Proof. We only consider $r, s \in \{\alpha, c, b\}$ as σ and α are interchangeable. By definition we have

$$
\begin{aligned}
Q^{\alpha,\alpha}_{l,m,t,n} &= \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\sum_{s=1}^{K}\sum_{b=1,b\neq s}^{K} \zeta^{\alpha}_{l,k,j,n}\zeta^{\alpha}_{m,s,b,n}\phi_k(\epsilon_{k,t})\epsilon_{j,t}\phi_s(\epsilon_{s,t})\epsilon_{b,t} \\
&\quad + \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\sum_{s=1}^{K} \zeta^{\alpha}_{l,k,j,n}\zeta^{\alpha}_{m,s,s,n}\phi_k(\epsilon_{k,t})\epsilon_{j,t}[\tau_{s,1}\epsilon_{s,t} + \tau_{s,2}\kappa(\epsilon_{s,t})] \\
&\quad + \sum_{s=1}^{K}\sum_{b=1,j\neq s}^{K}\sum_{k=1}^{K} \zeta^{\alpha}_{l,s,b,n}\zeta^{\alpha}_{m,k,k,n}\phi_s(\epsilon_{s,t})\epsilon_{b,t}[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})] \\
&\quad + \sum_{k=1}^{K}\sum_{s=1}^{K} \zeta^{\alpha}_{l,k,k,n}\zeta^{\alpha}_{m,s,s,n}[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})][\tau_{s,1}\epsilon_{s,t} + \tau_{s,2}\kappa(\epsilon_{s,t})]. \\
Q^{\alpha,c}_{l,m,t,n} &= -\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\sum_{s=1}^{K} \zeta^{\alpha}_{l,k,j,n}\phi_k(\epsilon_{k,t})\phi_s(\epsilon_{s,t})\epsilon_{j,t}A_{n,s\bullet}D_{c,m} \\
&\quad - \sum_{k=1}^{K}\sum_{s=1}^{K} \zeta^{\alpha}_{l,k,k,n}\left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\right]\phi_s(\epsilon_{s,t})A_{n,s\bullet}D_{c,m} \\
Q^{\alpha,b}_{l,m,t,n} &= -\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\sum_{s=1}^{K} \zeta^{\alpha}_{l,k,j,n}\phi_k(\epsilon_{k,t})\phi_s(\epsilon_{s,t})\epsilon_{j,t}A_{n,s\bullet}D_{b,m}(X_t - \mathbb{E}_{\theta_n}X_t) \\
&\quad - \sum_{k=1}^{K}\sum_{s=1}^{K} \zeta^{\alpha}_{l,k,k,n}\left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\right]\phi_s(\epsilon_{s,t})A_{n,s\bullet}D_{b,m}(X_t - \mathbb{E}_{\theta_n}X_t) \\
Q^{c,c}_{l,m,t,n} &= \sum_{k=1}^{K}\sum_{s=1}^{K} \phi_k(\epsilon_{k,t})\phi_s(\epsilon_{s,t})A_{n,k\bullet}D_{c,l}A_{n,s\bullet}D_{c,m} \\
Q^{c,b}_{l,m,t,n} &= \sum_{k=1}^{K}\sum_{s=1}^{K} \phi_k(\epsilon_{k,t})\phi_s(\epsilon_{s,t})A_{n,k\bullet}D_{c,l}A_{n,s\bullet}D_{b,m}(X_t - \mathbb{E}_{\theta_n}X_t) \\
Q^{b,b}_{l,m,t,n} &= \sum_{k=1}^{K}\sum_{s=1}^{K} \phi_k(\epsilon_{k,t})\phi_s(\epsilon_{s,t})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)A_{n,s\bullet}D_{b,m}(X_t - \mathbb{E}_{\theta_n}X_t)
\end{aligned}
$$

Hence, by Minkowski's inequality, the independence of X_i (with finite second moments) and Lemmas A.2 & A.3, $\|Q^{r,s}_{l,m,i,n}\|_{G,p} < \infty$. □

LEMMA A.5: *If Assumption 3.3.1 holds, then* $\|\hat{\tau}_{k,n} - \tau_{k,n}\|_2 = o_{P_{\theta_n}}(\nu_{n,p}) = o_{P_{\theta_n}}(\nu_n^{1/2})$, *where* θ_n *is as in Lemma A.1.*

Proof. Under P_{θ_n}, $A_{n,k}(Y_t - c_n - B_nX_t) \approx \epsilon_{k,t} \sim \eta_k$, hence the claim will follow if we show that $\check{\tau}_{k,n} - \check{\tau}_k = o_G(\nu_n^{1/2})$, where

$$
\check{\tau}_{k,n} \coloneqq \check{M}^{-1}_{k,n}\begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \text{where } \check{M}_{k,n} \coloneqq \begin{pmatrix} 1 & \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{k,t})^3 \\ \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{k,t})^3 & \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{k,t})^4 - 1 \end{pmatrix},
$$

and

$$\breve{\tau}_{k,n} \coloneqq \breve{M}_{k,n}^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \text{where } \breve{M}_{k,n} \coloneqq \begin{pmatrix} 1 & G(\epsilon_{k,i})^3 \\ G(\epsilon_{k,i})^3 & G(\epsilon_{k,i})^4 - 1 \end{pmatrix}.$$

