Essays in Macroeconometrics

Yiru Wang

TESI DOCTORAL UPF / year 2019

THESIS SUPERVISOR
Professor Barbara Rossi
Department Departament d'Economia i Empresa

To my parents.

Acknowledgments

This thesis represents a collection of my research at Universitat Pompeu Fabra (UPF) during the past years. I would like to express my sincere gratitude towards the faculty, the colleagues, and my family and friends.

First of all, I would like to express my very great appreciation to my advisor Barbara Rossi. I am indebted to her constant and invaluable guidance, encouragement, and support. Without her brilliant insights and patient guidance, this thesis wouldn't be possible. *Thank you, Barbara! I wish I could become a scholar and teacher just like you.*

I would like to offer my special thanks to Geert Mesters who has always been available when I need advice. His comments have considerably improved my work. His willingness to generously give his time and as his continued help has been very much appreciated.

Among the faculty and professors, I would like to thank in particular Christian Brownlees and Majid Al-Sadoon for their valuable comments and continued support during the past years. I received very helpful comments on my first and second chapters from the participants at the UPF econometrics seminars. Furthermore, it was a great pleasure to collaborate with Atsushi Inoue on the third chapter of this thesis.

I thank the Spanish Ministerio de Economia y Competitividad (FPI grant BES-2016-076868), and UPF and Barcelona GSE scholarships for the financial support. I further thank Marta Araque, Laura Agusti, and Mariona Novoa for being so kind and supportive in administrative duties.

I would like to thank my colleagues and friends Adam, Andre, Bjarni, Donghai, Erqi, Florens, Greg, Guohao, Ines, Luca, Lukas, Manlin, Philipp, Renbin, Shangyu, Shengliang, Wei, Yimei, Yucheng, Yuqi, to name only a few, for the comments, advice, and/or the friendly chats and hangingouts at various points. I thank my officemate and friend Christian H. for sharing all the good and bad news and cheering me up. I thank Qianqian, Shiyan, Yiting, and Yuxin for always being there, and the long-lasting friendship. I thank my friends at Yishu Coffee for the sweets and unforgettable moments. Very special thanks to my lovely flatmate and close friend Yining for the warmest company in the hardest of times.

Finally, I dedicate this thesis to my family, in particular, my parents. I can always count on your unconditional love, support, and encouragement throughout these intensive years. You are the reason for who I am today. I owe so much to both of you. Without you, I wouldn’t have completed this journey. I thank you very deeply for everything! *Love you.*

Abstract

This thesis consists of three chapters on topics in Macroeconometrics. Chapter 1 proposes a method to analyze the relationship between models' in-sample fit and their out-of-sample density forecasting performance. To this end, I further develop a formal test to capture density forecast breakdowns (DFBs); situations in which the out-of-sample density forecast performance is significantly worse than its anticipated performance. Chapter 2 proposes a novel methodology for identifying and estimating structural breaks in the factor loadings of a high dimensional approximate factor model with an unknown number of latent factors. The approach is robust to structural changes in the volatility of the factors (the second moment of the factors), applicable to multiple structural breaks, and easy to implement for practitioners. Chapter 3 introduces time variation into the local projections framework and proposes an impulse responses estimation methodology under unstable local projections.

Resum

Aquesta tesi consta de tres capítols sobre temes en Macroeconometria. El capítol 1 proposa un mètode per analitzar la relació entre l'ajust en mostra de models i el seu rendiment de previsió de densitat fora de mostra. Amb aquesta finalitat, desenvolupo una prova formal per capturar els desglossaments de previsió de densitat (DFB); situacions en què el rendiment previst de la densitat fora de mostra és significativament pitjor que el rendiment previst. El capítol 2 proposa una nova metodologia per identificar i estimar les ruptures estructurals en les càrregues de factors d'un model aproximat dimensional de factor aproximat amb un nombre desconegut de factors latents. L'enfocament és robust a canvis estructurals en la volatilitat dels factors (segon moment dels factors), aplicables a múltiples ruptures estructurals i fàcils d'implementar per als practicants. El capítol 3 introdueix la variació de temps en el marc de les projeccions locals i proposa una metodologia d'estimació de la resposta d'impuls en projeccions locals inestables.

Preface

This thesis consists of three chapters on topics in Macroeconometrics.

In Chapter 1, *Detecting Density Forecast Breakdowns*, I propose a method to analyze the relationship between models' in-sample fit and their out-of-sample density forecasting performance. To this end, I develop a formal test to capture density forecast breakdowns (DFBs); situations in which the out-of-sample density forecast performance is significantly worse than its anticipated performance. This test can be used to evaluate both parametric and non-parametric models. For parametric models, it allows for model misspecification and takes parameter estimation uncertainty into account. For non-parametric models, I provide conditions under which estimation uncertainty is asymptotically irrelevant. The proposed test accommodates conditional density forecast models, which nest unconditional density forecast models, and robust versions are provided for practical use. Monte Carlo results indicate that the test has good size properties in moderately large samples and has power against changes in mean and variance, as well as shifts in distribution type. The empirical study finds that: (i) DFBs occur sporadically in the lower quantiles of the one-quarter-ahead and one-year-ahead predictive densities of real GDP growth in the US, modeled with current financial and economic conditions and skewed-t distributed errors; and (ii) DFBs occur in the one-day-ahead predictive densities for the S&P 500, considering GARCH(1,1) and GARCH-t(1,1) models.

In Chapter 2, *Identification and Estimation of Parameter Instability in a High Dimensional Approximate Factor Model*, I propose a novel methodology for identifying and estimating structural breaks in the factor loadings of a high dimensional approximate factor model with an unknown number of latent factors. The methodology is based on the fact that the sum of the number of pseudo factors in the pre- and post-split subsamples will be minimal if the sample is split at the structural break. I demonstrate that an appropriate transformation of such criteria, based on the eigenvalue ratios of the covariance matrices of the pre- and post-split subsamples, provides a consistent estimator of structural break ratios in large panels.

Importantly, this new method is robust to structural changes in the volatility of factors, uncomplicated to implement for practitioners, and can easily be extended to multiple structural breaks. A Monte Carlo simulation confirms that the approach works well for estimating moderately large breaks under different data-generating processes, as well as compares favorably to an existing method ([Baltagi et al., 2017]) in moderately large samples. In an empirical study of disaggregated US inflation series, I find two structural breaks in the factor loadings around the 1973 oil price shock and the 2008 financial crisis, as well as one structural break in the volatility of the factors around January 1991.

In Chapter 3, *Impulse Responses Estimation under Unstable Local Projections*, we introduce local time variation into the local projections framework, in the sense that the instability in local projections will be detected with a probability smaller than 1 even in the limit, and propose an impulse responses estimation methodology under unstable local projections. Importantly, the local time variation is considered in both the coefficients and the variances, thus modeling and estimating changes both in structural shocks and in the transmission mechanism. The impulse responses estimation methodology builds upon [Muller and Petalas, 2010] path estimators in a multivariate system. The chapter contains ample Monte Carlo evidence illustrating that [Muller and Petalas, 2010] asymptotically WAR minimizing path estimators and WAP maximizing parameter stability test statistics perform well in the unstable local projections framework with flexible specifications. To illustrate the estimation methodology, we revisit the small quarterly time-varying SVAR model of the U.S. economy studied in [Primiceri, 2005].

Contents

Chapter 1

DETECTING DENSITY FORECAST BREAKDOWNS

1.1 Introduction

The emergence of density forecasting as an established method springs largely from its ability to fully characterize the uncertainty associated with a prediction in ways that point forecasting, its alternative, does not convey. Research such as [Adrian et al., 2019] demonstrates the additional information that density forecasting provides. Density forecasting has therefore become important to both macroeconomics and finance[1], with central banks putting great effort into generating density forecasts for key macroeconomic variables. The Survey of Professional Forecasters, conducted by the Federal Reserve Bank of Philadelphia, provides density forecasts for inflation, output growth, and the unemployment rate. Meanwhile, the Bank of England was among the first institutions to officially adopt density forecasts, published in the form of fan charts.

The finding that in-sample fit is not indicative of out-of-sample forecasting performance is widespread in economics and finance; [Rossi, 2013] provides a review. Considering the increasing importance of reliable den-

[1] See [Tay and Wallis, 2000] for a survey.

sity forecasting, I propose a new method for evaluating density forecast models by developing a formal test to answer the following question: *How to formally assess whether the in-sample fit of a predictive density is indicative of its out-of-sample forecasting ability?* Note that the goal of this paper is distinct from a predictive density evaluation, which examines the accuracy or the correct specification of predictive densities (e.g., [Corradi and Swanson, 2006] and [Rossi and Sekhposyan, 2013]). Rather, I focus on whether the in-sample fit is a reliable indicator of the out-of-sample density forecasting ability, that is, whether future density forecasting performance is in line with the anticipated performance based on past information. [Clements and Hendry, 1998] explored this question, describing a forecast failure as a deterioration in forecast performance relative to anticipated performance, while [Giacomini and Rossi, 2009a] formally implemented the concept in their point forecast breakdown test. I generalize the forecast breakdown concept to density forecasting and develop a formal test to capture density forecast breakdowns. I formalize the definition of a density forecast breakdown (DFB) as a situation in which the out-of-sample density forecast performance, judged by an evaluation function, is significantly worse than its anticipated performance. Further, a scoring surprise is defined as the difference between out-of-sample density forecast performance and average in-sample performance. Thus, the DFB test is built on the idea that, in the absence of a DFB, the expectation of the average scoring surprise should be zero. The DFB test is formulated as a problem of inference about expectations of the scoring surprise, and is constructed based on the conditional density forecast models, which nest the unconditional density forecast models regarding the choice of conditioning set. In particular, applying the DFB test to conditional density forecast models based on current economic conditions helps to determine whether a density forecast model adapts well to changes in the economy. This framework applies to both conditional and unconditional densities, which can be informative for different research objectives.

This paper's primary theoretical contribution is to propose the DFB test for evaluating density forecast models and establish theories for the asymptotic behavior of the test statistics. "Model" is understood in a wide

sense, including both parametric and non-parametric models, and I establish theories for both settings. In the parametric case, the test allows for model misspecification and takes into account parameter estimation uncertainty[2] under both hypotheses. In the non-parametric case, I provide conditions under which estimation uncertainty is asymptotically irrelevant.

The DFB test is valid under general assumptions. It permits parametric estimation procedures, including ordinary least squares (OLS), generalized method of moments (GMM), and (quasi-) maximum likelihood (QML), as well as non-parametric density estimation, such as kernel density estimation (KDE). It also considers different algorithms for rolling and recursive window schemes for density estimation and forecasting. Most importantly, it permits a wide range of evaluation functions, or scoring rules. I propose the following three evaluation functions. The first is the logarithmic score (LS; e.g., [Amisano and Giacomini, 2007]), which is closely related to the Kullback Leibler Information Criterion (KLIC; e.g., [White, 1982], [Vuong, 1989], [Fernandez-Villaverde and Rubio-Ramirez, 2004]), and rewards density forecasts that assign a high probability to events that actually occurred. Here, evaluating density forecast performance amounts to understanding whether the realized values are, indeed, low-probability events. The second suggested evaluation function is the continuous ranked probability score (CRPS), analogous to the distributional generalization of mean squared error (DMSE) proposed in [Corradi and Swanson, 2006], which, compared with the LS, is sensitive to distance. That is, no credit is given for assigning a high probability to the value near, but not identical to, the event that actually occurred. The third suggested evaluation function is a confidence interval version of the DMSE, proposed in [Corradi and Swanson, 2006]. The advantage of this evaluation function, in contrast to the LS and CRPS, is its focus on the specific regions of the distribution, rather than the overall performance of the whole distribution.

Extending the DFB test to a more robust version in a context of instability increases its practical use. Indeed, as originally formulated, the

[2]Here, I adopt a similar framework to that developed by [West, 1996].

test is inadequate for detecting DFBs in an environment characterized by instability (see [Giacomini and Rossi, 2010]). This is because the out-of-sample density forecast performance may vary over time relative to the in-sample fit. Thus, a fluctuation DFB test is proposed to analyze scoring surprises over the entire time path, which circumvents the problem of useful information potentially being lost when averaging the global scoring surprises. Making the DFB test robust to the choice of the estimation window size is also important, since using arbitrary window sizes may lead to spurious empirical results in practice; see [Inoue and Rossi, 2012].

From an empirical perspective, this paper investigates the presence of DFBs in the density forecasting of real gross domestic product (GDP) growth and daily Standard and Poor's (S&P) 500 returns in the US. The first application investigates the relationship between the in-sample fit and the out-of-sample density forecasting ability of the model adopted by [Adrian et al., 2019]. This model forecasts the conditional distribution of future real GDP growth using a function of current financial and economic conditions, while the error term is modeled with a skewed t-distribution. The density forecasts of one-quarter-ahead and one-year-ahead real GDP growth are constructed by applying the quantile regressions of [Koenker and Bassett, 1978] and fitting the skewed t-distribution developed by [Azzalini and Capitanio, 2003] to the empirical conditional quantile function, using data from the first quarter of 1971 (1971Q1) to the fourth quarter of 2017 (2017Q4) with both rolling and recursive window schemes. The main finding is that the model's overall in-sample fit is indicative of the out-of-sample density forecasting ability, but DFBs occur sporadically in specific regions of the conditional distribution. In particular, the model's in-sample fit is not a reliable indicator of the out-of-sample density forecasting ability for the left tails (i.e., the lower quantile of the conditional distribution) of either the one-quarter-ahead or the one-year-ahead density forecasts. Meanwhile, the model's density forecasting ability of the right tails (i.e., the upper quantile of the conditional distribution) is consistent with in-sample fit. Furthermore, I find that excluding the current financial condition from the model results in DFBs occurring more often and in more regions of the conditional distribution. Thus, including financial

conditions in the density forecast model of real GDP growth helps narrow the gap between the in-sample fit and the out-of-sample density forecasting ability.

These results align with [Adrian et al., 2019], who find a correlation between financial conditions and higher moments of real GDP growth. They assert that the estimated lower quantiles of the future GDP growth exhibit strong variation as a function of current financial conditions, while the upper quantiles are stable over time. My findings further support their conclusions in that asymmetry exists not only in the estimated distribution but also in the out-of-sample density forecasting ability relative to the in-sample fit. The model is therefore very likely to experience DFBs in the region of the distribution exhibiting strong variation, and the in-sample fit is more likely to provide reliable guidance for the out-of-sample density forecasting ability in the relatively stable region. Moreover, I find that even if the model includes the current financial condition as a predictor, the in-sample fit still provides unreliable guidance for the out-of-sample density forecasting ability for lower quantiles since the financial crisis, and that misspecification of the out-of-sample predictive densities may cause DFBs, even though the in-sample density estimators are correctly specified and have good fit. The DFB test therefore implies that the in-sample fit of the model adopted by [Adrian et al., 2019] provides reliable guidance for researchers focusing on the overall performance of out-of-sample predictive densities. However, if the focus is on specific regions of the distribution, for example the 25%-50% quantile, care should be taken in choosing predictors, especially during the financial crisis.

Another application of the DFB test is to evaluate the generalized autoregressive conditional heteroskedasticity (GARCH) model's ability to construct the density forecasts of daily S&P 500 returns. Daily S&P 500 returns, ranging from Janury 3, 1981 to December 31, 2018, are fitted with GARCH(1,1) and GARCH-t(1,1) models with both rolling and recursive window schemes, and the one-day-ahead density forecasts are constructed accordingly. The empirical results show that, DFBs occur in the one-day-ahead density forecasts of S&P 500 returns of both the GARCH(1,1) and GARCH-t(1,1) models. This means that the overall

in-sample fit of the GARCH(1,1) and GARCH-t(1,1) is not indicative of their out-of-sample density forecasting abilities.

The proposed test is closely related to the work of [Giacomini and Rossi, 2009a], who propose a theoretical framework to capture a point forecast breakdown. However, I introduce several novel aspects. First, evaluating predictive densities differs from evaluating point forecasts. The former considers distinct evaluation functions, focusing on either the overall performance or the specific regions of the distributions. In contrast, evaluating point forecasts does not need to consider "specific regions". Second, parameter estimation uncertainty is much more complicated in density estimation. Taking the simple linear model with normal errors as an example, constructing density forecasts requires both coefficient and variance estimators, while constructing point forecasts does not require the latter. Including higher moment estimators makes study of asymptotic distribution of the test statistic more complicated. Additionally, when evaluating point forecasts, the parameter estimation uncertainty is asymptotically irrelevant in the common situation where the loss function used for estimation is the same as that used for evaluation. However, in most density forecast evaluations, parameter estimation uncertainty component is non-negligible. Finally, the proposed test can be applied not only to parametric density forecast models, but also to non-parametric models in which correcting estimation uncertainty is not a technically feasible option. I consequently provide conditions under which estimation uncertainty is asymptotically irrelevant.

This paper differs from the specification testing literature (e.g., [Diebold et al., 1998], [Bai, 2003b], [Hong and Li, 2004] and [Corradi and Swanson, 2006] for testing the correct specification of parametric in-sample density estimators; [Rossi and Sekhposyan, 2017] for testing the correct specification of out-of-sample predictive densities; [Berkowitz, 2001], [Corradi and Swanson, 2006], [Hong et al., 2007] and [Knuppel, 2015] for additional approaches for assessing the correct calibration of predictive densities) in its focus on the consistency between future density forecasting performance and the expectations based on past performance, which allows for model misspecification and the consideration of situa-

tions closer to reality, as opposed to concentrating exclusively on correct specification.