Let $w \coloneqq (0, -2)'$. By the preceding definitions and the fact that the map $M \mapsto M^{-1}$ is Lipschitz at a positive definite matrix M_0 we have that for a positive constant C then for large enough n, with probability approaching one

$$\|\breve{\tau}_{k,n} - \breve{\tau}_{k,n}\|_2 = \|(\breve{M}_{k,n}^{-1} - \breve{M}_k^{-1})w\|_2 \leq 2\|\breve{M}_{k,n}^{-1} - \breve{M}_k^{-1}\|_2 \leq 2C\|\breve{M}_{k,n} - \breve{M}_k\|_2. \quad (33)$$

If $\upsilon \coloneqq \delta/4 \geq 4$, we have that by Theorem 2.5.11 in Durrett (2019)

$$\frac{1}{n}\sum_{t=1}^{n}[(\epsilon_{k,t})^3 - G(\epsilon_{k,t})^3] = o_G\left(n^{-1/2}\log(n)^{1/2+\iota}\right)$$
$$\frac{1}{n}\sum_{t=1}^{n}[(\epsilon_{k,t})^4 - G(\epsilon_{k,t})^4] = o_G\left(n^{-1/2}\log(n)^{1/2+\iota}\right)$$

for $\iota > 0$, which implies that

$$\|\breve{M}_{k,n} - \breve{M}_k\|_2 \leq \|\breve{M}_{k,n} - \breve{M}_k\|_F = o_G\left(n^{-1/2}\log(n)^{1/2+\iota}\right).$$

If $0 < \upsilon < 4$, we have by Theorems 2.5.11 & 2.5.12 in Durrett (2019) that for $\iota > 0$,

$$\frac{1}{n}\sum_{t=1}^{n}[(\epsilon_{k,t})^3 - G(\epsilon_{k,t})^3] = \begin{cases} o_G\left(n^{-1/2}\log(n)^{1/2+\iota}\right) & \text{if } \upsilon \in [2,4) \\ o_G\left(n^{\frac{1-p}{p}}\right) & \text{if } \upsilon \in (0,2) \end{cases},$$
$$\frac{1}{n}\sum_{t=1}^{n}[(\epsilon_{k,t})^4 - G(\epsilon_{k,t})^4] = o_G\left(n^{\frac{1-p}{p}}\right).$$

which together imply that

$$\|\breve{M}_{k,n} - \breve{M}_k\|_2 \leq \|\breve{M}_{k,n} - \breve{M}_k\|_F = o_G\left(n^{\frac{1-p}{p}}\right).$$

Combining these convergence rates with equation (33) yields the result in light of the observations made at the beginning of the proof. □

LEMMA A.6: *Suppose Assumptions 3.3.1 and 3.3.2 hold and* $\theta_n = (\alpha_0, \beta_n, \eta)$ *where* $\sqrt{n}(\beta_n - \beta) = O(1)$ *is a deterministic sequence. Then for each* $r \in \{\alpha, \sigma, c, b\}$ *and* l

$$\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\theta_n,r_l}(Y_t, X_t) - \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)\right)^2 = o_{P_{\theta_n}}(\nu_n).$$

Proof. We start by considering elements in $\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\theta_n,\alpha_l}(Y_t, X_t) - \tilde{\ell}_{\theta_n,\alpha_l}(Y_t, X_t)\right)^2$. We define $\tilde{\tau}_{k,n,q} \coloneqq \hat{\tau}_{k,n,q} - \tau_{k,q}$ and $V_{t,n} = Y_t - c_n - B_n X_t$. Since each $|\zeta_{n,l,k,j}^{\alpha}| < \infty$ and the

sums over k, j are finite, it is sufficient to demonstrate that for every $k, j, m, s \in [K]$, with $k \neq j$ and $s \neq m$,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})\right]\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})\right] \\ \times A_{n,j\bullet}V_{t,n}A_{n,m\bullet}V_{t,n} = o_{P_{\theta_n}}(\nu_n),\end{aligned} \tag{34}$$

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})\right]A_{n,j\bullet}V_{t,n} \\ \times \left[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{t,n} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{t,n})\right] = o_{P_{\theta_n}}(\nu_n),\end{aligned} \tag{35}$$

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{t,n} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{t,n})\right]\left[\tilde{\tau}_{k,n,1}A_{n,k\bullet}V_{t,n} + \tilde{\tau}_{k,n,2}\kappa(A_{n,k\bullet}V_{t,n})\right] \\ = o_{P_{\theta_n}}(\nu_n).\end{aligned} \tag{36}$$

For (36), let $\xi_1(x) = x$ and $\xi_2(x) = \kappa(x)$. Then, we can split the sum into 4 parts, each of which has the following form for some $q, w \in \{1, 2\}$

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w}\xi_q(A_{n,s\bullet}V_{t,n})\xi_w(A_{n,k\bullet}V_{t,n}) \\ &= \tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w}\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{t,n})\xi_w(A_{n,k\bullet}V_{t,n}) \\ &= o_{P_{\theta_n}}(\nu_n),\end{aligned}$$

since we have that each $\tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w} = o_{P_{\theta_n}}(\nu_n)$ by Lemma A.5.[19] For (35) we can argue similarly. Again let $\xi_1(x) = x$ and $\xi_2(x) = \kappa(x)$. Then, we can split the sum into 2 parts, each of which has the following form for some $q \in \{1, 2\}$