The break testing literature has largely been concerned with distributional change testing (e.g., [Inoue, 2001]) and structural break testing (e.g., [Andrews, 1993], [Bai and Perron, 1998], [Hansen, 2000] and [Elliott and Muller, 2006]). In contrast, I focus on the stability of forecasting performance instead of the stability of parameters or distributions; [Giacomini and Rossi, 2009a] and [Rossi, 2013] provide theoretical examples showing that structural breaks in the parameters are neither necessary nor sufficient to generate instabilities in forecasting performance. The test proposed in this study thus captures all the various changes that affect density forecasting performance, including parameter changes, distributional changes and other changes. Furthermore, this test can be applied to non-parametric predictive densities, which do not require assuming parametric forms. I use similar evaluation functions to those employed in the predictive density comparison literature (e.g., [Vuong, 1989], [Amisano and Giacomini, 2007], and Corradi and Swanson (2005a, 2006b)). However, I focus on model evaluation by comparing the performance of predictive densities with the anticipated performance, rather than by comparing the performances of various predictive density models.

A wealth of empirical studies have explored whether in-sample fit provides guidance for out-of-sample forecasting ability in predicting stock returns (e.g., [Ang and Bekaert, 2006])[3], exchange rates (e.g., [Meese and Rogoff, 1983], [Sarno and Valente, 2009], [Rossi and Sekhposyan, 2011]), and output growth (e.g., [Swanson and White, 1997], [Swanson, 1998], [Giacomini and Rossi, 2009b]). I investigate this relationship in the context of forecasting a target variable's entire density/distribution rather than its mean value. Furthermore, a large literature has documented GDP volatility change, especially during the Great Moderation

[3]Other literature in this area includes Campbell (1987), Campbell and Shiller (1988), Bekaert and Hodrick (1992), Fama and French (1988), Perez-Quiros and Timmermann (2000), and Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Cooper et al. (2005), Marquering andVerbeek (2004), Sullivan et al. (1999), Paye and Timmermann (2006), Goyal and Welch (2003).

(e.g., [Kim and Nelson, 1999], [McConnell and Perez-Quiros, 2000], [Blanchard and Simon, 2001], [Bernanke, 2004] and [Giannone et al., 2008]). Here I focus on the whole density/distribution rather than the second moment of real GDP growth and investigate the relationship between in-sample fit and out-of-sample density forecasting ability.

The remainder of this paper is organized as follows. Section 1.2 discusses the framework of the DFB test in parametric density forecast models. Section 1.3 extends the method to the non-parametric case. Section 2.5 introduces the robust and adapted versions of the DFB test. Section 3.3 gives the results of the Monte Carlo simulation study. Section 3.4 applies the DFB test to evaluate the density forecast models of real GDP growth and daily S&P 500 returns. Section 3.5 concludes.

1.2 Detecting Density Forecast Breakdowns for Parametric Models

This section introduces the theoretical framework of the density forecast breakdown test for parametric models.

1.2.1 Environment

Consider a stochastic process $\{Z_t : \Omega \to \mathbb{R}^{m+1}\}_{t=1}^{T}$ defined on a complete probability space $(\Omega; \mathbb{F}; P)$, the observed vector Z_t is partitioned as $Z_t = (y_t; \mathcal{X}_t')'$, where $y_t : \Omega \to \mathbb{R}$ is the variable of interest and $\mathbb{F}_t = \sigma(Z_1'; \ldots; Z_t')'$ is the true information set available at time t. The true but unknown h-step-ahead ($1 \leq h < \infty$) predictive density for the scalar[4] variable y_{t+h} conditional on $\mathcal{X}_t$ is denoted by $\phi_0(\cdot) = \phi(\cdot|\mathcal{X}_t, \theta_0)$, where θ_0 is the true parameter vector. The researcher may observe a subset of the true information set and characterize the unknown h-step-ahead conditional predictive density for the scalar variable y_{t+h} using the model $\phi(\cdot|\cdot, \cdot)$ based on the observed information set $\mathcal{F}_t$ ($\mathcal{F}_t \subseteq \mathbb{F}_t$). In this case,

[4]The multivariate case would be similar.

the model is misspecified unless $\mathcal{F}_t = \mathbb{F}_t$. Suppose X_t ($X_t \subseteq \mathcal{X}_t$) is a vector of k predictors included in the density forecast model, the pseudo-true predictive density of the scalar variable y_{t+h} conditional on X_t is denoted by $\phi^*_{t+h}(\cdot) = \phi(\cdot|X_t, \theta^*)$, where θ^* corresponds to the probability limits of the estimated parameters.[5] This possibly misspecified conditional density forecast model is the focus of this paper.

To construct the h-step-ahead predictive density of the variable of interest y_{t+h}, the available sample of size T is divided into an in-sample portion of size R and an out-of-sample portion of size P such that $R + P - 1 + h = T$. The data that comprise the in-sample window are based on the estimation schemes. The following two estimation schemes should be considered: (i) a rolling estimation scheme, in which the in-sample window at time t contains observations indexed $t - R + 1, \ldots, t$, and (ii) a recursive estimation scheme in which the in-sample window includes observations indexed $1, \ldots, t$.[6] At each period $t = R, \ldots, T - h$, the estimated parameter $\hat{\theta}_t$ is obtained based on the corresponding in-sample window data, and the predictive density function for the scalar variable y_{t+h} conditional on X_t is denoted as $\hat{\phi}^f_{t+h}(\cdot) \equiv \phi(\cdot|X_t, \hat{\theta}_t)$, where the superscript f refers to the "out-of-sample forecast." Similarly, the conditional density estimator for period j (j varying over the corresponding in-sample window) is denoted as $\hat{\phi}^e_j(\cdot) = \phi(\cdot|X_{j-h}, \hat{\theta}_t)$, where the superscript e refers to "estimated in-sample." Figure 1.1 gives an illustration of the out-of-sample conditional predictive density $\hat{\phi}^f_{t+h}(\cdot)$, as well as the in-sample conditional density estimators $\hat{\phi}^e_j(\cdot), j = t - R + 1, \ldots, t$, standing at period t using a rolling window estimation scheme.

The conditional density forecast model discussed above nests the unconditional forecast model based on the choice of conditioning set. For

[5] In this context, we are concerned with direct multi-step forecasting, where the predictors are lagged h periods; that is, the model specifies the relationship between y_{t+h} and X_t. The indirect approach, not considered here, is to compute iterated h-step-ahead density forecasts, i.e., first predict y_{t+1} using observations up to time t, and then use this predicted value in order to predict y_{t+2}, and so on.

[6] Although not widely used in practice, the fixed estimation scheme, in which the in-sample window at time t contains observations indexed $1, \ldots, R$, can also be nested in this context.

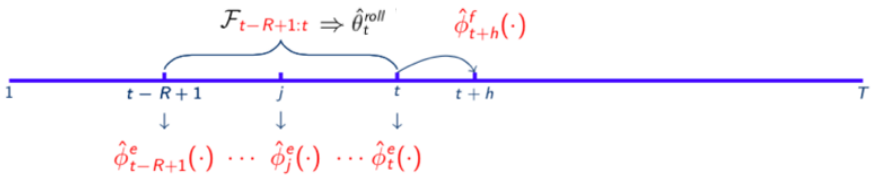


Figure 1.1: Illustration of estimation and forecasting at t

example, an unconditional predictive density for the scalar variable y_{t+h} constructed at period t would be $\hat{\phi}^f_{t+h}(\cdot) \equiv \phi(\cdot|\hat{\theta}_t)$.[7] For simplicity, the notation used here is that of the conditional case. I extend the setting to the non-parametric case in Section 1.3.

1.2.2 Evaluation functions

Evaluation functions, or scoring rules, are loss functions whose arguments are the density forecast and the actual outcome of the variable (e.g., [Winkler, 1967], [Diebold and Lopez, 1996], and [Lopez, 2001]). Scoring rules provide summary measurements for evaluating probabilistic forecasts by assigning a numerical score based on the predictive distribution and on the event or value that materializes (for a review and theory of proper scoring rules, see [Gneiting and Raftery, 2007]). I propose three evaluation functions in this paper: the logarithmic score (LS), the continuous ranked probability score (CRPS), and a confidence interval version of the distributional generalization of the mean squared error (CI).

Logarithmic score

The logarithmic scoring rule proposed by [Good, 1952] is defined as $S(f, Y) = \log f(Y)$, where Y is the observed value of the variable and

[7] Similarly, the density estimator for period j, with j varying over the corresponding in-sample window, would be $\hat{\phi}^e(\cdot) = \phi^e(\cdot|\hat{\theta}_t)$. There is no subscript j since there is no conditioning variable.

$f(.)$ is the density function (see also [Amisano and Giacomini, 2007]). This LS is closely related to the Kullback-Leibler Information Criterion (KLIC), since the divergence function associated with the expected LS is the classical Kullback-Leibler divergence. The LS intuitively rewards a density forecast that assigns a high probability to events that have actually occurred. The LS is also mathematically convenient, since it is the only scoring rule that is solely a function of the density's value at the variable's realization.

In the present context, the LS denotes the score for the h-step-ahead density forecast for the scalar variable y_{t+h} constructed at period t by

$$S_{t+h}(\hat{\theta}_t) \equiv S(\hat{\phi}^f_{t+h}(\cdot), y_{t+h}) = \log \phi(y_{t+h}|X_t, \hat{\theta}_t)$$

and also denotes the score for period j (with j varying over the in-sample window) as

$$S_j(\hat{\theta}_t) \equiv S(\hat{\phi}^e_j(\cdot), y_j) = \log \phi(y_j|X_{j-h}, \hat{\theta}_t)$$

Continuous ranked probability score

The CRPS is defined in terms of a cumulative distribution function (CDF) as $CRPS(F, x) = -\int_{-\infty}^{\infty} (F(y) - 1(y \geq x))^2 \, dy$, where x is the observed value of the variable and $F(\cdot)$ is the cumulative distribution function. The CRPS is closely related to the distributional generalization of the mean square error (DMSE) introduced by [Corradi and Swanson, 2006] as a measurement of distributional accuracy,[8] and the term

[8] Suppose the true distribution is $F_0(\cdot|\theta_0)$. [Corradi and Swanson, 2006] define the mean square error associated with $F_1(\cdot|\theta_1^*)$, $\theta_1^* = \theta_0$ if correctly specified, in terms of an average over U of $\mathrm{E}\left[(F_1(u|\theta_1^*) - F_0(u|\theta_0))^2\right]$, where $u \in U$. [Corradi and Swanson, 2006] adopt this DMSE to compare multiple (mis)specified predictive density models. In particular, Model 1 outperforms Model 2 if

$$\int_U \mathrm{E}[(F_1(u|\theta_1^*) - F_0(u|\theta_0))^2 - (F_2(u|\theta_2^*) - F_0(u|\theta_0))^2]\phi(u)du < 0$$

where $\int_U \phi(u)du = 1$ and $\phi(u) \geq 0$. A confidence interval version can be written as

$$\mathrm{E}[(F_1(\overline{u}|\theta_1^*) - F_1(\underline{u}|\theta_1^*)) - (F_0(\overline{u}|\theta_0) - F_0(\underline{u}|\theta_0)))^2 - (F_2(\overline{u}|\theta_2^*) - F_2(\underline{u}|\theta_2^*)) - (F_0(\overline{u}|\theta_0) - F_0(\underline{u}|\theta_0)))^2] < 0$$

$F(y) - 1(y \geq x)$ can be interpreted as the "error" term[9] for the cumulative distribution functions.

In the present context, the CRPS denotes the score for the h-step-ahead density forecast for the scalar variable y_{t+h} constructed at period t by

$$S_{t+h}(\hat{\theta}_t) \equiv S(\hat{\Phi}^f_{t+h}(\cdot), y_{t+h}) = -\int_U \Big(1(y_{t+h} \leq u) - \Phi(u|X_t, \hat{\theta}_t)\Big)^2 du$$

and also denotes the score for period j (with j varying over the in-sample window) as

$$S_j(\hat{\theta}_t) \equiv S(\hat{\Phi}^e_j(\cdot), y_j) = -\int_U \Big(1(y_j \leq u) - \Phi(u|X_{j-h}, \hat{\theta}_t)\Big)^2 du$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of $\phi(\cdot)$.

Compared with the LS, which is defined in terms of a probability density function (PDF), the CRPS (or DMSE) is defined directly in terms of a cumulative distribution function (CDF). Considering CRPS (or DMSE) complements the LS in a number of ways. First, it helps circumvent the problem that restrictions to predictive densities are often impractical. For example, not all distributions have PDFs, and PDFs may involve a point mass. Thus, a scoring rule defined directly in terms of a CDF can be more compelling. Furthermore, CRPS (or DMSE) is also sensitive to distance; no credit is given for assigning a high probability to the value close but not identical to the event that actually occurred. In this sense, the CRPS (or DMSE) is a more accurate measurement than the LS.

A confidence interval version of measurement

A CI is more natural to use when measuring performance in specific regions. In the present context, the CI denotes the score for the h-step-ahead density forecast for the scalar variable y_{t+h} constructed at period t by

$$\begin{aligned} S_{t+h}(\hat{\theta}_t) &\equiv S(\hat{\Phi}^f_{t+h}(\cdot), y_{t+h}) \\ &= -\Big(1(\underline{u} \leq y_{t+h} \leq \overline{u}) - (\Phi(\overline{u}|X_t, \hat{\theta}_t) - \Phi(\underline{u}|X_t, \hat{\theta}_t))\Big)^2 \end{aligned}$$

[9] $\mathrm{E}\left[1(y \geq x)\right] = Pr(x \leq y) = F_0(y)$, where $F_0(y)$ is the true cumulative distribution function. See [Corradi and Swanson, 2005] for a similar explanation.

and also denotes the score for period j (with j varying over the in-sample window) as

$$\begin{aligned} S_j(\hat{\theta}_t) &\equiv S(\hat{\Phi}_j^e(\cdot), y_j) \\ &= -\left(1(\underline{u} \leq y_j \leq \overline{u}) - (\Phi(\overline{u}|X_{j-h}, \hat{\theta}_t) - \Phi(\underline{u}|X_{j-h}, \hat{\theta}_t))\right)^2 \end{aligned}$$

Compared with the LS and CRPS, which measure the overall performance of the density function or distribution function, CI measures the performance of the distribution function over a specific region. Thus, tests involving a CI complement those that consider the LS and CRPS by capturing DFBs in specific regions, whereas performances of different regions may offset each other in the latter tests. An alternative method for considering the performance of a specific region is adding a weighting function so as to focus on certain regions of U that are of interest (see [Corradi and Swanson, 2006] and [Amisano and Giacomini, 2007]).[10]

It is worth noting that the evaluation function $S(\cdot)$ can take a general form, and other types of evaluation functions can be considered, including scoring rules for quantile (e.g., value-at-risk) and interval forecasts (see [Gneiting and Raftery, 2007] for interval scores). Furthermore, the probability integral transform (PIT) may also be employed as an evaluation function[11], although in this context, the CRPS (or DMSE) is a better

[10]For example, the adjusted score for the h-step-ahead density forecast for the scalar variable y_{t+h} constructed at period t by

$$S_{t+h}(\hat{\theta}_t) \equiv S(\hat{\phi}_{t+h}^f(\cdot), y_{t+h}) = -\int_U \left(1(y_{t+h} \leq u) - \Phi(u|X_t, \hat{\theta}_t)\right)^2 g(u)du$$

and denotes the score for period j, with j varying over the in-sample window, by

$$S_j(\hat{\theta}_t) \equiv S(\hat{\phi}_j^e(\cdot), y_j) = -\int_U \left(1(y_j \leq u) - \Phi(u|X_{j-h}, \hat{\theta}_t)\right)^2 g(u)du$$

where $\int_U g(u)du = 1$ and $g(u) \geq 0$. By setting the value of $g(u)$, one can measure the performances using varying weights given for different regions.

[11]In this case, $S_{t+h}(\hat{\theta}_t) \equiv S(\hat{\phi}_{t+h}^f(\cdot), y_{t+h}) = \int_{-\infty}^{y_{t+h}} \phi(y|\mathcal{F}_t, \hat{\theta}_t)dy$, and $S_j(\hat{\theta}_t) \equiv S(\hat{\phi}_j^e(\cdot), y_j) = \int_{-\infty}^{y_j} \phi(y|\mathcal{F}_t, \hat{\theta}_t)dy$.

choice. Although both the PIT and CRPS (or DMSE) are CDF-based measurements, accuracy measurements based on the PIT may cause a problem in estimation uncertainty correction, which requires the derivative of the measurements. Moreover, the test herein cannot make use of the PIT's properties of being uniform, independent and identically distributed if the density forecast is correctly specified as I focus on consistency between the future density forecasting performance and the in-sample fit, rather than the correct specification. In addition, evaluating the out-of-sample performance relative to the in-sample fit is similar to model comparisons that commonly use the LS and CRPS (or DMSE) to compare relative performance (e.g., [Corradi and Swanson, 2005] and [Amisano and Giacomini, 2007]).

For simplicity, $S_{t+h}(\hat{\theta}_t)$ and $S_j(\hat{\theta}_t)$ are hereon denoted as S_{t+h} and S_j, and a superscript $*$ is used to refer to counterparts evaluated by the pseudo-true parameter θ^*.