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})\right]A_{n,j\bullet}V_{t,n}\tilde{\tau}_{s,n,q}\xi_q(A_{n,s\bullet}V_{t,n}) \\ &\quad\leq \tilde{\tau}_{s,n,q}\left(\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})\right]^2(A_{n,j\bullet}V_{t,n})^2\right)^{1/2} \\ &\qquad\times\left(\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{t,n})^2\right)^{1/2} \\ &\quad= o_{P_{\theta_n}}(1).\end{aligned}$$

[19] The fact that $\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{t,n})\xi_w(A_{n,k\bullet}V_{t,n}) = O_{P_{\theta_n}}(1)$ can be seen to hold using the moment and i.i.d. assumptions from Assumption 3.3.1 and Markov's inequality, noting once more that $A_{n,k\bullet}V_{t,n} \simeq \epsilon_{k,t}$ under P_{θ_n}.

by Assumption 3.3.2 applied with $W_{t,n} = A_{n,j\bullet}V_{t,n}$ and $\tilde{\tau}_{s,n,q} = o_{P_{\theta_n}}(\nu_n^{1/2})$.[20] For (34) use Cauchy-Schwarz with Assumption 3.3.2:

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})\right]\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})\right] \\
&\qquad \times A_{n,j\bullet}V_{t,n}A_{n,m\bullet}V_{t,n} \\
&\leq \left(\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})\right]^2 (A_{n,j\bullet}V_{t,n})^2\right)^{1/2} \\
&\qquad \times \left(\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})\right]^2 (A_{n,m\bullet}V_{t,n})^2\right)^{1/2} \\
&= o_{P_{\theta_n}}(\nu_n).
\end{aligned}$$

This completes the proof for the components corresponding to α_l. We note that the components corresponding to σ_l follow identically.

Next, we consider the elements in $\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\ell}_{\theta_n,c_l}(Y_t,X_t) - \tilde{\ell}_{\theta_n,c_l}(Y_t,X_t)\right)^2$. By Cauchy-Schwarz with Assumption 3.3.2 we have that for any $k, s \in [K]$

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})][\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})]A_{n,k\bullet}D_{c,l}A_{n,s\bullet}D_{c,l} \\
&\qquad \left(\frac{1}{n}\sum_{t=1}^{n}[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})]^2(A_{n,k\bullet}D_{c,l})^2\right)^{1/2} \\
&\qquad\qquad \times \left(\frac{1}{n}\sum_{t=1}^{n}[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})]^2(A_{n,s\bullet}D_{c,l})^2\right)^{1/2} \\
&\qquad = o_{P_{\theta_n}}(\nu_n)
\end{aligned}$$

Finally, we consider the elements in $\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\theta_n,b_l}(Y_t,X_t) - \tilde{\ell}_{\theta_n,b_l}(Y_t,X_t)\right)^2$, where we note that

$$\begin{aligned}
\hat{\ell}_{\theta_n,b_l}(Y_t,X_t) - \tilde{\ell}_{\theta_n,b_l}(Y_t,X_t) &= \sum_{k=1}^{K}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t) \\
&\quad - \phi_k(A_{n,k\bullet}V_{t,n})A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t) \\
&= \sum_{k=1}^{K}[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})]A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)
\end{aligned}$$

[20] See Footnote 19.

By Cauchy-Schwarz with Assumption 3.3.2 we have for any $k, s \in [K]$

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})][\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})] \\
&\qquad \times A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t)A_{n,s\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t) \\
&\leq \left(\frac{1}{n}\sum_{t=1}^{n}[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{t,n}) - \phi_k(A_{n,k\bullet}V_{t,n})]^2(A_{n,k\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t))^2\right)^{1/2} \\
&\qquad \times \left(\frac{1}{n}\sum_{t=1}^{n}[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{t,n}) - \phi_s(A_{n,s\bullet}V_{t,n})]^2(A_{n,s\bullet}D_{b,l}(X_t - \mathbb{E}_{\theta_n}X_t))^2\right)^{1/2} \\
&= o_{P_{\theta_n}}(\nu_n)
\end{aligned}$$

□

B Density score estimation

In this section, we discuss the estimation of the log density scores using the B-spline methodology developed in Jin (1992) and Chen and Bickel (2006). We first discuss the estimator and then show that this estimate satisfies Assumption 3.3.2 under mild additional assumptions on the densities η_k.

B-spline estimator

Let $\xi_1 < \cdots < \xi_N$ be a knot sequence, the first order B-splines are defined according to $b_i^{(1)}(x) \coloneqq \mathbf{1}_{[\xi_i,\xi_{i+1})}(x)$. Subsequent order B-splines can be computed according to the recurrence relation

$$b_i^{(\kappa)}(x) = \frac{x - \xi_i}{\xi_{i+\kappa-1} - \xi_i}b_i^{(\kappa-1)}(x) + \frac{\xi_{i+\kappa} - x}{\xi_{i+\kappa} - \xi_{i+1}}b_{i+1}^{(\kappa-1)}(x), \tag{37}$$

for $\kappa > 1$ and $i = 1, \ldots, N - \kappa$. A κ-th order B-spline is $\kappa - 2$ times differentiable in x with first derivative

$$c_i^{(\kappa)}(x) = \frac{\kappa - 1}{\xi_{i+\kappa-1} - \xi_i}b_i^{(\kappa-1)}(x) - \frac{\kappa - 1}{\xi_{i+\kappa} - \xi_{i+1}}b_{i+1}^{(\kappa-1)}(x). \tag{38}$$

See de Boor (2001) for more details on B-splines.