1.2.3 Test

In this paper, DFBs are defined as situations in which the out-of-sample density forecast performance, judged by an evaluation function, is significantly worse than its anticipated performance as measured by average in-sample performance. I further formalize this idea by defining the "scoring surprise" at time $t+h$ as the difference between the out-of-sample density forecast performance (the score for the density forecast for y_{t+h}) and the averaged in-sample performances (the average score over the in-sample window). This definition is as follows:

$$SS_{t+h} = S_{t+h} - \overline{S}_t \tag{1.1}$$

where $\overline{S}_t$ is the average in-sample score computed over the in-sample window implied by the estimation schemes: $\overline{S}_t = \frac{1}{R}\sum_{j=t-R+1}^{t} S_j$ for the rolling scheme and $\overline{S}_t = \frac{1}{t}\sum_{j=1}^{t} S_j$ for the recursice scheme.[12] A positive scoring surprise indicates better out-of-sample density forecast

[12] $\overline{S}_t = \frac{1}{R}\sum_{j=1}^{R} S_j$ for the fixed scheme.

performance relative to the in-sample fit while, conversely, a negative scoring surprise indicates worse relative out-of-sample density forecast performance.

Consider the out-of-sample mean of the scoring surprise:

$$\overline{SS}_{R,P} \equiv \frac{1}{P} \sum_{t=R}^{T-h} SS_{t+h} \tag{1.2}$$

The DFB test is based on the concept that, in the absence of a DFB, this sum should be close to zero. More formally, the expectation of the averaged scoring surprise evaluated by the pseudo-true model should be zero. This implies a no-DFB null hypothesis constructed as follows:

$$H_0 : E\left[\frac{1}{P} \sum_{t=R}^{T-h} SS^*_{t+h}\right] = 0 \tag{1.3}$$

against the alternative that

$$H_a : E\left[\frac{1}{P} \sum_{t=R}^{T-h} SS^*_{t+h}\right] < 0 \tag{1.4}$$

where SS^*_{t+h} is the counterpart of SS_{t+h} evaluated with the pseudo-true parameter θ^*. The DFB test is a one-sided test, since a breakdown refers to worse out-of-sample density forecast performance relative to the in-sample fit. Similarly, if one is interested in testing whether there is a density forecast improvement (DFI), then the alternative should be $H_a : E\left[\frac{1}{P} \sum_{t=R}^{T-h} SS^*_{t+h}\right] > 0$. If one is interested in testing whether there is a density forecast break, either a breakdown or an improvement, then a two-sided test can be constructed with the alternative $H_a : E\left[\frac{1}{P} \sum_{t=R}^{T-h} SS^*_{t+h}\right] \neq 0$.

The DFB test statistic is

$$t_{R,P,h} = \frac{P^{1/2}\overline{SS}_{R,P}}{\hat{\sigma}_{R,P}} \tag{1.5}$$

where $\hat{\sigma}^2_{R,P}$ is the asymptotic variance estimator of $P^{1/2}\overline{SS}_{R,P}$. A level α test rejects the null hypothesis whenever $t_{R,P,h} > z_\alpha$, where z_α is the $1-\alpha$-th quantile of a standard normal distribution. The asymptotic distribution for the DFB test is provided by Theorem 1.

1.2.4 Assumptions

The following assumptions are made for the asymptotic distribution of the DFB test.

Assumption P1. *$\{Z_t\}$ is a mixing sequence with α of size $-r/(r-2)$, $r > 2$.*

Assumption P2. *(a) S_t is measurable and twice continuously differentiable with respect to θ; (b) Under H_0, in the neighbourhood $\mathcal{N}$ of θ^*, there exists a constant $D < \infty$ such that for all t, $sup_{\theta\in\mathcal{N}}|\frac{\partial^2 S_t}{\partial\theta\partial\theta'}| < m_t$, for a measurable m_t such that $E[m_t] < D$.*

Assumption P3. *Under H_0, $sup_{t\geq R}\left\|\hat{\theta}_t - \theta^* - B_t^* H_t^*\right\| \to^{a.s} 0$, where $\hat{\theta}$ is $k\times 1$. B_t^* is a $k\times q$ matrix of column k such that $sup_{t\geq 1}|B_t^*| < \infty$. $H_t^* = 1/R\sum_{s=1}^{R} h_s^*$(fixed scheme); $H_t^* = 1/R\sum_{s=t-R+1}^{t} h_s^*$(rolling scheme); $H_t^* = 1/t\sum_{s=1}^{t} h_s^*$(recursive scheme); for a $q\times 1$ orthogonality condition h_s^* such that $\mathrm{E}[h_s^*] = 0$.*

Assumption P4. *$sup_{t\geq 1}\left\|[S_t^*, \partial S_t^*/\partial\theta, h_t^{*\prime}]'\right\|^{2r} < \infty$, where $\partial S_t^*/\partial\theta$ is $1\times k$.*

Assumption P5. *$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\frac{\partial S_t^*}{\partial\theta}\right] < \infty$ for all T.*

Assumption P6. *$\mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T} S_t^*\right) > 0$ for all T sufficiently large.*

Assumption P7. *$R, P \to \infty$, $\frac{P}{R} \to \pi$, $0 \leq \pi < \infty$.*

Assumption P1 restricts the dependence and allows heterogeneity in the data. This assumption follows Assumption A1 made by [Giacomini

and Rossi, 2009a]. Assumption P2 is adapted from Assumption A1 made by [West, 1996] and Assumption A2 made by [Giacomini and Rossi, 2009a], implying that (i) $\phi(y|X_t,\theta) \neq 0$ for all y, and (ii) $\phi(y|X_t,\theta)$ is twice continuously differentiable w.r.t θ.[13] Assumption P3 follows Assumption A2 made by [West, 1996] and Assumption A3 made by [Giacomini and Rossi, 2009a]). It permits a number of estimation procedures for the model's parameters, including ordinary least squares (OLS), generalized method of moments (GMM), and (quasi-) maximum likelihood (QML). The important difference between Assumption P3 and the assumptions made by [West, 1996] and [Giacomini and Rossi, 2009a] is that the point forecast only requires coefficient parameters, while the density forecast also requires the variance parameter. For example, for the OLS estimation of the parameters in the linear model $Y_s = X_s'\beta^* + \epsilon_s, s = 1,\ldots,t$, the estimation uncertainty adopted in point forecast is $\left\|\hat{\beta}_t - \beta^* - B_t^* H_t^*\right\| \to^{a.s} 0$ with $B_t^* = (\mathrm{E}(\frac{1}{t}\sum_{s=1}^t X_s X_s'))^{-1}$ and $h_s^* = X_s\epsilon_s$; while the density forecast requires the estimation uncertainty of the parameter vector $\theta = [\beta'\sigma^2]'$. In the density forecast,

$$\begin{aligned}\hat{\theta}_t - \theta^* &= \begin{bmatrix}\hat{\beta}_t - \beta^* \\ \hat{\sigma}_t^2 - \sigma^{*2}\end{bmatrix} \\ &= \begin{bmatrix}(\frac{1}{t}\sum_{s=1}^t X_s X_s')^{-1} & 0 \\ -\frac{1}{t}\sum_{s=1}^t \epsilon_s X_s'(\frac{1}{t}\sum_{s=1}^t X_s X_s')^{-1} & 1\end{bmatrix}\begin{bmatrix}\frac{1}{t}\sum_{s=1}^t X_s\epsilon_s' \\ \frac{1}{t}\sum_{s=1}^t \epsilon_s\epsilon_s' - \sigma^{*2}\end{bmatrix} = \overline{B}_t H_t^*\end{aligned}$$

and requires that $B_t^* = \begin{bmatrix}(E[\frac{1}{t}\sum_{s=1}^t X_s X_s'])^{-1} & 0 \\ 0 & 1\end{bmatrix}$ and $h_s^* = \begin{bmatrix}X_s\epsilon_s' \\ \epsilon_s\epsilon_s' - \sigma^{*2}\end{bmatrix}$.
[14]

Assumption P5 is a relaxed version of Assumption A5 made by [Giacomini and Rossi, 2009a]. In the point forecast, estimation uncertainty

[13] The reason being that, for a parametric model, the second derivative of $S_t = \log\phi(y_t|X_t,\theta)$ with respect to θ is $\frac{\partial^2 S_t}{\partial\theta\partial\theta'} = \frac{\frac{\partial^2\phi(y_t|X_t,\theta)}{\partial\theta\partial\theta'}\cdot\phi(y_t|X_t,\theta) - (\partial\phi(y_t|X_t,\theta)/\partial\theta)^2}{\phi(y_t|X_t,\theta)^2}$.

[14] $\overline{B}_t = \begin{bmatrix}(\frac{1}{t}\sum_{s=1}^t X_s X_s')^{-1} & 0 \\ -\frac{1}{t}\sum_{s=1}^t \epsilon_s X_s'(\frac{1}{t}\sum_{s=1}^t X_s X_s')^{-1} & 1\end{bmatrix} \to \begin{bmatrix}(E[\frac{1}{t}\sum_{s=1}^t X_s X_s'])^{-1} & 0 \\ 0 & 1\end{bmatrix} = B_t^*$

can be irrelevant if the loss function is the same for the estimation and evaluation.[15] However, in the density forecast, this can only hold when the LS and QMLE are adopted simultaneously, and the derivatives of the evaluation functions are the derivatives of the log-likelihood functions (see Corollary 1). For all other cases, estimation uncertainty cannot be omitted. Assumption P5 ensures that estimation uncertainty converges. Assumption P7 follows Assumption A7 made by [Giacomini and Rossi, 2009a]. It ensures that the test statistic has an asymptotically normal distribution when the in-sample and out-of-sample sizes reach infinity at the same rate or when the in-sample size grows faster than the out-of-sample size. In particular, when $\pi = 0$ (i.e., when the in-sample size grows faster than the out-of-sample size), the estimation uncertainty component can be asymptotically irrelevant (see Theorem 1).

1.2.5 Main results

In what follows, I show how to construct a valid asymptotic variance estimator for the DFB test statistic. The averaged scoring surprise is rewritten as a weighted average of in-sample and out-of-sample scoring with weights dependent on (R, P) and the estimation schemes. Additionally, the averaged estimation uncertainty component is rewritten as a weighted average of in-sample and out-of-sample $h_t(\theta^*)$ as defined in Assumption P3, with weights that are dependent on (R, P) and the estimation scheme and that are computed using B_t^* as defined in Assumption P3. Therefore, the asymptotic variance estimator is simply a (rescaled) heteroskedasticity and autocorrelation consistent (HAC) estimator of the variance and covariance of these two weighted averages. A HAC estimator is used for the asymptotic variance due to the possible presence of serial correlation in both the scores and the estimation uncertainty.

For simplification, the algorithm for the fixed scheme is given here in detail, and the algorithms for the rolling and recursive schemes are

[15] If the loss function is the same for the estimation and evaluation, then $\mathrm{E}[\frac{\partial L_t(\beta^*)}{\partial \beta}] = 0$. To this regard, [Giacomini and Rossi, 2009a] demonstrate that estimation uncertainty is irrelevant if this assumption holds.

provided in Appendix 1.8.2.

Algorithm 1 *Construct the following: (1) a* $1 \times T$ *vector* S^* *of in-sample and out-of-sample scores, with element* $S_t^*, t = 1, \ldots, T$*, and its counterpart* S *with element* $S_t, t = 1, \ldots, T$*:*

$$S^* \equiv \left[\underbrace{S_1^*(\theta^*), \ldots, S_R^*(\theta^*)}_{R}, \underbrace{S_{R+1}(\theta^*), \ldots, S_{R+h-1}(\theta^*)}_{h-1}, \underbrace{S_{R+h}^*(\theta^*), \ldots, S_T^*(\theta^*)}_{P} \right]$$

$$S \equiv \left[\underbrace{S_1(\hat{\theta}_R), \ldots, S_R(\hat{\theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\theta}_R), \ldots, S_{R+h-1}(\hat{\theta}_R)}_{h-1}, \underbrace{S_{R+h}(\hat{\theta}_R), \ldots, S_T(\hat{\theta}_R)}_{P} \right]$$

and the corresponding vector $\tilde{S}$ *of demeaned scorings, where* $\tilde{S}_t \equiv S_t - T^{-1} \sum_{j=1}^{T} S_j$*; (2) a* $1 \times T$ *vector of weights, depending on the estimation scheme, with elements* $\omega_t^S, t = 1, \ldots, T$*:*

$$\omega^S \equiv \left[\underbrace{-\frac{P}{R}, \ldots, -\frac{P}{R}}_{R}, \underbrace{0, \ldots, 0}_{h-1}, \underbrace{1, \ldots, 1}_{P} \right]$$

so that

$$\sum_{t=1}^{T} \omega_t^S S_t^* = \sum_{t=R}^{T-h} S S_{t+h}^*$$

Algorithm 2 *Construct the following: (1) a* $(k+1) \times T$ *vector of in-sample and out-of-sample* h^**, as defined in Assumption P3, with element* $h_t^*, t = 1, \ldots, T$*, and its counterpart* h *with element* $h_t, t = 1, \ldots, T$*:*

$$h^* \equiv \left[\underbrace{h_1^*(\theta^*), \ldots, h_R^*(\theta^*)}_{R}, \underbrace{h_{R+1}^*(\theta^*), \ldots, h_{T-h}^*(\theta^*)}_{P-1}, \underbrace{O, \ldots, O}_{h} \right]$$

$$h \equiv \left[\underbrace{h_1(\hat{\theta}_R), \ldots, h_R(\hat{\theta}_R)}_{R}, \underbrace{h_{R+1}(\hat{\theta}_R), \ldots, h_{T-h}(\hat{\theta}_R)}_{P-1}, \underbrace{O, \ldots, O}_{h} \right]$$

(2) a $1 \times T$ *vector of weights* ω^{h*}*, depending on the estimation scheme, with elements* $\omega_t^{h*}, t = 1, \ldots, T$ *and its counterpart* ω^h *with elements* $\omega_t^h, t = 1, \ldots, T$*:*[16]

$$\omega^{h*} \equiv \left[\underbrace{-\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_R^*}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_R^*}{R}}_{R}, \underbrace{O, \ldots, O}_{T-R} \right] \qquad D_{t+h}^* = \frac{\partial SS_{t+h}^*(\theta^*)}{\partial \theta}$$

$$\omega^{h} \equiv \left[\underbrace{-\frac{\sum_{t=R}^{T-h} D_{t+h} B_R}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h} B_R}{R}}_{R}, \underbrace{O, \ldots, O}_{T-R} \right] \qquad D_{t+h} = \frac{\partial SS_{t+h}(\hat{\theta}_R)}{\partial \theta}$$

so that

$$\sum_{t=1}^{T} \omega_t^{h*} h_t^* = \sum_{t=R}^{T-h} D_{t+h}^* B_t^* H_t^*$$

The asymptotic variance estimator is then simply a (rescaled) heteroskedasticity and autocorrelation consistent (HAC) estimator of the variance and covariance of these two weighted averages.

$$\hat{\sigma}_{R,P}^2 = \frac{T}{P}(V_T^{SS} + V_T^{hh} + 2V_T^{Sh}) \tag{1.6}$$

with V_T^{SS},V_T^{hh},V_T^{Sh} defined as follows:

$$\begin{aligned}
V_T &= \begin{bmatrix} V_T^{SS} & V_T^{Sh} \\ V_T^{Sh} & V_T^{hh} \end{bmatrix} \\
V_T^{SS} &= \frac{1}{T}\sum_{t=1}^{T}(\omega_t^S \tilde{S}_t)^2 + \frac{2}{T}\sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} \omega_t^S \tilde{S}_t \omega_{t-j}^S \tilde{S}_{t-j} \\
V_T^{hh} &= \frac{1}{T}\sum_{t=1}^{T} \omega_t^h h_t h_t' \omega_t^{h'} + \frac{2}{T}\sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} (\omega_t^h h_t h_{t-j}' \omega_{t-j}^{h'} + \omega_{t-j}^h h_{t-j} h_t' \omega_t^{h'}) \\
V_T^{Sh} &= \frac{1}{T}\sum_{t=1}^{T} \omega_t^S \tilde{S}_t h_t' \omega_t^{h'} + \frac{2}{T}\sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} (\omega_t^S \tilde{S}_t h_{t-j}' \omega_{t-j}^{h'} + \omega_{t-j}^S \tilde{S}_{t-j} h_t' \omega_t^{h'}),
\end{aligned} \tag{1.7}$$

[16] B_t is the consistent estimate of B_t^*.

where bandwidth p_T and weights $v_{T,j}$ are appropriately chosen as per [Newey and West, 1987].

Theorem 1. ***(Generalization of Asymptotic Distribution of the DFB Test for Parametric Models)***
(i) Given Assumption P 1-7, if V_T defined in eq (1.7) is p.d., then under H_0 in (1.3), $t_{R,P,h} \xrightarrow{d} \mathcal{N}(0,1)$, where $t_{R,P,h}$ is defined in (1.5) *and $\hat{\sigma}^2_{R,P}$ defined in* (1.6)*.*
(ii) In the special case that $\pi = 0$ in Assumption P7, $\hat{\sigma}^2_{R,P} = \frac{T}{P} V_T^{SS}$ with V_T^{SS} defined in (1.7)*.*

When the LS and QMLE are adopted simultaneously, the evaluation function and the estimation objective function are the same, such that $\mathrm{E}\left[\frac{\partial S_t^*}{\partial \theta}\right] = \mathrm{E}\left[\frac{\partial \log \phi_t^*}{\partial \theta}\right] = 0$ for all t. Therefore, it is satisfied that $\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left[\frac{\partial S_t^*}{\partial \theta}\right] < \infty$ for all t, leading to a simple expression for $\hat{\sigma}^2_{R,P}$, where the estimation uncertainty is asymptotically irrelevant. The following corollary illustrates the situation in which the LS is adopted as the evaluation function and the QMLE.