Let $b_{k,n} = (b_{k,n,1}, \ldots, b_{k,n,B_{k,n}})'$ be a collection of $B_{k,n}$ cubic B-splines and let $c_{k,n} = (c_{k,n,1}, \ldots, c_{k,n,B_{k,n}})'$ be their derivatives: $c_{k,n,i}(x) \coloneqq \frac{\mathrm{d}b_{k,n,i}(x)}{\mathrm{d}x}$ for each $i \in [B_{k,n}]$. Let $\gamma_k \in \mathbb{R}^{B_{k,n}}$. The knots of the splines, $\xi_{k,n} = (\xi_{k,n,i})_{i=1}^{K_{k,n}}$ are equally spaced in $[\Xi_{k,n}^L, \Xi_{k,n}^U]$ with $\delta_{k,n} \coloneqq \xi_{k,n,i+1} - \xi_{k,n,i} > 0$.[21] For each (k, n) pair, the relationship between the number of knots ($K_{k,n}$), the number of spline functions ($B_{k,n}$) and $\delta_{k,n}$ are

[21] For each $k \in [K]$ the sequences $(\Xi_{k,n}^L)_{n\in\mathbb{N}}$, $(\Xi_{k,n}^U)_{n\in\mathbb{N}}$, $(B_{k,n})_{n\in\mathbb{N}}$ and $(\delta_{k,n})_{n\in\mathbb{N}}$ are deterministic.

given by $B_{k,n} = K_{k,n} - 4$ and $K_{k,n} = 1 + (\Xi^U_{k,n} - \Xi^L_{k,n})/\delta_{k,n}$.[22]

Since the B-splines vanish at infinity for any $n \in \mathbb{N}$, integration by parts gives that

$$\begin{aligned}\int (\phi_k(z) - \gamma_k' b_{k,n}(z))^2 \eta_k(z)\,\mathrm{d}z &= \int \phi_k^2\,\mathrm{d}G_k + \int (\gamma_k' b_{k,n})^2\,\mathrm{d}G_k \\ &\quad + 2\int \gamma_k' c_{k,n}(z)\eta_k(z)\,\mathrm{d}z \\ &= G_k\phi_k^2 + \gamma_k' G_k[b_{k,n}b_{k,n}']\gamma_k + 2\gamma_k' G_k c_{k,n}.\end{aligned} \tag{39}$$

The solution to minimizing this mean-squared error is given by:[23]

$$\gamma_{k,n} = -G_k[b_{k,n}b_{k,n}']^{-1}G_k c_{k,n}. \tag{40}$$

Replacing the population expectations with sample counterparts we arrive at our estimate of γ_k

$$\hat{\gamma}_{k,n} \coloneqq -\left[\frac{1}{n}\sum_{t=1}^n b_{k,n}(\epsilon_{k,t})b_{k,n}(\epsilon_{k,t})'\right]^{-1}\frac{1}{n}\sum_{t=1}^n c_{k,n}(\epsilon_{k,t}), \tag{41}$$

where $\epsilon_{k,t}$ is set equal to $A_{n,k\bullet}(Y_t - c_n - B_n X_t)$, which under P_{θ_n} has the same distribution. Our estimate for ϕ_k is given by

$$\hat{\phi}_{k,n}(z) \coloneqq \hat{\gamma}_{k,n}' b_{k,n}(z)\ . \tag{42}$$

We note that computing (42) effective only requires computing the B-spline regression coefficients $\hat{\gamma}_{k,n}$ in (41). To implement the score test we need to estimate K density scores, hence the computational costs is quite modest.

Theoretical results

In this section we demonstrate that the log density score estimate (42) satisfies Assumption 3.3.2 under Assumption 3.3.1 if the following conditions hold. The proof for the following lemma is given in Lee and Mesters (2021), see their Appendix B.

LEMMA B.1: *Let $\phi_{k,n} \coloneqq \phi_k \mathbf{1}_{[\Xi^L_{k,n}, \Xi^U_{k,n}]}$ and $\Delta_{k,n} \coloneqq \Xi^U_{k,n} - \Xi^L_{k,n}$ and suppose that for ν_n as in Assumption 3.3.2, $[\Xi^L_{k,n}, \Xi^U_{k,n}] \uparrow \tilde{\Xi} \supset \mathrm{supp}(\eta_k)$ and $\delta_{k,n} \downarrow 0$ such that*

(i) $G_k(\epsilon_k \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]) = o(\nu_n^2)$;

(ii) *For some* $\iota > 0$, $n^{-1}\Delta_{k,n}^{2+2\iota}\delta_{k,n}^{-(8+2\iota)} = o(\nu_n)$;

(iii) η_k *is bounded (*$\|\eta_k\|_\infty < \infty$*) and differentiable, with a bounded derivative:* $\|\eta_k'\|_\infty < \infty$;

[22] Implicitly we choose $K_{k,n}$ and the endpoints and $\delta_{k,n}$ adjusts such that these formulae hold; this way we do not need to adjust anything to ensure these are integers.

[23] This differs from the expression in Chen and Bickel (2006) by a factor of -1 as they estimate $-\phi_k$.

(iv) For each n, $\phi_{k,n}$ is three-times continuously differentiable on $[\Xi_{k,n}^L, \Xi_{k,n}^U]$ and $\|\phi_{k,n}^{(3)}\|_\infty^2 \delta_{k,n}^6 = o(\nu_n)$;[24]

(v) There are $c > 0$ and $N \in \mathbb{N}$ such that for $n \geq N$ we have $\inf_{t \in [\Xi_{k,n}^L, \Xi_{k,n}^U]} |\eta_k(t)| \geq c\delta_{k,n}$.