Corollary 1. ***(Special case: variance estimator under the LS and QMLE)***
Given Assumption P 1-7, under H_0 in (1.3), $t_{R,P,h} \xrightarrow{d} \mathcal{N}(0,1)$, where $t_{R,P,h}$ is defined in (1.5) *and $\hat{\sigma}^2_{R,P} = \frac{T}{P} V_T^{SS}$ with V_T^{SS} defined in* (1.7)*.*

1.2.6 Causes of DFBs

It is of interest to investigate why the in-sample fit provides unreliable guidance for the out-of-sample density forecasting ability. To this end, Proposition 1 considers the expectation of the DFB test statistic's numerator so as to analyze the causes of the DFBs.

Proposition 1. ***Causes of density forecast breakdowns for parametric***

models

$$
\begin{aligned}
E\left[P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\phi})\right] &= \underbrace{E\left[P^{-1/2}\sum_{t=R}^{T-h}\left(S_{t+h}(\theta^*_{t+h}) - \overline{\sum}_j S_j(\theta^*_j)\right)\right]}_{\textit{Other Instabilities}} \\
&+ \underbrace{P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\frac{\partial S_{t+h}(\theta^*_{t+h})}{\partial\theta}\right]\left(\theta^*_t - \theta^*_{t+h}\right) - P^{-1/2}\sum_{t=R}^{T-h}\overline{\sum}_j \mathrm{E}\left[\frac{\partial S_j(\theta^*_j)}{\partial\theta}\right]\left(\theta^*_t - \theta^*_j\right)}_{\textit{Parameter Instability I}} \\
&+ \underbrace{\frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}\left\{\left(\theta^*_t - \theta^*_{t+h}\right)'\mathrm{E}\left[\frac{\partial^2 S_{t+h}(\bar{\theta}^*_{t+h})}{\partial\theta\partial\theta'}\right]\left(\theta^*_t - \theta^*_{t+h}\right) - \overline{\sum}_j\left(\theta^*_t - \theta^*_j\right)'\mathrm{E}\left[\frac{\partial^2 S_j(\bar{\theta}^*_j)}{\partial\theta\partial\theta'}\right]\left(\theta^*_t - \theta^*_j\right)\right\}}_{\textit{Parameter Instability II}} \\
&+ \underbrace{P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\frac{\partial S_{t+h}(\theta^*_t)}{\partial\theta}\left(\hat{\theta}_t - \theta^*_t\right)\right]}_{\textit{Estimation Uncertainty I}} + \underbrace{P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left\{\left[\left(\hat{\theta}_t - \theta^*_t\right)'\frac{\partial^2\overline{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'} + \frac{\partial\overline{L}_t(\theta^*_t)}{\partial\theta} - \frac{\partial\overline{S}_t(\theta^*_t)}{\partial\theta}\right]\left(\hat{\theta}_t - \theta^*_t\right)\right\}}_{\textit{Estimation Uncertainty II}} \\
&+ \underbrace{\frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\left(\hat{\theta}_t - \theta^*_t\right)'\left(\frac{\partial^2 S_{t+h}(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'} - \frac{\partial^2 S_j(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right)\left(\hat{\theta}_t - \theta^*_t\right)\right]}_{\textit{Estimation Uncertainty III}}
\end{aligned}
$$

where $L(\cdot)$ denotes the loss function/objective function for estimation, $\overline{\sum}_j$ denotes the in-sample average that corresponds to the estimation scheme[17], and $(\bar{\theta}^*_{t+h}, \bar{\theta}^*_j, \bar{\hat{\theta}}_t)$ denote some intermediate points between the intervals $(\theta^*_t, \theta^*_{t+h}), (\theta^*_t, \theta^*_j), (\hat{\theta}_t, \theta^*_t)$, respectively.

Proposition 1 implies that the DFB can result from parameter instability, estimation uncertainty and other instabilities. "Other Instabilities" include any changes, other than parameter instabilities, that cause an expected scoring surprise. "Parameter Instability I" captures changes in parameters with magnitude $\theta^*_t - \theta^* = Op(P^{1/2})$, which is the same magnitude as that considered in the structural break literature. "Parameter Instability II" indicates changes in parameters with magnitude $\theta^*_t - \theta^* = Op(P^{1/4})$, which is instead smaller than that considered in the aforementioned literature.

When the LS and QMLE are adopted simultaneously, "Estimation Uncertainty II" degenerates to $P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left\{\left(\hat{\theta}_t - \theta^*_t\right)'\frac{\partial^2\overline{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\left(\hat{\theta}_t - \theta^*_t\right)\right\}$, which is a quadratic form and is always positive. This form results from

[17]For example, $\overline{\sum}_j = \frac{1}{R}\sum_{j=1}^{R}$ for the fixed scheme.

the average in-sample score computed at the parameter estimates being maximized by construction according to the QMLE. It is therefore greater than the expected out-of-sample score and can be interpreted as a measure of "overfitting." In this case, Corollary 2 in the appendix proposes a simple correction to the DFB test statistic by subtracting an approximated "Estimation Uncertainty II" from the numerator of the DFB test statistic in Proposition 1.

1.3 Detecting Density Forecast Breakdowns for Non-parametric Models

This section extends the DFB test to non-parametric models. The researcher may characterize the h-step-ahead predictive density for the scalar variable y_{t+h} based on the conditioning information set $\mathcal{F}_t$ (denoted by $\phi(\cdot|\mathcal{F}_t)$) without imposing a parametric form. Hence, the sequence of P out-of-sample density forecasts evaluated at the ex-post realizations would be denoted as $\hat{\phi}^f_{t+h}(y_{t+h}) = \phi(y_{t+h}|\mathcal{F}_t), t = R, \ldots, T-h$. Similarly, the density estimator evaluated at the realization for period j (with j varying over the corresponding in-sample window) would be denoted by $\hat{\phi}^e_j(y_j) = \phi(y_j|\mathcal{F}_t)$. Considering that only a subset of the true information set ($\mathcal{F}_t \subset \mathbb{F}_t$) may be observed, the pseudo-true conditional density of the scalar variable y_{t+h} would be denoted by $\phi^*_{t+h}(\cdot) = \phi(\cdot|\mathcal{F}_t)$. In the parametric case, the evaluation function becomes a function of the parameter vector θ. However, in this case, the evaluation function for the h-step-ahead density forecast for the scalar variable y_{t+h} constructed at period t would be a function of the predictive density function evaluated at the corresponding realization. The LS, for example, denotes the score for the h-step-ahead density forecast for the scalar variable y_{t+h} constructed at period t as

$$S_{t+h} \equiv S(\hat{\phi}^f_{t+h}(\cdot), y_{t+h}) = \log \phi(y_{t+h}|\mathcal{F}_t)$$

and denotes the score for period j (with j varying over the in-sample window) as

$$S_j \equiv S(\hat{\phi}_j^e(\cdot), y_j) = \log \phi(y_j|\mathcal{F}_t)$$

Similarly, $\cdot^*$ is used to refer to counterparts evaluated by the pseudo-true density $\phi^*(\cdot)$.

A fundamental difference between the parametric case and the non-parametric case is that correcting the estimation uncertainty is difficult in the non-parametric case, since the estimation uncertainty is a function of the derivatives of the density estimators. To solve this problem, I provide the following conditions to ensure that the estimation uncertainty component is asymptotically irrelevant for the non-parametric case.

Assumption NP1. *(a) The Hadamard Derivative of the scoring $S_t(\phi^*)$ exists, denoted as $D^*_{t,\phi}$, and $sup_{t\geq 1} \left\|D^*_{t,\phi}\right\| < \infty$. (b) $sup_{t\geq 1} \left\|S_t(\phi^*)\right\|^{2r} < \infty$. (c) $\mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T} S_t(\phi^*)\right) > 0$ for all sufficiently large T.*

Assumption NP2. $R, P \to \infty$. $\sqrt{P}\sup_y |\hat{\phi}_t(y) - \phi^*(y)| = op(1), \forall t = R, \ldots, T$.

Assumption NP1 is similar to Assumption P2, P4, P5 and P6 in the parametric case. Apart from the regular conditions, Assumption NP1 ensures that the functional derivative of the scoring exists and is bounded. Assumption NP2 is then a general assumption, ensuring that the estimation uncertainty is asymptotically irrelevant.

These high level assumptions are very realistic and can be replaced with more detailed assumptions when adopting different non-parametric estimation methods. Below is an example list of detailed assumptions for the kernel density estimation (KDE), one of the most commonly used non-parametric estimation methods.

Example (KDE): To determine the h-step-ahead unconditional predictive density for the scalar variable y_{t+h} based on the information set $\mathcal{F}_t = \sigma(y_1', \ldots; y_t')'$ available at time t, KDE with kernel function $K(\cdot)$ is adopted at each period t to obtain the density estimator $\hat{\phi}_t(\cdot)$ and used as the density forecast for y_{t+h}. In this case, the following assumptions are considered.

Assumption NP-KDE1. $\{y_t\}$ *is i.i.d, and have a three-time differentiable pdf* $\phi(y)$ *with* $\inf_y \phi(y) \geq \delta > 0$.

Assumption NP-KDE2. *The kernel function* $K(.)$ *is bounded and satisfies: (a)* $\int K(u)du = 1$. *(b)* $K(u) = K(-u)$. *(c)* $\int u^2 K(u)du = \kappa_2 > 0$.

Assumption NP-KDE3. *The estimation window size* R *and the corresponding bandwidth parameter* h_R *satisfy that, as* $R \to \infty$, $h_R \to 0$ *and* $Rh_R \to \infty$.

Assumption NP-KDE4. $R, P \to \infty$, $\frac{P \ln R}{Rh_R} \to 0$ *and* $Ph_R^4 \to 0$.

Assumption NP-KDE5. *(a) The Hadamard Derivative of the scoring* $S_t(\phi^*)$ *exists, denoted as* $D^*_{t,\phi}$, *and* $sup_{t\geq 1} \left\| D^*_{t,\phi} \right\| < \infty$. *(b)* $sup_{t\geq 1} \left\| S_t(\phi^*) \right\|^{2r} < \infty$. *(c)* $\text{Var}\left(T^{-1/2} \sum_{t=1}^{T} S_t(\phi^*) \right) > 0$ *for all sufficiently large* T.

Assumption NP-KDE1 can be relaxed to allow the data to be mixing, because time series dependence has no effect on the asymptotic bias and variance of the kernel estimator. In addition, KDE averages the data locally in the y-dimension, where there is no time-series dependence. Assumptions NP-KDE2 and NP-KDE3 are standard assumptions for the kernel function and bandwidth parameter, respectively, which ensure the consistency of the KDE. Assumption NP-KDE5 ensures that the density estimation uncertainty component converges to zero asymptotically and works similarly to Assumptions P4, P5 and P6 in the parametric case.

Theorem 2. ***(Generalization of Asymptotic Distribution of the DFB Test for Non-parametric Models)*** *Given Assumptions NP1 and NP2, if* V_T *defined below is p.d., then under* H_0 *in (1.3),* $t_{R,P,h} \xrightarrow{d} \mathcal{N}(0,1)$, *where* $t_{R,P,h}$ *is defined in* (1.5) *and* $\hat{\sigma}^2_{R,P}$ *is defined as*

$$\hat{\sigma}^2_{R,P} = \frac{T}{P}\left\{ \frac{1}{T}\sum_{t=1}^{T}(\omega_t^S \tilde{S}_t)^2 + \frac{2}{T}\sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} \omega_t^S \tilde{S}_t \omega_{t-j}^S \tilde{S}_{t-j} \right\}$$

where bandwidth p_T *and weights* $v_{T,j}$ *are appropriately chosen as per [Newey and West, 1987].*

Similar to the parametric case, Proposition 2 below again considers the expectation of the DFB statistic's numerator in order to analyze the causes of the DFB.

Proposition 2. ***Causes of density forecast breakdowns for non-parametric models***

$$
\begin{aligned}
\mathrm{E}\left[P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\phi})\right] &= \underbrace{\mathrm{E}\left[P^{-1/2}\sum_{t=R}^{T-h}\left(S_{t+h}(\phi^*_{t+h}) - \overline{\Sigma}_j S_j(\phi^*_j)\right)\right]}_{\textit{Other Instabilities}} \\
&+ \underbrace{P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\frac{\partial S_{t+h}(\phi^*_{t+h})}{\partial\phi(y)}\right]\left(\phi^*_t(y_{t+h}) - \phi^*_{t+h}(y_{t+h})\right) - P^{-1/2}\sum_{t=R}^{T-h}\overline{\Sigma}_j\mathrm{E}\left[\frac{\partial S_j(\phi^*_j)}{\partial\phi(y)}\right]\left(\phi^*_t(y_j) - \phi^*_j(y_j)\right)}_{\textit{Density Instability I}} \\
&+ \underbrace{\frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}\left\{\left(\phi^*_t(y_{t+h}) - \phi^*_{t+h}(y_{t+h})\right)'\mathrm{E}\left[\frac{\partial^2 S_{t+h}(\bar{\phi}^*_{t+h})}{\partial\phi(y)\partial\phi(y')'}\right]\left(\phi^*_t(y'_{t+h}) - \phi^*_{t+h}(y'_{t+h})\right) - \overline{\Sigma}_j\left(\phi^*_t(y_j) - \phi^*_j(y_j)\right)'\mathrm{E}\left[\frac{\partial^2 S_j(\bar{\phi}^*_j)}{\partial\phi(y)\partial\phi(y')'}\right.\right.}_{\textit{Density Instability II}} \\
&+ \underbrace{P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\frac{\partial S_{t+h}(\phi^*_t)}{\partial\phi(y)}\left(\hat{\phi}_t(y_{t+h}) - \phi^*_t(y_{t+h})\right)\right]}_{\textit{Estimation Uncertainty I}} - \underbrace{P^{-1/2}\sum_{t=R}^{T-h}\overline{\Sigma}_j\mathrm{E}\left[\frac{\partial S_j(\phi^*_t)}{\partial\phi(y)}\left(\hat{\phi}_t(y_j) - \phi^*_t(y_j)\right)\right]}_{\textit{Estimation Uncertainty II}} \\
&+ \underbrace{\frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\left(\hat{\phi}_t(y_{t+h}) - \phi^*_t(y_{t+h})\right)'\left(\frac{\partial^2 S_{t+h}(\bar{\bar{\phi}}_t)}{\partial\phi(y)\partial\phi(y')'} - \frac{\partial^2 S_j(\bar{\bar{\phi}}_t)}{\partial\phi(y)\partial\phi(y')'}\right)\left(\hat{\phi}_t(y'_j) - \phi^*_t(y'_j)\right)\right]}_{\textit{Estimation Uncertainty III}}
\end{aligned}
$$

where $(\bar{\phi}^*_{t+h}, \bar{\phi}^*_j, \bar{\bar{\phi}}_t)$ denotes some intermediate functionals evaluated at the corresponding realizations between the intervals (ϕ^*_t, ϕ^*_{t+h}), (ϕ^*_t, ϕ^*_j), $(\hat{\phi}_t, \phi^*_t)$ respectively.

Proposition 2 implies that the DFB can result from density instability, estimation uncertainty and other instabilities. Similar to the parametric case, "Other Instbility" captures any changes, other than a shift in the density itself, that cause an expected scoring surprise. "Density Instability I" captures changes in the density function with magnitude $\phi^*_t(y) - \phi^*(y) = Op(P^{1/2})$, and "Density Instability II" captures changes in the density function with magnitude $\phi^*_t(y) - \phi^*(y) = Op(P^{1/4})$.

1.4 Extensions

1.4.1 Fluctuation DFB test

The DFB test discussed above focuses on the global performance of the density forecast model. However, the relative out-of-sample performance, compared with the in-sample performance, may itself be time-varying. The fluctuation test formulated by [Giacomini and Rossi, 2010] can thus be used to construct a fluctuation DFB test for this case[18]. The fluctuation DFB test statistics are based on measurements of the entire time path of the local averaged scoring surprise, computed over the rolling out-of-sample windows of size M (M is defined formally below). Compared with the DFB test, this fluctuation DFB test considers the possibility that useful information may be lost when averaging the global scoring surprise.

The fluctuation DFB test statistic is built on measurements of the local averaged scoring surprise and assesses whether the averaged scoring surprise equals zero at each point in time. In contrast to the null hypothesis in (1.3) for the DFB test, it focuses on the robust null hypothesis H_0^F:

$$H_0^F : \mathrm{E}\left[SS_{t+h}^*\right] = 0, \forall t = R, \ldots, T-h \tag{1.8}$$

Define the local out-of-sample mean of the scoring surprise:

$$\overline{SS}_{R,M,t} \equiv \frac{1}{M} \sum_{j=t-M/2}^{t+M/2-1} SS_j, \qquad t = R+M/2, \ldots, T-M/2+1 \tag{1.9}$$

The following proposition describes the asymptotic distribution of the fluctuation DFB test.

Proposition 3. ***The Fluctuation DFB Test*** *Suppose the assumptions for the DFB test hold. Suppose the following conditions are satisfied: (i)*

[18] [Giacomini and Rossi, 2010] elaborated a fluctuation test for forecast comparisons in an unstable environment.