Then, under Assumption 3.3.1, the estimates $\hat{\phi}_{k,n}$ satisfy Assumption 3.3.2.

C Algorithm to compute IRF confidence bands

This appendix provides details on the algorithm that is used to compute the identification-robust confidence-bands for the impulse responses using the semi-parametric inference approach.

Recall the definition of the SVAR model from Section 3.3:

$$Y_t = c + B(b)X_t + A(\alpha, \sigma)^{-1} \epsilon_t$$

where the parameter vectors are defined as $\gamma = (\alpha, \beta)$ and $\beta = (\sigma, c, b)$.

Let $\delta \coloneqq (\sigma, b)$, then for a fixed value $\alpha = \alpha_0$, the $(K^2 \times 1)$ vector function collecting all impulse responses at horizon h is defined as a function of δ:

$$\frac{\partial Y_{t+h}}{\partial \epsilon_t} \coloneqq g_h(\delta) \equiv \text{vec}\left(J\boldsymbol{B}(b)^h J' A^{-1}(\alpha_0, \sigma) \right) \tag{43}$$

where $J \coloneqq \begin{bmatrix} I_K, & 0_{K \times K(p-1)} \end{bmatrix}$ and $\boldsymbol{B}$ is the companion matrix:

$$\boldsymbol{B} \equiv \begin{pmatrix} B_1 & B_2 & \dots & B_{p-1} & B_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{pmatrix} \tag{44}$$

Define the gradient of $g_h(\delta)$ as the $(K^2 \times (K_\sigma + K_b))$ matrix

$$\nabla g_h(\delta) \coloneqq \begin{pmatrix} \partial g_h(\delta)/\partial \sigma', & \partial g_h(\delta)/\partial b' \end{pmatrix} \tag{45}$$

The formulae to compute the elements of the gradient matrix are reported at the end of this section.

[24] The differentiability and continuity requirements at the end-points are one-sided.

Algorithm

For a set of sign restrictions fixing the column permutation and signs of the elements in $A^{-1}(\sigma, \alpha)$, the algorithm to compute the $(1-\zeta)$-level confidence interval for the impulse responses to each of the structural shocks, ϵ_t, is as follows.

1. Define a K_α-dimensional grid of permissible values for α and discretize it into G_α gridpoints, $\alpha_1^g, \ldots, \alpha_{G_\alpha}^g$. If possible, construct the grid so that all grid points satisfy the sign restrictions on $A^{-1}(\alpha, \sigma)$ that are required in Step 4.

2. For a given gridpoint, α_j^g, estimate the nuisance parameter vector $\beta = (\sigma, c, b)$ by the one-step efficient estimate which updates the OLS estimate of β using one Gauss-Newton iteration based on the efficient information matrix, evaluated at the OLS estimates.

3. Using the nuisance parameter estimates $\hat{\beta}$, conduct the semi-parametric score test for $H_0 : \alpha = \alpha_j^g$ vs. $H_A : \alpha \neq \alpha_j^g$ defined in Section 3.4. If the test for α_j^g rejects, repeat Step 2 with the next gridpoint, otherwise continue with Step 4.

4. If the test for α_j^g does not reject (and $A^{-1}(\alpha_j^g, \hat{\sigma})$ satisfies the sign restrictions), carry out the following steps for each $h = 0, 1, 2, \ldots, H$.

   (a) Compute the vector of impulse responses $g_h(\delta)$ using the one-step estimates of δ (which is a subvector of β) and $\alpha = \alpha_j^g$.

   (b) Compute and store the $(K^2 \times 1)$ vectors of upper and lower $(1-\zeta/2)$ pointwise asymptotic confidence intervals obtained as

   $$\mathrm{vec}(g_h(\hat{\delta})) \pm z_{\zeta/2}\, \hat{\sigma}_{\mathrm{vec}(g_h)} \qquad h = 0, 1, 2, \ldots, H$$

   where

   $$\hat{\sigma}_{\mathrm{vec}(g_h)} \coloneqq \sqrt{\mathrm{diag}\left(\nabla g_h(\hat{\delta})\, \Sigma_{\hat{\delta}} \nabla\, g_h(\hat{\delta})'\right)/T}$$

   and $z_{\zeta/2}$ denotes the $\zeta/2$ quantile point of the standard normal distribution. Note that $\Sigma_{\hat{\delta}}$ can be computed as the inverse of the efficient information matrix evaluated at the one-step efficient estimates of δ. The formulae to compute $\nabla g_h(\hat{\delta})$ are reported in the section below.

5. Repeat steps 2-4 for every of the G_α gridpoints.

6. Compute the $(1-\zeta)$ IRF confidence interval as the infimum and supremum of the stored upper and lower bounds in Step 4.

Formulae to compute IRF gradient matrix

Recall the definition of the impulse response function at horizon h in eq. (43) above. To derive the expressions for the components of $\nabla g_h(\delta)$, defined in eq. (45), recall that for matrices A, B, C of dimensions $(k \times l), (l \times m), (m \times n)$, the following vectorization rules hold:

$$\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B) \qquad\qquad \operatorname{vec}(ABC) = (I_n \otimes AB)\operatorname{vec}(C)$$

Using the first vectorization rule to compute the gradient of the impulse response function with respect to b, we get:

$$\begin{aligned} \frac{\partial g_h(\delta)}{\partial b'} &\equiv \left[(A(\alpha_0,\sigma)^{-1})' \otimes I_K\right] \frac{\partial \operatorname{vec}(J\boldsymbol{B}^h J')}{\partial b'} \\ &= \left[(A(\alpha_0,\sigma)^{-1})' \otimes I_K\right] \left\{ \sum_{j=0}^{h-1} \left[J(\boldsymbol{B}')^{h-1-j} \otimes (J\boldsymbol{B}^j J')\right] \right\} \end{aligned} \tag{46}$$

Similarly, using the second vectorization rule to compute the gradient of the impulse response function with respect to σ, we get:

$$\frac{\partial g_h(\delta)}{\partial \sigma'} \equiv \left[I_K \otimes J\boldsymbol{B}^h J'\right] \frac{\partial \operatorname{vec}(A(\alpha_0,\sigma)^{-1})}{\partial \sigma'} \tag{47}$$

Note that the expression for $(\partial \operatorname{vec}(A(\alpha_0,\sigma)^{-1}))/(\partial \sigma')$ depends on the specific parameterization that is chosen for $A^{-1}(\alpha,\sigma)$. For instance, using the parameterization defined in example 3.3.1 that $A^{-1}(\alpha,\sigma) = \Sigma^{1/2}R(\alpha)$, we get

$$\frac{\partial \operatorname{vec}(A(\alpha_0,\sigma)^{-1})}{\partial \sigma'} = \left[\nabla_{\sigma_l} \Sigma^{1/2}\ R(\alpha_0)\right]_{l=1,\ldots,K_\sigma} \tag{48}$$

D Non-Gaussian distributions

This section reports the non-Gaussian distributions that are used to simulate the structural errors in the simulation studies in Section 3.5. Apart from the standard Gaussian and student-t distributions, these distributions are taken from Marron and Wand (1992). For convenience, they are reproduced in Table D.4 below together with their respective abbreviations that are used in the column headers of the tables in Section 3.5. Note that the table reports the unstandardized densities. When we simulate from these distributions, we standardize the random variates by subtracting the theoretical mean and dividing by the respective theoretical standard deviation of the respective distribution.

Table D.4: Non-Gaussian distributions

Abbreviation	Name	Definition
SKU	Skewed Unimodal	$\frac{1}{5}\mathcal{N}\big(0,1\big) + \frac{1}{5}\mathcal{N}\big(\frac{1}{2},(\frac{2}{3})^2\big) + \frac{3}{5}\mathcal{N}\big(\frac{13}{12},(\frac{5}{9})^2\big)$
KU	Kurtotic Unimodal	$\frac{2}{3}\mathcal{N}\big(0,1\big) + \frac{1}{3}\mathcal{N}\big(0,(\frac{1}{10})^2\big)$
BM	Bimodal	$\frac{1}{2}\mathcal{N}\big(-1,(\frac{2}{3})^2\big) + \frac{1}{2}\mathcal{N}\big(1,(\frac{2}{3})^2\big)$
SPB	Separated Bimodal	$\frac{1}{2}\mathcal{N}\big(-\frac{3}{2},(\frac{1}{2})^2\big) + \frac{1}{2}\mathcal{N}\big(\frac{3}{2},(\frac{1}{2})^2\big)$
SKB	Skewed Bimodal	$\frac{3}{4}\mathcal{N}\big(0,1\big) + \frac{1}{4}\mathcal{N}\big(\frac{3}{2},(\frac{1}{3})^2\big)$
TRI	Trimodal	$\frac{9}{20}\mathcal{N}\big(-\frac{6}{5},(\frac{3}{5})^2\big) + \frac{9}{20}\mathcal{N}\big(\frac{6}{5},(\frac{3}{5})^2\big) + \frac{1}{10}\mathcal{N}\big(0,(\frac{1}{4})^2\big)$

Notes: The table reports the (unstandardized) distributions that are used in the simulation studies in Section 3.5 to draw the structural errors. The distributions are taken from Marron and Wand (1992) and reproduced here for convenience, see their Table 1.

Bibliography

Amari, S. and Cardoso, J.-F. (1997). Blind source separation - Semiparametric statistical approach. *IEEE Transactions On Signal Processing*, 45(11).

Andrade, P. and Ferroni, F. (2021). Delphic and odyssean monetary policy shocks: Evidence from the euro area. *Journal of Monetary Economics*, 117:816–832.

Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica: Journal of the Econometric Society*, 55(6):1465–1471.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Andrews, D. W. K. and Cheng, X. (2012). Estimation and inference with weak, semi-strong and strong identification. *Econometrica*, 80(5).

Andrews, I. and Mikusheva, A. (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, 84(4).

Andrews, I., Stock, J., and Sun, L. (2019). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*, 11:727–753.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, 20(3):651–707.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, 91(2):299–323.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Bai, J. and Perron, P. (2006). *Multiple structural change models: A simulation analysis*, page 212–238. Cambridge University Press.

Bauer, M. and Swanson, E. (2020). The fed's response to economic news explains the 'fed information effect'. *CESifo Working Paper Series*, 8151.

Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999.

Bekaert, G., Engstrom, E., and Ermolov, A. (2019). Macro Risks and the Term Structure of Interest Rates. *Working paper*.

Bekaert, G., Engstrom, E., and Ermolov, A. (2020). Aggregate Demand and Aggregate Supply Effects of COVID-19: A Real-time Analysis. *Working paper*.

Ben-Israel, A. and Greville, T. N. E. (2003). *Generalized Inverses: Theory and Applications*. Springer, New York, NY, USA.