$\{P^{-1/2}\sum_{t=R}^{R+[\tau P]} SS^*_{t+h}\}$ *obeys a Functional Central Limited theorem; (ii)* $M/P \to \mu \in (0,\infty)$ *as* $M \to \infty$, $P \to \infty$. *Let*

$$\mathcal{F}_{t,R,M} = \frac{M^{-1/2}\sum_{j=t-M/2}^{t+M/2-1} SS_j}{\hat{\sigma}_{R,P}}, \qquad t = R + M/2, \ldots, T - M/2 + 1 \tag{1.10}$$

where $\hat{\sigma}^2_{R,P}$ *is the asymptotic variance estimator of* $P^{1/2}\overline{SS}_{R,P}$. *Under the null hypothesis* H_0^F *in* (1.8)*:*

$$\mathcal{F}_{t,R,M} \Rightarrow [\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu} \tag{1.11}$$

where $t = [\tau P]$, $M = [\mu P]$ *and* $\mathcal{B}(\cdot)$ *is a standard univariate Brownian motion. Then,*

$$\sup_t \mathcal{F}_{t,R,M} \xrightarrow{d} \sup_\tau [\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu} \tag{1.12}$$

The critical value κ_α *for a significance level* α *solves the problem that* $Pr\left[\sup_\tau |[\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu}| > \kappa_\alpha\right] = \alpha$.

The fluctuation DFB test is a [Giacomini and Rossi, 2010] fluctuation test applied to the DFB test. Additionally, it can be applied to both parametric and non-parametric cases, with the corresponding scoring surprises as defined in sections 1.2.3 and 1.3.

1.4.2 DFB test robust to the choice of R

The choice of window size presents an additional issue, since arbitrarily choosing the size of the estimation window might lead to different empirical results in practice. For example, satisfactory results might simply be obtained after data snooping over window sizes. To solve this problem, [Inoue and Rossi, 2012] proposed a methodology for evaluating the out-of-sample forecasting performance of models that is robust to the choice of in-sample window size. The DFB test can be similarly adapted to be robust to the choice of estimation window size by regarding the window size R as a nuisance parameter.

Proposition 4. ***(Asymptotic Distribution of the Test Statistics Robust to the Choice of*** R***)*** *Let $\mathcal{T}_T(R)$ denote the DFB test statistic as a function of the in-sample size R, that is*

$$\mathcal{T}_T(R) \equiv t_{R,P,h} = \frac{P^{1/2}\overline{SS}_R}{\hat{\sigma}_{R,P}} = \frac{P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_{t,R})}{\hat{\sigma}_{R,P}} \tag{1.13}$$

where $\hat{\sigma}^2_{R,P}$ is the asymptotic variance estimator of $P^{1/2}\overline{SS}_{R,P}$, and $\hat{\theta}_{t,R}$ is the estimator at period t with the in-sample window size R.

Suppose the test statistic $\mathcal{T}_T(\cdot)$ satisfies

$$\mathcal{T}_T([\imath(\cdot)T]) \Rightarrow \mathcal{T}(\cdot) \tag{1.14}$$

where $\imath(\cdot)$ is the identity function, that is, $\imath(x) = x$ and $\Rightarrow$ denotes weak convergence in the space of cadlag functions on $[0,1]$ equipped with the Skorokhod metric. Then,

$$\sup_{[\underline{\mu}T]\leq R\leq[\bar{\mu}T]} \mathcal{T}_T(R) \xrightarrow{d} \sup_{[\underline{\mu}T]\leq R\leq[\bar{\mu}T]} \mathcal{T}(\mu) \tag{1.15}$$

$$\frac{1}{[\bar{\mu}T]-[\underline{\mu}T]+1}\sum_{R=\underline{\mu}T}^{[\bar{\mu}T]} \mathcal{T}_T(R) \xrightarrow{d} \int_{\underline{\mu}}^{\bar{\mu}} \mathcal{T}(\mu)d\mu \tag{1.16}$$

where $0 < \underline{\mu} < \bar{\mu} < 1$.

This approach assumes that R is growing with the sample size and asymptotically becomes a fixed fraction of the total sample size, an assumption that is consistent with the approaches used by [West, 1996], [West and McCracken, 1998], [McCracken, 2000], and [Inoue and Rossi, 2012]. In the existing DFB test, $\mu = \lim_{T\to\infty} R/T$ is fixed, and condition (1.14) holds pointwise for a given μ. In Proposition 4, however, condition (1.14) requires that the convergence holds uniformly in μ rather than pointwise.

1.4.3 Other extensions

Other extensions of the DFB test are possible. For example, while the DFB test discussed above focuses on the averaged scoring surprises, which give equal weight to each out-of-sample period, there may be a need to weight periods differently. Unlike the null hypothesis used in (1.3) for the DFB test, a weighted version of the DFB test would focus on a null hypothesis H_0^w: $E\left[\frac{1}{P}\sum_{t=R}^{T-h} w_t SS_{t+h}^*\right] = 0$ where the weight function w_t depends on t and can be chosen to weight the desired periods. It could, for instance, be of interest to capture a DFB over a particular period of time. In this case, targeted periods could be given more weight and other less weight. A simple example of investigating whether there is a DFB between period T_1 and T_2 ($R < T_1 < T_2 < T - h$) could be

$$w_t = \begin{cases} 1 & \text{if} \quad T_1 \leqslant t < T_2 \\ 0 & \text{otherwise,} \end{cases}$$

which keeps the scoring surprises between period T_1 and T_2 and drops those outside this range. Furthermore, since the weighting w_t is determined separately from the evaluation function and the data, the only additional assumption required to retain the properties of the DFB test statistic is $w_t < \infty, \forall t$.

Additionally, the DFB test can be easily extended to evaluate joint density forecast models, since the evaluation functions can be adjusted to evaluate joint predictive densities. For example, the DFB test statistic can be constructed based on a multivariate CPRS in the same way as with the univariate setting; one such multivariate CRPS is defined by [Gneiting and Raftery, 2007].

1.5 Monte Carlo Evidence

This section analyzes the size and power properties of the DFB test in finite samples, considering both parametric and non-parametric models.

1.5.1 Size properties

In what follows, I describe the simulation study investigating the size properties of the DFB test. The data generating process (DGP) considered here is:

$$y_{t+h} = c^* + x_t\gamma^* + y_t\beta^* + \epsilon_{t+h} \tag{1.17}$$

for $t = 1, \ldots, T$, with i.i.d. regressors $x_t \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$, and error terms $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^{*2})$. The parameters are set as follows: $(c^*, \gamma^*, \beta^*) = (0.5, 0.2, 0.3)$, $\sigma^{*2} = 1$. The choice of this autoregressive model that contains additional regressors (i.e., an ARX model) is inspired by the empirical application in Section 3.4. Additional examples of the ARX model include the Phillips curve, which, as a forecast model of inflation, relates changes in inflation to past values of the unemployment gap and of inflation (see [Stock and Watson, 2003], [Corradi and Swanson, 2006], and [Giacomini and Rossi, 2009a]).

I consider three scenarios.

Scenario 1: Conditional density forecasts with a correctly specified model

This scenario focuses on the conditional density forecasts of y_{t+h} where the model is correctly specified. The OLS estimation is applied to eq (1.17) at each period with the corresponding in-sample observations and the OLS estimators $\hat{\theta}_t = (\hat{c}_t, \hat{\gamma}_t, \hat{\beta}_t, \hat{\sigma}_t^2)$ are obtained at each period t. Given the OLS estimators, the conditional density forecast of y_{t+h} is

$$\hat{\phi}^f_{t+h}(y) = \phi(y|x_t, y_t, \hat{\theta}_t) = \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} e^{-\frac{(y-\hat{c}_t - x_t\hat{\gamma}_t - y_t\hat{\beta}_t)^2}{2\hat{\sigma}_t^2}}$$

Scenario 2: Conditional density forecasts with a misspecified model

This scenario focuses on the conditional density forecasts of y_{t+h} where the model is misspecified. The OLS estimation is applied on the following regression at each period with the corresponding in-sample observations, and the OLS estimators $\hat{\theta}_t = (\hat{c}_t, \hat{\gamma}_t, \hat{\beta}_t, \hat{\sigma}_t^2)$ are obtained at each period t.

$$y_{t+h} = c^* + y_t\beta^* + \epsilon_{t+h}$$

Given the OLS estimators, the conditional density forecast of y_{t+h} is

$$\hat{\phi}^f_{t+h}(y) = \phi(y|y_t, \hat{\theta}_t) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2_t}} e^{-\frac{(y-\hat{c}_t - y_t\hat{\beta}_t)^2}{2\hat{\sigma}^2_t}}$$

Scenario 3: Unconditional density estimation using the KDE

This final scenario focuses on the unconditional density forecasts of y_{t+h}. In this case, the KDE is applied at each period with the corresponding in-sample observations, and the KDEs $\hat{\phi}^e_t(y)$ are obtained at each period t. Given the KDEs, the unconditional density forecast of y_{t+h} is:

$$\hat{\phi}^f_{t+h}(y) = \hat{\phi}^e_t(y) = \frac{1}{Rh_R} \sum_{t=1}^{R} \kappa(\frac{y_t - y}{R}), \qquad h_R = cR^{-1/5}$$

where h_R is the bandwidth parameter. In this simulation study, the kernel function $\kappa(\cdot)$ is a standard normal kernel $\kappa(\nu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2}$, [19] and the constant $c = 1.06$.[20]

All three scenarios consider different combinations given by $R, P \in \{50, 100, 200, 400\}$ for the in-sample and out-of-sample sizes (R, P) along with $h = 1, 4$ for the forecast horizons. For each grouping (R, P, h), 5000 Monte Carlo replications are conducted, generating $T = R + P + h - 1$ data points as in (1.17). All three evaluation functions discussed in Section 1.2.2 are considered: LS, CRPS and CI. The DFB test is considered for the rolling window scheme and recursive window scheme using the general asymptotic variance estimator of Theorem 1, Corollary 1 and Theorem 2, with truncation lags for the HAC estimator set to be $T^{1/3}$.

Tables 1.1, 1.2, and 1.3 provide the rejection frequencies of the DFB test for various (R, P) pairs for each scenario. The results imply that the DFB test has good size properties for (i) moderate-large samples (i.e., both R and P are large), (ii) both parametric and nonparametric models, and (iii) all three evaluation functions. When the in-sample size is

[19] Other kernel functions can also be used.

[20] See the pilot estimate of the optimal bandwidth for a standard normal kernel function in [Li and Racine, 2007].

much smaller than the out-of-sample size, the test performed poorly for the rolling window scheme.[21] However, the test performed well for the recursive window scheme.

1.5.2 Power properties

This section describes the simulation study investigating the power properties of the DFB test. The following scenarios with various sources of DFBs (DGP-P1, DGP-P2, DGP-P3) are considered and are described in detail below.

DGP-P1: Change in mean Consider a one-time change in mean in DGP:

$$y_t = \alpha_\beta \cdot 1(t > T_b) + \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1) \tag{1.18}$$

Let $(R, P) = (200, 100)$, and let $T_b = R + 0.15P$. α_β controls the size of the change in mean. The power curve is obtained by letting α_β vary between 0 and 2.

DGP-P2: Change in variance Consider a one-time change in variance in DGP:

$$\begin{aligned} y_t &= \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_t^{*2}) \\ \sigma_t^{*2} &= 1 + \alpha_\sigma \cdot 1(t > T_b) \end{aligned} \tag{1.19}$$

Let $(R, P) = (200, 100)$, and let $T_b = R + 0.15P$. α_σ controls the size of the change in variance. The power curve is obtained by letting α_σ vary between 0 and 2.5.

DGP-P3: Change in type of distribution Consider a one-time change in the type of the distribution in DGP:

$$\begin{aligned} &y_t = \epsilon_t \\ &\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_t^{*2}), && t = 1, \ldots, T_b \\ &\epsilon_t \overset{i.i.d}{\sim} (1 - \alpha_d)\mathcal{N}(0, \sigma_t^{*2}) + \alpha_d \chi^2_{k^*}, && t = T_b + 1, \ldots, T \end{aligned} \tag{1.20}$$

[21] This is due to the overfitting component, which becomes asymptotically relevant when the out-of-sample size P grows faster than the $R^{3/2}$, causing the test statistics to diverge (for a detailed discussion, see [Giacomini and Rossi, 2009a]). An overfitting corrected DFB test statistic is provided in Corollary 2 in the appendix.

Table 1.1: Size of DFB test, Scenario 1, Normal size 0.05

		$h=1$						$h=4$					
		LS		CRPS		CI		LS		CRPS		CI	
R	P	roll	rec	roll	rec	roll	rec	roll	rec	roll	rec	roll	rec
50	50	0.068	0.082	0.052	0.062	0.090	0.100	0.076	0.082	0.064	0.074	0.084	0.084
50	100	0.068	0.060	0.042	0.066	0.056	0.060	0.070	0.064	0.044	0.062	0.058	0.094
50	200	0.084	0.050	0.036	0.040	0.054	0.078	0.094	0.066	0.032	0.048	0.050	0.080
50	400	0.104	0.040	0.024	0.048	0.046	0.066	0.114	0.054	0.026	0.048	0.056	0.080
100	50	0.056	0.066	0.042	0.046	0.068	0.072	0.076	0.074	0.064	0.068	0.090	0.094
100	100	0.072	0.068	0.062	0.074	0.050	0.062	0.074	0.078	0.048	0.062	0.068	0.082
100	200	0.082	0.070	0.044	0.062	0.048	0.066	0.052	0.060	0.046	0.048	0.074	0.098
100	400	0.090	0.066	0.034	0.054	0.062	0.062	0.056	0.054	0.028	0.054	0.066	0.070
200	50	0.062	0.066	0.054	0.056	0.068	0.066	0.060	0.062	0.064	0.068	0.064	0.072
200	100	0.050	0.066	0.048	0.050	0.056	0.070	0.064	0.062	0.044	0.046	0.092	0.072
200	200	0.044	0.052	0.072	0.050	0.072	0.090	0.048	0.050	0.052	0.044	0.060	0.064
200	400	0.070	0.068	0.042	0.044	0.040	0.046	0.060	0.060	0.054	0.060	0.062	0.056
400	50	0.030	0.034	0.03	0.032	0.048	0.042	0.024	0.030	0.048	0.042	0.052	0.052
400	100	0.076	0.070	0.046	0.054	0.048	0.054	0.054	0.044	0.050	0.048	0.078	0.070
400	200	0.056	0.042	0.044	0.048	0.048	0.064	0.042	0.058	0.054	0.064	0.072	0.074
400	400	0.060	0.052	0.042	0.044	0.064	0.068	0.052	0.074	0.054	0.068	0.064	0.070

Note: This table reports rejection frequencies over 5000 Monte Carlo replications of the Density Forecast Breakdown test, using either the general asymptotic variance estimator of Corollary 1 (LS) or of Theorem 1 (CRPS, CI), implemented with rolling and recursive schemes. R and P denote the in-sample and out-of-sample sizes, respectively. For CI, $(\underline{u}, \overline{u})$ are set to be $(-2, 2)$.

Table 1.2: Size of DFB test, Scenario 2, Normal size 0.05

		$h=1$						$h=4$					
		LS		CRPS		CI		LS		CRPS		CI	
R	P	roll	rec	roll	rec	roll	rec	roll	rec	roll	rec	roll	rec
50	50	0.058	0.074	0.072	0.090	0.068	0.076	0.080	0.074	0.060	0.084	0.114	0.136
50	100	0.086	0.066	0.062	0.062	0.042	0.064	0.056	0.050	0.062	0.078	0.076	0.094
50	200	0.078	0.046	0.032	0.060	0.044	0.072	0.066	0.048	0.040	0.054	0.064	0.068
50	400	0.118	0.076	0.036	0.042	0.046	0.082	0.114	0.066	0.022	0.048	0.050	0.044
100	50	0.078	0.098	0.056	0.062	0.078	0.090	0.074	0.074	0.090	0.084	0.070	0.068
100	100	0.056	0.064	0.070	0.062	0.072	0.084	0.054	0.052	0.052	0.066	0.076	0.098
100	200	0.050	0.062	0.048	0.056	0.064	0.104	0.058	0.058	0.046	0.058	0.070	0.092
100	400	0.054	0.044	0.042	0.044	0.050	0.048	0.088	0.050	0.040	0.066	0.060	0.084
200	50	0.080	0.078	0.056	0.052	0.052	0.058	0.066	0.068	0.048	0.040	0.060	0.060
200	100	0.062	0.056	0.050	0.056	0.070	0.070	0.050	0.056	0.052	0.058	0.082	0.082
200	200	0.060	0.062	0.054	0.076	0.068	0.076	0.072	0.072	0.064	0.054	0.084	0.076
200	400	0.058	0.070	0.058	0.050	0.048	0.064	0.068	0.056	0.062	0.060	0.074	0.082
400	50	0.046	0.046	0.028	0.034	0.038	0.044	0.050	0.044	0.044	0.044	0.062	0.060
400	100	0.058	0.052	0.076	0.068	0.068	0.074	0.068	0.068	0.058	0.048	0.058	0.056
400	200	0.076	0.072	0.066	0.068	0.042	0.036	0.066	0.052	0.058	0.054	0.058	0.064
400	400	0.054	0.050	0.052	0.060	0.062	0.060	0.042	0.056	0.052	0.060	0.066	0.052

Note: This table reports rejection frequencies over 5000 Monte Carlo replications of the Density Forecast Breakdown test, using either the general asymptotic variance estimator of Corollary 1 (LS) or of Theorem 1 (CRPS, CI), implemented with rolling and recursive schemes. R and P denote the in-sample and out-of-sample sizes, respectively. For CI, $(\underline{u}, \overline{u})$ are set to be $(-2, 2)$.

Table 1.3: Size of DFB test, Scenario 3, Normal size 0.05

		$h=1$						$h=4$					
		LS		CRPS		CI		LS		CRPS		CI	
R	P	roll	rec	roll	rec	roll	rec	roll	rec	roll	rec	roll	rec
50	50	0.076	0.104	0.078	0.108	0.087	0.090	0.256	0.136	0.112	0.102	0.058	0.080
50	100	0.092	0.086	0.060	0.064	0.087	0.084	0.680	0.218	0.194	0.092	0.048	0.070
50	200	0.100	0.076	0.128	0.062	0.093	0.085	0.974	0.184	0.482	0.088	0.052	0.076
50	400	0.210	0.046	0.212	0.078	0.126	0.066	1.000	0.194	0.910	0.082	0.088	0.064
100	50	0.068	0.074	0.058	0.062	0.068	0.069	0.094	0.072	0.072	0.078	0.100	0.100
100	100	0.048	0.058	0.058	0.068	0.066	0.071	0.232	0.150	0.092	0.058	0.056	0.062
100	200	0.048	0.068	0.066	0.074	0.075	0.077	0.510	0.154	0.132	0.076	0.064	0.060
100	400	0.086	0.056	0.068	0.070	0.079	0.067	0.956	0.154	0.258	0.052	0.062	0.088
200	50	0.046	0.052	0.048	0.042	0.055	0.057	0.070	0.056	0.060	0.066	0.116	0.116
200	100	0.072	0.056	0.058	0.054	0.062	0.061	0.100	0.076	0.082	0.096	0.070	0.072
200	200	0.068	0.070	0.076	0.064	0.064	0.065	0.160	0.102	0.062	0.048	0.044	0.056
200	400	0.074	0.096	0.056	0.064	0.055	0.060	0.428	0.116	0.082	0.054	0.082	0.088
400	50	0.042	0.044	0.046	0.046	0.033	0.034	0.050	0.052	0.048	0.048	0.094	0.094
400	100	0.056	0.070	0.058	0.062	0.057	0.057	0.074	0.070	0.060	0.070	0.076	0.086
400	200	0.050	0.048	0.056	0.060	0.058	0.060	0.078	0.072	0.060	0.054	0.062	0.056
400	400	0.050	0.058	0.056	0.064	0.055	0.048	0.118	0.094	0.069	0.086	0.056	0.062

Note: This table reports rejection frequencies over 5000 Monte Carlo replications of the Density Forecast Breakdown test, using either the general asymptotic variance estimator of Theorem 2 (LS, CRPS, CI), implemented with rolling and recursive schemes. R and P denote the in-sample and out-of-sample sizes, respectively. For CI, $(\underline{u}, \overline{u})$ are set to be $(-2, 2)$.

Let $(R, P) = (200, 100)$, and let $T_b = R + 0.15P$. The parameter k^* controls the degree of freedom of the χ^2 distribution. k^* is set to be 2, and the power curve is obtained by letting α_d vary between 0 to 1.

For each scenario, both the parametric and the non-parametric models are considered for the h-step-ahead (un)conditional density forecast of y_{t+h}. For the parametric case, the OLS estimation is applied to the following regression at each period with the corresponding in-sample observations:

$$y_{t+h} = \beta_0^* + x_t\beta_1^* + \epsilon_{t+h}$$

As for forecasting, given the OLS estimators $\hat{\beta}_{0,t}$, $\hat{\beta}_{1,t}$ and $\hat{\sigma}_t$ at each period t, the h-step-ahead density forecast of y_{t+h}, assuming normality of ϵ_t, is

$$\hat{\phi}^f_{t+h}(y) = \phi(y|\hat{\beta}_{0,t}, \hat{\beta}_{1,t}, \hat{\sigma}_t^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} e^{-\frac{(y-\hat{\beta}_{0,t}-x_t\hat{\beta}_{1,t})^2}{2\hat{\sigma}_t^2}}$$

For the non-parametric case, the KDE is applied at each period with the corresponding in-sample observations, and the KDEs $\hat{\phi}^e_t(y)$ are obtained at each period t. As for forecasting, the h-step-ahead density forecast is the same as the density estimator at period t:

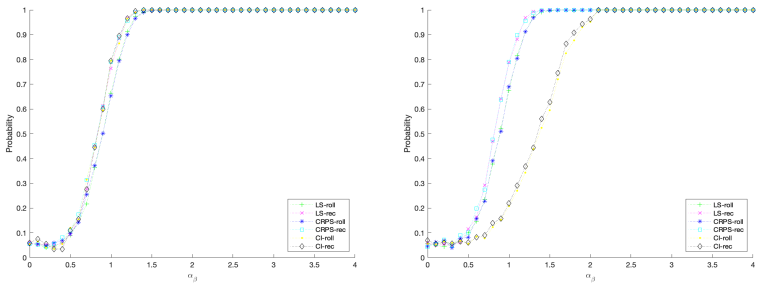
$$\hat{\phi}^f_{t+h}(y) = \hat{\phi}^e_t(y) = \frac{1}{Rh_R} \sum_{s=t-R+1}^{t} \kappa(\frac{y_s - y}{R}), \qquad h_R = cR^{-1/5}$$

where h_R is the bandwidth parameter. The kernel function $\kappa(\cdot)$ is a standard normal kernel ($\kappa(\nu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\nu^2}$) and the constant $c = 1.06$.

I conducted 5000 Monte Carlo replications for each scenario for both the rolling and the recursive window schemes, as well as for the general asymptotic variance estimator of Theorem 1, Corollary 1, and Theorem 2, with truncation lags for the HAC estimator set to be 0.

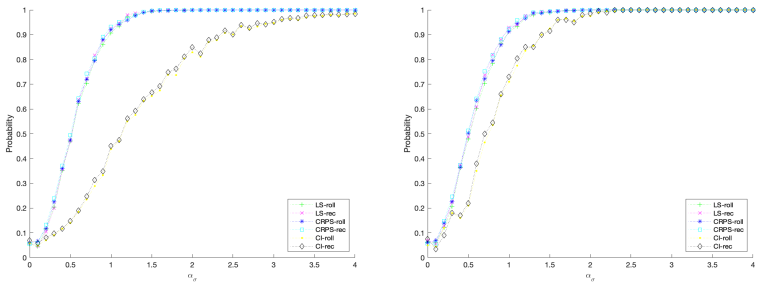
Figures 1.2, 1.3, and 1.4 display the power curves of the DFB test with DGP-P1, P2, and P3 using both parametric and nonparametric models. These power curves imply that the DFB test has power against changes

in mean, variance, and distribution type for both parametric and non-parametric models. In most cases, the CI evaluation function has relatively less power compared with the LS and the CRPS evaluation functions. This could be due to the fact that the CI, especially compared to the CRPS, focuses on specific regions of the distributions. Hence, it is less powerful against certain changes in mean and variance.



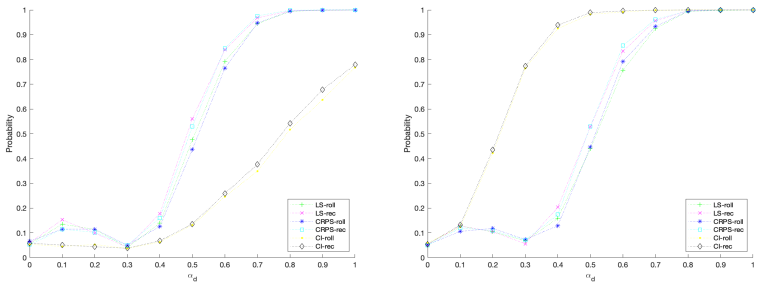
(a) Parametric (OLS) (b) Non-parametric (KDE)

Figure 1.2: Power function, a one-time change in mean



(a) Parametric (OLS) (b) Non-parametric (KDE)

Figure 1.3: Power function, a one-time change in variance



(a) Parametric (OLS) (b) Non-parametric (KDE)

Figure 1.4: Power function, a one-time shift in distribution type

1.6 Empirical Analysis

To illustrate the empirical usefulness of the proposed test, I investigate whether there are DFBs in the density forecast models of real GDP growth and daily S&P 500 stock price returns in the US.

1.6.1 Real GDP growth in the US

This section evaluates the density forecast model of real GDP growth in the US adopted by [Adrian et al., 2019], where the distribution of the future annualized average real GDP growth rate between t and $t+h$, denoted as y_{t+h}, is modeled as a function of current financial and economic conditioning variables, denoted as x_t, with the constant included.[22]

Real GDP growth is calculated as the percentage change from the preceding quarter.[23] The dependent variable is the annualized average of the quarter-to-quarter real GDP growth. As for the conditioning variables, the economic condition is characterized by the quarter-to-quarter real GDP growth and the financial condition is characterized by the National Fi-

[22] That is $P[y_{t+h} < y|x_t] = F_{y_{t+h}|x_t}(y - x_t\beta)$, where the shape of $F.(\cdot)$ is not precisely known.

[23] Downloaded from FRED.

nancial Conditions Index (NFCI)[24], which is a weighted average of 105 measures of financial activity, each expressed relative to their sample averages and scaled by their sample standard deviations. The NFCI provides a weekly estimate of US financial conditions in money markets, debt and equity markets, and the traditional and shadow banking systems. Positive NFCI values indicate that financial conditions are tighter than average. [Brave and Butters, 2012] show that the NFCI is a highly predictive and robust leading indicator of financial stress for horizons of up to one year. The NFCI is converted to a quarterly frequency by averaging weekly observations over each quarter. The data covers 188 quarters ($T = 188$), ranging from 1971Q1 to 2017Q4. Hereafter, y_{t+h} denotes the annualized average real GDP growth rate between t and $t+h$, and x_t denotes a vector containing the conditioning variables (the current financial and economic conditioning variables), with the constant included.

Estimation and forecasting procedure

I follow the two-step linear quantile regression estimation procedure adopted by [Adrian et al., 2019] to construct the conditional predictive distribution. In the first step, the quantile regressions of [Koenker and Bassett, 1978] is used to estimate the conditional quantile function of future real GDP growth as a function of current financial and economic conditioning variables. The regression slope $\hat{\beta}_\tau$ in a quantile regression of y_{t+h} on x_t is obtained as follows:

$$\hat{\beta}_\tau = \arg\min_{\beta_\tau} \sum_{t=1}^{T-h} \left(\tau \cdot 1_{(y_{t+h} \geq x_t\beta)} |y_{t+h} - x_t\beta_\tau| + (1-\tau) \cdot 1_{(y_{t+h} < x_t\beta_\tau)} |y_{t+h} - x_t\beta_\tau|\right)$$

where $1(\cdot)$ denotes the indicator function. The predicted value from the regression is the quantile of y_{t+h} conditional on x_t, that is $\hat{Q}_{y_{t+h}|x_t}(\tau|x_t) = x_t\hat{\beta}_\tau$.

In the second step, the estimated quantile distribution is smoothed for each quarter by interpolating between the estimated quantiles using the

[24]The NFCI is computed by the Federal Reserve Bank of Chicago.

skewed t-distribution developed by [Azzalini and Capitanio, 2003] and the probability density function (PDF) is recovered accordingly.

$$f(y;\mu,\sigma,\alpha,\nu) = \frac{2}{\sigma} t\left(\frac{y-\mu}{\sigma};\nu\right) T\left(\alpha\frac{y-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+\frac{y-\mu}{\sigma}}};\nu+1\right)$$

where $t(\cdot)$ and $T(\cdot)$ respectively denote the PDF and CDF of the Student t-distribution. The four parameters of the distribution pin down the location μ, scale σ, fatness ν, and shape α. Relative to the t-distribution, the skewed t-distribution adds the shape parameter, which regulates the skewness effect of the PDF.

For each quarter, the parameters of the skewed t-distribution function are obtained by minimizing the squared distance between the estimated quantile function $\hat{Q}_{y_{t+h}|x_t}(\tau|x_t)$ and the quantile function of the skewed t-distribution $F^{-1}(\tau;\mu_t,\sigma_t,\alpha_t,\nu_t)$ to match the 5, 25, 75, and 95 percent quantiles.

$$\{\hat{\mu}_{t+h},\hat{\sigma}_{t+h},\hat{\alpha}_{t+h},\hat{\nu}_{t+h}\} = \underset{\mu,\sigma,\alpha,\nu}{\arg\min} \sum_{\tau} \left(\hat{Q}_{y_{t+h}|x_t}(\tau|x_t) - F^{-1}(\tau;\mu,\sigma,\alpha,\nu)\right)^2$$

Using the procedure discussed above, I construct the one-quarter-ahead and the one-year-ahead density forecasts of real GDP growth using both recursive and rolling window schemes. Following [Adrian et al., 2019], the predictive distributions for 1993Q1 (one-quarter-ahead) and 1993Q4 (one-year-ahead) are estimated using data from 1971Q1 to 1992Q4, the initial in-sample range, via the two-step linear quantile regression estimation procedure introduced above. I then repeat the same procedure until the end of the sample. The total size of the out-of-sample is 99 quarters for one-quarter-ahead density forecasts and 96 quarters for one-year-ahead density forecasts.

Results for the model with financial conditions

I use the DFB test to evaluate the model, denoted as "GDP+NFCI", and Table 1.4 displays the results for the one-quarter-ahead and one-year-

ahead density forecasts using the recursive window scheme. The findings of the rolling window scheme are provided in the appendix.[25]

The main finding is that DFBs occur sporadically in specific regions of the conditional distribution. The DFB test using the LS and CRPS as evaluation functions indicate no DFB. This means that the overall in-sample fit of the model, including the financial conditions (NFCI) in the predictors, along with the skewed-t errors, is indicative of the overall out-of-sample density forecasting ability. But more strikingly, the DFB test using the CI as the evaluation function to focus on specific intervals, indicates that the left tails (especially the lower quantile between 25% and 50%) of the conditional distribution do experience DFBs. The right tails of the conditional distribution are stable (with the upper quantile between 75% and 100% experiencing improvement). Thus, the in-sample fit, with respect to specific intervals, does not provide reliable guidance for the out-of-sample density forecasting ability. The in-sample fit overstates the out-of-sample density forecast ability for lower quantiles while understating it for the upper quantiles.

I further apply the fluctuation DFB test to study the entire time path of the local out-of-sample density forecast performance relative to its in-sample fit. Figure 1.5 and 1.6 display the fluctuation DFB test statistics $F_{R,M,t}$ built on the CI evaluation function for the one-quarter-ahead and one-year-ahead density forecasts using the recursive window scheme. Results for the rolling window scheme are provided in the appendix. The fluctuation DFB test indicates that both the 0%-25% quantile and the 25%-50% quantile sporadically experience DFBs. DFBs occur during the financial crisis.

My findings are consistent with those of [Adrian et al., 2019], who find that asymmetry exists between the upper and lower conditional quan-

[25] [Adrian et al., 2019] adopted a two-step semi-parametric estimation procedure such that the estimation uncertainty component cannot be corrected in a regular way. I thus rely more on the results of the recursive scheme where the in-sample window size R expands, while the out-of-sample size P remains constant and the estimation uncertainty component is asymptotically irrelevant. That said, I also provide the results of the rolling window scheme as a complement.

Table 1.4: P-values of Density Forecast Breakdown test, GDP+NFCI, recursive

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	0.961	0.884	0.248	0.003	0.637	0.954
$h = 4$	0.688	0.803	0.124	0.005	0.341	0.986

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a recursive estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 188$, $R = 89$. The HAC truncation is $T^{1/3}$.

tiles when characterizing future real GDP growth as a function of current economic and financial conditions. They demonstrate that the estimated lower quantiles of the distribution of future GDP growth exhibit stronger variation, while the upper quantiles are stable over time. The DFB test mirrors these findings, since the density forecasts using the model are more likely to experience DFBs in the more volatile region of the distribution, while the in-sample fit is more likely to be indicative of the out-of-sample forecasting ability in the relatively more stable region.

Results for the model without financial conditions

I further use the DFB test to evaluate the model without financial conditions, i.e., modeling the future annualized average real GDP growth rate as a function of current economic conditioning variables with the constant included. This model is denoted as "GDPonly" in contrast to "GDP+NFCI". Real GDP growth data considered in this "GDPonly" model ranges from 1948Q1-2017Q4 ($T = 280$), and the initial in-sample ranges from 1948Q1 to 1983Q4 ($R = 144$). Findings from different

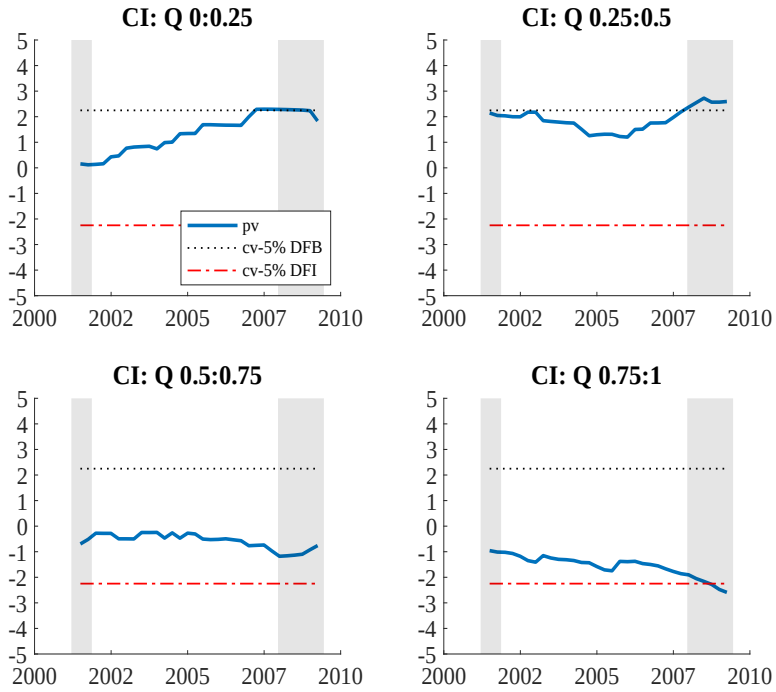


Figure 1.5: $F_{R,M,t}$ statistics, GDP+NFCI, $h = 1$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 188$, $R = 89$, $\frac{M}{P} = 0.7$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

choices of R are provided in the appendix. Table 1.5 displays the results for the DFB test for the recursive scheme. Figure 1.7 and 1.8 further display the results for the fluctuation DFB test for specific regions. Findings from the rolling window scheme are provided in the appendix.