Berge, T. J., Chang, A. C., and Sinha, N. R. (2019). Evaluating the conditionality of judgmental forecasts. *International Journal of Forecasting*, 35(4):1627–1635.

Bhatia, R. (1997). *Matrix Analysis*. Springer, New York, NY, USA.

Bickel, P. J., Klaasen, C. A. J., Ritov, Y., and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York, NY, USA.

Bickel, P. J., Ritov, Y., and Stoker, T. M. (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *Annals of Statistics*, 34(2):721–741.

Billingsley, P. (1995). *Probability and Measure*. Wiley.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley.

Braun, R. (2021). The importance of supply and demand for oil prices: Evidence from a svar identified by non-gaussianity. *Working paper*.

Bundick, B. and Smith, A. L. (2020). The dynamic effects of forward guidance shocks. *The Review of Economics and Statistics*, 102(5):946—-965.

Campbell, J. R., Evans, C. L., Fisher, J. D., Justiniano, A., Calomiris, C. W., and Woodford, M. (2012). Macroeconomic effects of federal reserve forward guidance [with comments and discussion]. *Brookings Papers on Economic Activity*, pages 1–80.

Campbell, J. R., Fisher, J. D., Justiniano, A., and Melosi, L. (2017). Forward guidance and macroeconomic outcomes since the financial crisis. *NBER Macroeconomics Annual*, 31(1):283–357.

Campbell, J. Y. and Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012). Model selection when there are multiple breaks. *Journal of Econometrics*, 169(2):239–246.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews*, 23(3):259–292.

Chen, A. and Bickel, P. J. (2006). Efficient Independent Component Analysis. *Annals of Statistics*, 34(6).

Chinco, A., Clark-Joseph, A. D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492.

Choi, S., Hall, W. J., and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Annals of Statistics*, 24(2):841–861.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics*, 1:65–148.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2018). On dsge models. *Journal of Economic Perspectives*, 32(3):113–40.

Cieslak, A. and Schrimpf, A. (2019). Non-monetary news in central bank communication. *Journal of International Economics*, 118(C):293–315.

Clark, T. and McCracken, M. (2013). Chapter 20 - advances in forecast evaluation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 1107 – 1201. Elsevier.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *The Review of Financial Studies*, 21(4):1533–1575.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics*, 4(2):1–32.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, 36:287–314.

Croushore, D. (2012). Comment-forecast rationality tests based on multi-horizon bounds. *Journal of Business and Economic Statistics*, 30(1):17.

D'Agostino, A. and Whelan, K. (2008). Federal reserve information during the great moderation. *Journal of the European Economic Association*, 6(2-3):609–620.

Dagum, L. and Menon, R. (1998). Openmp: An industry standard api for shared-memory programming. *IEEE computational science and engineering*, 5(1):46–55.

de Boor, C. (2001). *A Practical Guide to Splines*. Springer, New York, NY, USA.

Durrett, R. (2019). *Probability Theory and Examples*. Cambridge University Press, Cambridge, UK, 5th edition.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Elliott, G. and Müller, U. K. (2006). Efficient tests for general persistent time variation in regression coefficients. *The Review of Economic Studies*, 73(4):907–940.

Ericsson, N. R. (2016). Predicting fed forecasts. *IFDP Notes*.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25.

Farmer, L., Schmidt, L., and Timmermann, A. (2019). Pockets of predictability. *Available at SSRN 3152386*.

Faust, J. and Wright, J. (2008). Efficient forecast tests for conditional policy forecasts. *Journal of Econometrics*, 146(2):293–303.

Faust, J. and Wright, J. H. (2009). Comparing greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business & Economic Statistics*, 27(4):468–479.

Fiorentini, G. and Sentana, E. (2020). Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. *Working paper*.

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2018). Testing for parameter instability in predictive regression models. *Journal of Econometrics*, 204(1):101–118.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451.

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

Gonzalo, J. and Pitarakis, J.-Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics*, 30(2):229–241.

Gonzalo, J. and Pitarakis, J.-Y. (2017). Inferring the predictability induced by a persistent regressor in a predictive threshold model. *Journal of Business & Economic Statistics*, 35(2):202–217.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics*, 196.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2019). Identification and Estimation in Non-Fundamental Structural VARMA Models. *The Review of Economic Studies*, 87(4):1915–1953.

Guay, A. (2020). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics*. Forthcoming.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American Economic Review*, 95(1):425–436.

Hallin, M. and Saidi, A. (2007). Optimal tests of noncorrelation between multivariate time series. *Journal of the American Statistical Association*, 102(479):938–951.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, pages 489–500.

Harvey, D. I., Leybourne, S. J., Sollis, R., and Taylor, A. R. (2020). Real-time detection of regimes of predictability in the u.s. equity premium. *Journal of Applied Econometrics*, 36(1):45–70.

Hawkins, D. M. (1976). Point estimation of the parameters of piecewise regression models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 25(1):51–57.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560–580.

Herwartz, H. (2019). Long-run neutrality of demand shocks: Revisiting blanchard and quah (1989) with independent structural shocks. *Journal of Applied Econometrics*, 34(5):811–819.

Hoesch, L., Rossi, B., and Sekhposyan, T. (2020). Has the information channel of monetary

policy disappeared? revisiting the empirical evidence. *CEPR Discussion Paper No. DP14456*.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press, 2 edition.