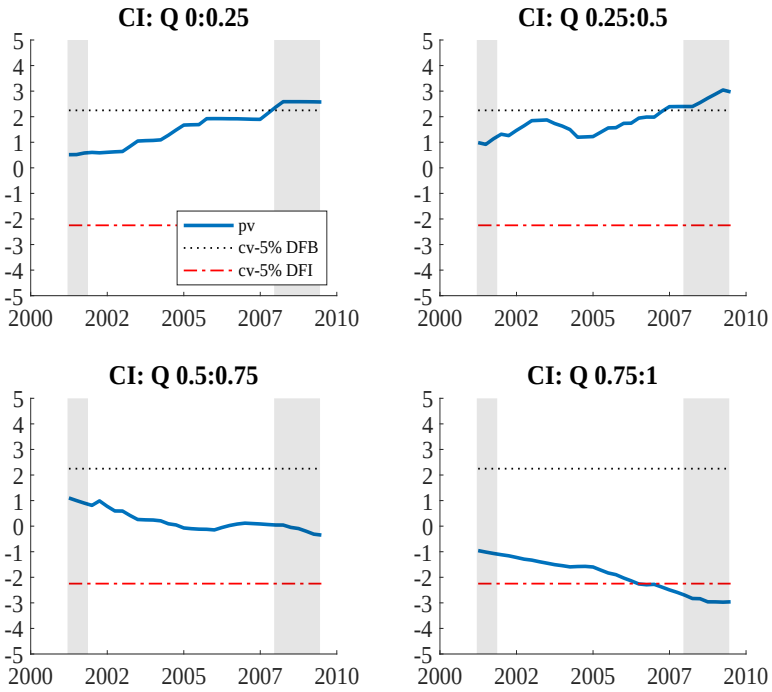


Figure 1.6: $F_{R,M,t}$ statistics, GDP+NFCI, $h = 4$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 188$, $R = 89$, $\frac{M}{P} = 0.7$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

Table 1.5: P-values of Density Forecast Breakdown test, GDPonly, recursive

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	1.000	1.000	0.925	0.000	0.002	0.997
$h = 4$	1.000	1.000	0.472	0.002	0.003	1.000

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a recursive estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 280$, $R = 144$. The HAC truncation is $T^{1/3}$.

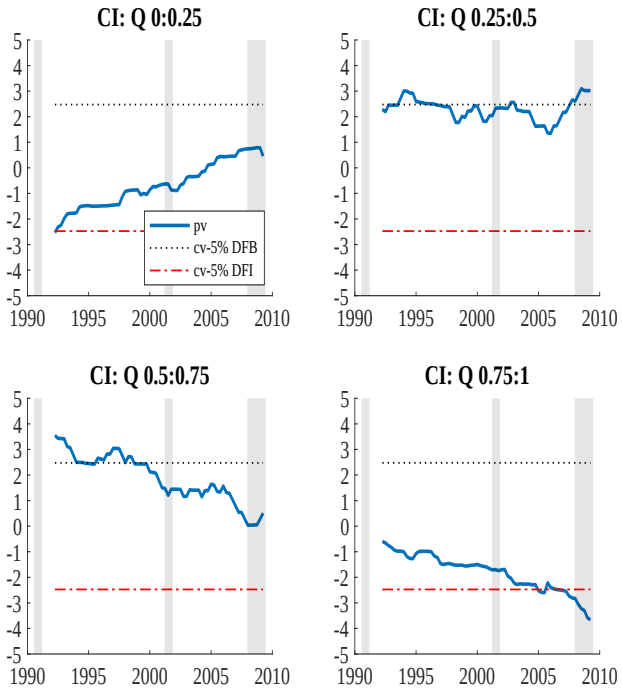


Figure 1.7: $F_{R,M,t}$ statistics, GDPonly, $h = 1$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 144$, $\frac{M}{P} = 0.5$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

Table 1.5 indicates that the model without financial conditions experiences DFBs in both the 25%-50% and 50%-75% quantiles. Figure 1.7 and 1.8 further show that DFBs frequently occur, i.e., the in-sample fit of the model without financial conditions provides unreliable guidance for the out-of-sample density forecasting ability in the 25%-50% and 50%-75% quantiles most of the time.

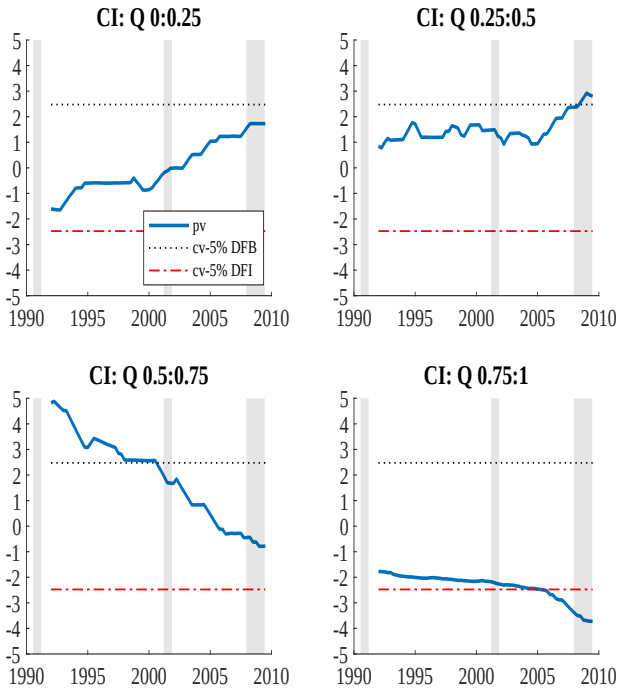


Figure 1.8: $F_{R,M,t}$ statistics, GDPonly, $h = 4$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 144$, $\frac{M}{P} = 0.5$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

Compared with the results for the model that includes financial conditions (see Figure 1.5 and 1.6), this finding confirms that the financial sector plays an important role in shaping macroeconomic performance. Indeed, including financial conditions helps narrow the gap between the in-sample fit and the out-of-sample density forecasting ability before the 2008 financial crisis. That said, discussion is warranted over which pre-

dictors to include after the financial crisis as the model including financial conditions still experiences DFBs.

Alternative test results

I also carry out alternative tests to evaluate either the in-sample density estimators or the out-of-sample predictive densities constructed using the model adopted by [Adrian et al., 2019]. These alternative tests include the [Inoue, 2001] change in distribution tests[26], as well as the [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests. Table 1.6 displays the results of the alternative tests. The [Inoue, 2001] tests do not reject the null hypothesis that there is no distributional change in either the in-sample density estimators or the out-of-sample density forecasts at the 5% significance level. However, in some cases, the null is rejected at the 10% significance level. The [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests imply that the in-sample density estimators are correctly specified, but the out-of-sample density forecasts are misspecified at the 5% significance level.

These results suggest that the DFBs in certain regions of the distribution may result from the misspecification of the out-of-sample density forecasts; that is, when using in-sample data to construct density forecasts, the predictions look poor, even though the in-sample density estimators are correctly specified and have good fit. Nonetheless, the in-sample fit can still be indicative of the out-of-sample density forecasting ability in some regions (e.g., the 50%-75% quantile), in spite of the out-of-sample density forecasts being misspecified.

Implications

Whether or not the density forecast model should be used is an important question for researchers. As I have shown in this paper, the answer de-

[26] While Inoue's (2001) approach is designed for in-sample density estimators, I apply an out-of-sample version of Inoue's (2001) test to evaluate the out-of-sample predictive densities.

Table 1.6: P-values of alternative tests, GDP+NFCI, recursive

	$h = 1$					
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.228	0.068	0.994	0.970	-	-
out-of-sample	0.052	0.104	0.000	0.000	0.000	0.000
	$h = 4$					
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.128	0.056	0.998	0.998	-	-
out-of-sample	0.198	0.098	0.026	0.000	0.018	0.000

Note: The table reports p-values of the [Inoue, 2001] distributional change tests based on Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics (denoted as Inoue-KS and Inoue-CvM), the [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests based on with Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics respectively (denoted as CS-KS, CS-CvM, RS-KS, and RS-CvM). The in-sample density estimators are constructed using the whole sample (1971Q1 to 2017Q4, $T = 188$), and the out-of-sample predictive densities are constructed using a recursive estimation scheme with $R = 89$.

pends on the regions in question. The DFB test indicates that the model proposed in [Adrian et al., 2019] is a good choice for applications assessing whether the overall in-sample fit is a good indicator of the out-of-sample density forecasting ability. However, studies looking at specific regions of the distribution, e.g., the 25%-50% quantile, may find that the out-of-sample density forecasting ability is not as good as the in-sample fit. Furthermore, my results provide insights relative to which predictors to include in forecasting real GDP growth. Including financial conditions helps narrow the gap between the in-sample fit and the out-of-sample density forecasting ability before the financial crisis. However, care should be taken in choosing predictors thereafter.

1.6.2 S&P 500

Density forecasting is popular not only in macroeocnomics but also in finance, where a large literature has analyzed the uncertainty associated with asset and portfolio returns, with empirical evidence showing that the distributions of stock returns, interest rates and other financial series have non-normal higher moments - see [Tay and Wallis, 2000] for an overview. Density forecasting in finance has roots in the literature on modeling and forecasting stock market volatility, e.g., the ARCH model proposed in [Engle, 1982] and the generalized ARCH proposed (GARCH) model proposed in [Bollerslev, 1986] and [Taylor, 1986].

In this section, I employ the DFB test to evaluate the density forecast models of daily stock price returns. Daily stock price returns are calculated as the percentage change of daily S&P 500 adjusted prices[27]. The data cover 9,581 daily returns, ranging from January 5, 1981 through December 31, 2018. The sample is split such that there are 4,298 in-sample observations (January 5, 1981 - December 31, 1997) and 5,283 out-of-sample observations (January 1, 1998 - December 31, 2018). In what follows, I consider the GARCH(1,1) and GARCH-t(1,1) models[28] for fitting the data and constructing the one-day-ahead density forecasts of the S&P 500 returns, with both the rolling and recursive window schemes.

Model 1: GARCH(1,1)

$$\begin{aligned} y_t &= \sigma_t \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0,1) \\ \sigma_t^2 &= w + a y_{t-1}^2 + b \sigma_{t-1}^2 \end{aligned}$$

At each period t, the MLE is applied with the corresponding in-sample observations. As for forecasting, given the MLE estimators $\hat{w}_t$, $\hat{a}_t$ and $\hat{b}_t$

[27] Downloaded from Yahoo! Finance

[28] Applications of (G)ARCH models on financial data date back to [Engle, 1982], [Bollerslev, 1986], [Taylor, 1986], and have become widespread tools for the analysis of financial data. [Diebold et al., 1997], for instance, studied stock price returns using a GARCH model, selecting an MA(1)-GARCH(1,1) model based on both the Akaike and Schwarz information criteria. More generally, numerous studies show the effectiveness of GARCH type models for forecasting stock market volatility.

at each period t, the one-day-ahead density forecast of y_{t+1} is

$$\hat{\phi}^f_{t+1}(y) = \phi(y|\hat{w}_t, \hat{a}_t, \hat{b}_t) = \frac{1}{\sqrt{2\pi\hat{\sigma}^{f2}_{t+1}}} e^{-\frac{y^2}{2\hat{\sigma}^{f2}_{t+1}}}$$

where $\hat{\sigma}^f_{t+1}$ is constructed using MLE estimators $\hat{w}_t$, $\hat{a}_t$ and $\hat{b}_t$.

Model 2: GARCH-t(1,1)

$$\begin{aligned} y_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim t_{df} \\ \sigma^2_t &= w + a y^2_{t-1} + b\sigma^2_{t-1} \end{aligned}$$

At each period t, the MLE is applied with the corresponding in-sample observations. As for forecasting, given the MLE estimators $\hat{w}_t$, $\hat{a}_t$, $\hat{b}_t$ and $\hat{\nu}$ at each period t, the one-day-ahead density forecast of y_{t+1} is

$$\hat{\phi}^f_{t+1}(y) = \phi(y|\hat{w}_t, \hat{a}_t, \hat{b}_t, \hat{\nu}) = \frac{\Gamma(\frac{\hat{\nu}_t+1}{2})}{\sqrt{(\hat{\nu}_t-2)\hat{\sigma}^{f2}_{t+h}\pi}\Gamma(\frac{\hat{\nu}_t}{2})} \left(1 + \frac{y^2}{(\hat{\nu}_t-2)\hat{\sigma}^{f2}_{t+1}}\right)^{-\frac{\hat{\nu}_t+1}{2}}$$

where $\hat{\sigma}^f_{t+1}$ is constructed using MLE estimators $\hat{w}_t$, $\hat{a}_t$ and $\hat{b}_t$.

Table 1.7 shows the findings of the DFB tests for GARCH(1,1) and GARCH-t(1,1), using the LS and CRPS as evaluation functions. Most of the test results demonstrate that, both the GARCH(1,1) and GARCH-t(1,1) models experience DFBs in their one-day-ahead density forecasts of S&P 500 returns. Thus, the performance of the in-sample density estimators is not informative for the performance of the out-of-sample density forecasts in GARCH(1,1) and GARCH-t(1,1).

Table 1.7: P-values of Density Forecast Breakdown test

	GARCH(1,1)		GARCH-t(1,1)	
	LS	CRPS	LS	CRPS
rec	0.0495	0.0078	0.0027	0.0081
roll	0.1074	0.0008	0.2434	0.0006

Note: The table reports the p-values of the one-sided DFB test, considering different evaluation functions (LS and CRPS) for one-day-ahead ($h = 1$) predictive densities. A small value implies a DFB, and a large value implies a DFI. $T = 9581$, $R = 4298$. The HAC truncation is $T^{1/3}$.

1.7 Conclusion

This paper proposes the DFB test to investigate the relationship between in-sample fit and out-of-sample density forecasting ability. For the parametric case, the test allows for model misspecification and takes into account parameter estimation uncertainty under both hypotheses. For the non-parametric case, I provide conditions under which estimation uncertainty is asymptotically irrelevant. The DFB test is valid under general assumptions. It permits a wide range of estimation procedures (e.g., OLS, GMM, QML, and KDE), different estimation window schemes, as well as a plethora of evaluation functions focusing on either the overall or the specific regions of the distribution. In addition, robust versions of this test are provided for practical use.

The DFB test I introduce in this study makes a number of novel contributions to the literature on density forecast evaluation. (i) Rather than specification testing, I focus on the extent to which the future density forecasting performance is consistent with in-sample fit. (ii) The test captures all the various changes that affect density forecasting performance, including parameter changes and distributional changes. This is important to forecasters, since it also allows the test to be robust to cases involving the effects of different structural breaks that offset one another. (iii)

Moreover, it can be used to evaluate both parametric and non-parametric density forecast models.

The Monte Carlo results indicate that the test has good size property in moderately large samples and has power against changes in mean and variance, as well as shifts in distribution type.

To illustrate the usefulness of this test, I carry out an empirical study that investigates whether there are DFBs in the density forecast models of real GDP growth and daily S&P 500 returns in the US. I find that (i) DFBs occur sporadically in the lower quantiles of the one-quarter-ahead and one-year-ahead predictive conditional densities of real GDP growth in the US, modeled with current financial and economic conditions as well as skewed-t distributed errors; and (ii) DFBs occur in the one-day-ahead predictive densities for S&P 500, using GARCH(1,1) and GARCH-t(1,1) models.

1.8 Proof Appendix

1.8.1 Notation

Let $\cdot^*$ denote the counterparts evaluated at the pseudo parameter θ^*; let $\tilde{\cdot}$ denote the demeaned counterparts. We omit the parameters of a function for simplicity. For example, $\overline{S}_t$, S_{t+h} and SS_{t+h} is used for $\overline{S}_t(\hat{\theta}_t)$, $S_{t+h}(\hat{\theta}_t)$ and $SS_{t+h}(\hat{\theta}_t)$.

Define $D_{t+h} \equiv \frac{\partial S_{t+h}}{\partial \theta} - \frac{\partial \overline{S}_t}{\partial \theta} = \frac{\partial SS_{t+h}}{\partial \theta}$, for $t = R, \ldots, T-h$, which is a $1 \times k$ vector; define $D^*_{t+h} \equiv \frac{\partial SS^*_{t+h}}{\partial \theta}$; $\tilde{D}^*_{t+h} \equiv D^*_{t+h} - E[D^*_{t+h}]$.