Inoue, A. and Rossi, B. (2021). A new approach to measuring economic policy shocks, with an application to conventional and unconventional monetary policy. *Quantitative Economics*. Forthcoming.

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43.

Jin, K. (1992). Empirical smoothing parameter selection in adaptive estimation. *The Annals of Statistics*, 20(4):1844–1874.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4).

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Lanne, M. and Luoto, J. (2019). Useful prior information in sign-identified structural vector autoregression: Replication of baumeister and hamilton. *Working Paper*.

Lanne, M. and Luoto, J. (2020). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, 39(1):69–81.

Lanne, M. and Lütkepohl, H. (2010). Structural vector autoregressions with nonnormal residuals. *Journal of Business & Economic Statistics*, 28(1):159–168.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structual vector autoregressions. *Journal of Econometrics*, 196(2).

Le Cam, L. M. and Yang, G. L. (2000). *Asypmtotics in Statistics: Some Basic Concepts*. Springer, New York, NY, USA, 2nd edition.

Lee, A. and Mesters, G. (2021). Robust non-gaussian identification and inference for simultaneous equations. *Working Paper*.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849.

Lunsford, K. G. (2020). Policy language and information effects in the early days of federal reserve forward guidance. *American Economic Review*, 110(9):2899–2934.

Magnus, J. R. and Neudecker, H. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics*, pages 712–736.

Maxand, S. (2018). Identification of independent structural shocks in the presence of multiple gaussian components. *Econometrics and Statistics*.

McLeish, D. (1975). Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(3):165–178.

Melosi, L. (2017). Signalling effects of monetary policy. *The Review of Economic Studies*, 84(2):853–884.

Miranda-Agrippino, S. and Ricco, G. (2020). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*. Forthcoming.

Moneta, A., Entner, D., Hoyer, P. O., and Coad, A. (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics*, 75(5):705–730.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Newey, W. K. and McFadden, D. (1994). Chapter 36: Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, pages 2111 – 2245. Elsevier.

Odendahl, F., Rossi, B., and Sekhposyan, T. (2020). Comparing forecast performance with state dependence. *Working Paper*.

Patton, A. J. and Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, 30(1):1–17.

Paul, P. (2020). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics*, 102(4):690–704.

Perron, P. and Qu, Z. (2006). Estimating restricted structural change models. *Journal of Econometrics*, 134(2):373–399.

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance*, 50(4):1201–1228.

Pesaran, M. H. and Timmermann, A. (2000). A recursive modelling approach to predicting uk stock returns. *The Economic Journal*, 110(460):159–191.

Phillips, P. C. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.

Pitarakis, J.-Y. (2017). A simple approach for diagnosing instabilities in predictive regressions. *Oxford Bulletin of Economics and Statistics*, 79(5):851–874.

Pitarakis, J.-Y. and Gonzalo, J. (2019). Predictive regressions. *Oxford Research Encyclopedias (Economics and Finance)*.

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.

Ramey, V. (2016). Macroeconomic Shocks and Their Propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*. Elsevier, Amsterdam, North Holland.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications*. John Wiley & Sons, Inc., New York, NY, USA.

Rapach, D. E. and Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance*, 13(2):231–247.

Romer, C. D. and Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, 90(3):429–457.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4):1055–1084.

Rossi, B. (2005). Optimal tests for nested model selection with underlying parameter instability. *Econometric theory*, 21(5):962–990.

Rossi, B. (2006a). Advances in forecasting under model instability. In *Handbook of Economic Forecasting II*, pages 1203–1324. Elsevier North Holland.

Rossi, B. (2006b). Are exchange rates really random walks? some evidence robust to parameter instability. *Macroeconomic dynamics*, 10(1):20–38.

Rossi, B. (2013). Advances in forecasting under instability. In *Handbook of economic forecasting*, volume 2, pages 1203–1324. Elsevier.

Rossi, B. (2020). Forecasting in the presence of instabilities: How do we know whether

models predict well and how to improve them. *Journal of Economic Literature*. Forthcoming.

Rossi, B. and Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. *Journal of Applied Econometrics*, 31(3):507–532.

Sanderson, C. and Curtin, R. (2016). Armadillo: A template-based c++ library for linear algebra. *Journal of Open Source Software*, 1(2):26.

Sen, A. (2012). On the Interrelation Between the Sample Mean and the Sample Variance. *The American Statistician*, 66(2).

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, pages 1–48.

Sims, C. A. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity*, 2002(2):1–40.

Sims, C. A. (2021). Svar identification through heteroskedasticity with misspecified regimes. *Working paper*.

Sowell, F. (1996). Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica: Journal of the Econometric Society*, pages 1085–1107.

Tank, A., Fox, E. B., and Shojaie, A. (2019). Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series. *Biometrika*, 106(2):433–452.

Timmermann, A. (2006). Forecast combinations. In *Handbook of Economic Forecasting*, pages 135–196. Elsevier North Holland.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24(1):1–18.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York, NY, USA, 1st edition.

van der Vaart, A. W. (2002). Semiparametric statistics. In Bernard, P., editor, *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999*. Springer, Berlin, Germany.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition.

Velasco, C. (2020). Identification and estimation of structural varma models using higher order dynamics. *Working paper*.

Welch, I. and Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

West, K. and McCracken, M. (1998). Regression-based tests of predictive ability. *International Economic Review*, 39(4):817–40.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, 64(5):1067–1084.

Winkler, R. L. and Clemen, R. T. (1992). Sensitivity of weights in combining forecasts. *Operations Research*, 40(3):609–614.