1.8.2 Algorithm

Algorithm 1 *Construct the following: (1) a $1 \times T$ vector S^* of in-sample and out-of-sample scorings, with element $S^*_t, t = 1, \ldots, T$, and its counterpart S with element $S_t, t = 1, \ldots, T$:*

$$S^* \equiv \left[\underbrace{S_1^*(\theta^*), \ldots, S_R^*(\theta^*)}_{R}, \underbrace{S_{R+1}(\theta^*), \ldots, S_{R+h-1}(\theta^*)}_{h-1}, \underbrace{S_{R+h}^*(\theta^*), \ldots, S_T^*(\theta^*)}_{P} \right]$$

$$\text{fixed} \quad S \equiv \left[\underbrace{S_1(\hat{\theta}_R), \ldots, S_R(\hat{\theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\theta}_R), \ldots, S_{R+h-1}(\hat{\theta}_R)}_{h-1}, \underbrace{S_{R+h}(\hat{\theta}_R), \ldots, S_T(\hat{\theta}_R)}_{P} \right]$$

$$\text{rolling} \quad S \equiv \left[\underbrace{S_1(\hat{\theta}_R), \ldots, S_R(\hat{\theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\theta}_{R+1}), \ldots, S_{R+h-1}(\hat{\theta}_{R+h-1})}_{h-1}, \underbrace{S_{R+h}(\hat{\theta}_R), \ldots, S_T(}_{P} \right.$$

$$\text{recursive} \quad S \equiv \left[\underbrace{S_1(\hat{\theta}_R), \ldots, S_R(\hat{\theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\theta}_{R+1}), \ldots, S_{R+h-1}(\hat{\theta}_{R+h-1})}_{h-1}, \underbrace{S_{R+h}(\hat{\theta}_R), \ldots, S_T(}_{P} \right.$$

and the corresponding vector $\tilde{S}$ of demeaned scorings, where $\tilde{S}_t \equiv S_t - T^{-1} \sum_{j=1}^{T} S_j$; (2) a $1 \times T$ vector of weights, depending on estimation scheme, with elements $\omega_t^S, t = 1, \ldots, T$:

$$\text{fixed} \quad \omega^S \equiv \left[\underbrace{-\frac{P}{R},\ldots,-\frac{P}{R}}_{R},\underbrace{0,\ldots,0}_{h-1},\underbrace{1,\ldots,1}_{P}\right]$$

$$\text{rolling}(P<R) \quad \omega^S \equiv \left[\underbrace{-\frac{1}{R},\ldots,-\frac{P}{R}}_{P},\underbrace{-\frac{P}{R},\ldots,-\frac{P}{R}}_{R-P},\underbrace{-\frac{P-1}{R},\ldots,-\frac{P-h+1}{R}}_{h-1},\underbrace{1-\frac{P-h}{R},\ldots}_{P-h}\right.$$

$$\text{rolling}(P\geq R) \quad \omega^S \equiv \left[\underbrace{-\frac{1}{R},\ldots,-\frac{R}{R}}_{R},\underbrace{-\frac{R}{R},\ldots,-\frac{R}{R}}_{h-1},\underbrace{0,\ldots,0}_{P-R-h-1},\underbrace{1-\frac{R-1}{R},\ldots,1-\frac{1}{R}}_{R-1},\underbrace{1,\ldots,1}_{h}\right.$$

$$\text{recursive} \quad \omega^S \equiv \left[\underbrace{-a_{R,0},\ldots,-a_{R,0}}_{R},\underbrace{a_{R,1},\ldots,a_{R,h-1}}_{h-1},\underbrace{1-a_{R,h},\ldots,1-a_{R,P-1}}_{P-h},\underbrace{1,\ldots,1}_{h}\right]$$

$$a_{R,j} = \frac{1}{R+j}+\frac{1}{R+j+1}+\cdots+\frac{1}{T-h}$$

so that

$$\sum_{t=1}^{T}\omega_t^S S_t^* = \sum_{t=R}^{T-h} SS_{t+h}^*$$

Algorithm 2 *Construct the following: (1) a $(k+1)\times T$ vector of in-sample and out-of-sample h^*), as defined in Assumption P3, with element $h_t^*, t=1,\ldots,T$, and its counterpart h with element $h_t, t=1,\ldots,T$:*

$$h^* \equiv \left[\underbrace{h_1^*(\theta^*),\ldots,h_R^*(\theta^*)}_{R},\underbrace{h_{R+1}^*(\theta^*),\ldots,h_{T-h}^*(\theta^*)}_{P-1},\underbrace{O,\ldots,O}_{h}\right]$$

$$\text{fixed} \quad h \equiv \left[\underbrace{h_1(\hat{\theta}_R), \ldots, h_R(\hat{\theta}_R)}_{R}, \underbrace{h_{R+1}(\hat{\theta}_R), \ldots, h_{T-h}(\hat{\theta}_R)}_{P-1}, \underbrace{O, \ldots, O}_{h}\right]$$

$$\text{rolling} \quad h \equiv \left[\underbrace{h_1(\hat{\theta}_R), \ldots, h_R(\hat{\theta}_R)}_{R}, \underbrace{h_{R+1}(\hat{\theta}_{R+1}), \ldots, h_{T-h}(\hat{\theta}_{T-h})}_{P-1}, \underbrace{O, \ldots, O}_{h}\right]$$

$$\text{recursive} \quad h \equiv \left[\underbrace{h_1(\hat{\theta}_R), \ldots, h_R(\hat{\theta}_R)}_{R}, \underbrace{h_{R+1}(\hat{\theta}_{R+1}), \ldots, h_{T-h}(\hat{\theta}_{T-h})}_{P-1}, \underbrace{O, \ldots, O}_{h}\right]$$

(2) a $1 \times T$ vector of weights ω^{h*}, depending on estimation scheme, with elements $\omega_t^{h*}, t = 1, \ldots, T$

$$\text{fixed} \quad \omega^{h*} \equiv \left[\underbrace{-\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_R^*}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_R^*}{R}}_{R}, \underbrace{O, \ldots, O}_{T-R}\right]$$

$$\text{rolling}(P < R) \quad \omega^{h*} \equiv \left[\underbrace{-\frac{D_{R+h}^* B_R^*}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_t^*}{R}}_{P}, \underbrace{-\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_t^*}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_t^*}{R}}_{R-P}, \underbrace{-\frac{\sum_{t=R+1}^{T-h} D_{t+h}^* B_t^*}{R}, \ldots, -\frac{D_T^* B}{R}}_{P-1}\right.$$

$$\text{rolling}(P \geq R) \quad \omega^{h*} \equiv \left[\underbrace{-\frac{D_{R+h}^* B_R^*}{R}, \ldots, -\frac{\sum_{t=R}^{2R-1} D_{t+h}^* B_t^*}{R}}_{R}, \underbrace{-\frac{\sum_{t=R}^{2R} D_{t+h}^* B_t^*}{R}, \ldots, -\frac{\sum_{t=P}^{T-h} D_{t+h}^* B_t^*}{R}}_{P-R}, \underbrace{-\frac{\sum_{t=P+1}^{T-h} D_{t+h}^* B_t^*}{R}, \ldots, -\frac{D_T^* B}{R}}_{R-1}\right.$$

$$\text{recursive} \quad \omega^{h*} \equiv \left[\underbrace{b_{R,0}^*, \ldots, b_{R,0}^*}_{R}, \underbrace{b_{R,1}^*, \ldots, b_{R,P-1}^*}_{P-1}, \underbrace{0, \ldots, 0}_{h}\right]$$

$$b_{R,j}^* = \frac{D_{R+h+j}^* B_{R+j}^*}{R+j} + \frac{D_{R+h+j+1}^* B_{R+j+1}^*}{R+j+1} + \cdots + \frac{D_T^* B_{T-h}^*}{T-h}$$

so that

$$\sum_{t=1}^{T} \omega_t^{h*} h_t^* = \sum_{t=R}^{T-h} D_{t+h}^* B_t^* H_t^*$$

and the counterpart ω^h with elements $\omega_t^h, t = 1, \ldots, T$ by replacing $D_\cdot^*, B_\cdot^*$ with $D_\cdot, B_\cdot$ evaluated with estimated parameters.

1.8.3 Lemmas

Lemma 1. *(a) For* $0 \leq a < 0.5$, $P^{-1/2}\sum_{t=R}^{T-h} t^{-1+a} \to 0$*; (b)* $P^{-1/2}\sum_{t=R}^{T-h} t^{-1/2} = O(1)$.

Proof: See [West, 1996] Lemma 1.

Lemma 2. *For* $0 \leq a < 0.5$, *(a)* $\sup_t |t^a H_t^*| \xrightarrow{p} 0$*; (b)* $\sup_t |t^a(\hat{\theta} - \theta^*)| \xrightarrow{p} 0$.

Proof: (a) is shown in [West, 1996] Lemma 3(a). (b) By Assumption 3 with new definition of B_t and H_t^*, it still holds that $\sup_t |B_t - B_t^*| \xrightarrow{p} 0$.

$$\begin{aligned}
\sup_t |t^a(\hat{\theta} - \theta^*)| &\equiv \sup_t |t^a B_t H_t^*| \\
&\leq \sup_t |t^a(B_t - B_t^*)H_t^*| + \sup_t |t^a B_t^* H_t^*| \\
&\leq \sup_t |B_t - B_t^*| \sup_t |t^a H_t^*| + \sup_t |B_t^*| \sup_t |t^a H_t^*| \\
&\xrightarrow{p} 0
\end{aligned}$$

Lemma 3. *(a)* $R_1 \equiv P^{-1/2}\sum_{t=R}^{T-h} \tilde{D}_{t+h}^* B_t^* H_t^* = op(1)$*; (b)* $R_2 \equiv \frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}(\hat{\theta}_t - \theta^*)' \frac{\partial^2 SS_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'}(\hat{\theta}_t - \theta^*) = op(1)$*, where* $\bar{\theta}^*$ *is an intermediate point between* $\hat{\theta}_t$ *and* θ^*.

Proof: (a) is analogous to proof of Lemma 1 (a) in [Giacomini and Rossi, 2009a] by replacing L with S in this context and redefining B, D in this context.

(b) For some a, $0 < a < 0.5$, C a positive constant, m_t defined in Assumption A2(b) and denoting by $\overline{m}_t$ the mean of the m_t's over the

relevant in-sample window at time t, we have

$$
\begin{aligned}
R_2 &= \frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h} t^{1-a}(\hat{\theta}_t - \theta^*)' \left(t^{a-1}\frac{\partial^2 SS_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'} \right)(\hat{\theta}_t - \theta^*) \\
&\leq C \sup_{R\leq t\leq T} |t^{0.5-0.5a}(\hat{\theta}_t - \theta^*)|^2 P^{-1/2}\sum_{t=R}^{T-h} t^{a-1}|\frac{\partial^2 SS_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'}| \\
&\leq C \sup_{R\leq t\leq T} |t^{0.5-0.5a}(\hat{\theta}_t - \theta^*)|^2 P^{-1/2}\sum_{t=R}^{T-h} t^{a-1}\left(|\frac{\partial^2 S_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'}| + |\frac{\partial^2 \overline{S}_t(\bar{\theta}^*)}{\partial\theta\partial\theta'}| \right) \\
&\leq C \sup_{R\leq t\leq T} |t^{0.5-0.5a}(\hat{\theta}_t - \theta^*)|^2 P^{-1/2}\sum_{t=R}^{T-h} t^{a-1}(m_{t+h} + \overline{m}_t) = op(1)
\end{aligned}
$$

using Lemma 1(a) and Lemma 2(b), Assumption 2(b).

Lemma 4. $\frac{T}{P}V_T^{SS*} \equiv var\left(P^{-1/2}\sum_{t=1}^{T}\omega_t^S \tilde{S}_t^*\right) > 0$ *for all* T *sufficiently large.*

Proof: The proof is analogous to proof of Lemma 2 in [Giacomini and Rossi, 2009a] by replacing L with S in this context.

1.8.4 Proof of Theorem 1

Apply a second order mean value expansion on the numerator of $t_{R,P,h}$ in (1.5), we have

$$\begin{aligned}
&\sqrt{P}\left[\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*)]\right] \\
&= P^{-1/2}\sum_{t=R}^{T-h}(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]) + P^{-1/2}\sum_{t=R}^{T-h}\frac{\partial SS_{t+h}(\theta^*)}{\partial\theta}(\hat{\theta}_t - \theta^*) \\
&\quad + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}(\hat{\theta}_t - \theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'}(\hat{\theta}_t - \theta^*) \\
&= P^{-1/2}\sum_{t=R}^{T-h}(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]) + P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[\frac{\partial SS_{t+h}(\theta^*)}{\partial\theta}\right](\hat{\theta}_t - \theta^*) \\
&\quad + P^{-1/2}\sum_{t=R}^{T-h}\left(\frac{\partial SS_{t+h}(\theta^*)}{\partial\theta} - \mathrm{E}\left[\frac{\partial SS_{t+h}(\theta^*)}{\partial\theta}\right]\right)(\hat{\theta}_t - \theta^*) + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}(\hat{\theta}_t - \theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'} \\
&= P^{-1/2}\sum_{t=R}^{T-h}(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]) + P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + P^{-1/2}\sum_{t=R}^{T-h}\tilde{D}^*_{t+h}B^*_t H^*_t \\
&\quad + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}(\hat{\theta}_t - \theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\theta}^*)}{\partial\theta\partial\theta'}(\hat{\theta}_t - \theta^*)
\end{aligned} \tag{1.21}$$

Apply Lemma 3, it holds that

$$\begin{aligned}
&\sqrt{P}\left(\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*)]\right) \\
&= P^{-1/2}\sum_{t=R}^{T-h}(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]) + P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + op(1)
\end{aligned}$$

Given Assumption P5, the term $P^{-1/2}\sum_{t=R}^{T-h} E\left[D^*_{t+h}\right]B^*_t H^*_t$ retains. So it holds that

$$\sqrt{P}\left(\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*)]\right)$$
$$= P^{-1/2}\sum_{t=R}^{T-h}\left(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]\right) + P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + op(1)$$
$$\overset{under H_0}{=} P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*) + P^{-1/2}\sum_{t=R}^{T-h}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + op(1)$$
$$\overset{under H_0}{=} \begin{bmatrix}1 & 1\end{bmatrix} P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*) \\ \sum_{t=R}^{T-h}\mathrm{E}[D^*_{t+h}]B^*_t H^*_t\end{bmatrix} + op(1)$$

Now we show that, under H_0,

$$\left(\frac{T}{P}V_T\right)^{-1/2} P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*) \\ \sum_{t=R}^{T-h}\mathrm{E}[D^*_{t+h}]B^*_t H^*_t\end{bmatrix} \xrightarrow{d} \mathcal{N}(O, I_2)$$

With notations defined above, we can write

$$\left(\frac{T}{P}V_T\right)^{-1/2} P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T-h} SS_{t+h}(\theta^*) \\ \sum_{t=R}^{T-h} E[D^*_{t+h}]B^*_t H^*_t\end{bmatrix} = V_T^{-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega^S_t S^*_t \\ \sum_{t=1}^{T}\omega^{h*}_t h^*_t\end{bmatrix} = V_T^{-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=}^{T} \\ \sum_{t=}^{T}\end{bmatrix}$$

where the last equality holds under H_0 as $T^{-1/2}\sum_{t=1}^{T}\omega^S_t\mathrm{E}[S^*_t] = T^{-1/2}\sum_{t=1}^{T} E[\frac{1}{P}\sum_{t=R}^{R} SS^*_{t+h}] = 0$.

Now we show that,

$$V_T^{*-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega^S_t\tilde{S}^*_t \\ \sum_{t=1}^{T}\omega^{h*}_t h^*_t\end{bmatrix} \xrightarrow{d} \mathcal{N}(O, I_2), \qquad V^*_T = var\left(T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega^S_t\tilde{S}^*_t \\ \sum_{t=1}^{T}\omega^{h*}_t h^*_t\end{bmatrix}\right)$$

We verify that the zero-mean vector sequence $\left\{[V_T^{*-1/2}\omega^S_t\tilde{S}^*_t, V_T^{*-1/2}\omega^{h*}_t h^*_t]'\right\}_{t=1}^{T}$ satisfies the conditions of [Wooldridge and White, 1988] Central Limit Theorem for mixing processes (also see Theorem 5.20 in [White, 2001]).

- $g_t = [V_T^{*-1/2}\omega_t^S \tilde{S}_t^*, V_T^{*-1/2}\omega_t^{h*} h_t^*]'$ is a function of only a finite number of leads and lags of Z_t, it follows from Lemma 2.1 of [White and Domowitz, 1984] that it is mixing of the same size as Z_t, which is mixing with α of size $-r/(r-2), r > 2$.

- By Assumption P4, we have $|\tilde{S}_t^*|^{2r} < \infty$. And $|\omega_t^S| < \infty$ for all t.(It holds for the fixed and rolling schemes by Assumption P7; it holds for recursive scheme by the fact that $a_{m,j} < a_{m,0} \to ln(1+\pi) < \infty$.) Since V_T^* is p.d., it holds that $E|V_T^{*-1/2}\omega_t^S \tilde{S}_t^*|^{2r} < \infty$.

- By defining $\lambda_t = V_T^{*-1/2}\omega_t^{h*} = V_T^{*-1/2}\frac{1}{T}\sum_{j=1}^T \mathrm{E}[\partial S_j^*]P_{t,j}$, where $P_{t,j}$ is defined in the proof Lemma 1, and by the fact that $|\lambda_{t,i}| < \infty$ for all t, i (using Assumption P5, $P_{t,j}$ having bounded component shown in the proof of Lemma 1, and that V_T^* is p.d.), we can write

$$E|V_T^{*-1/2}\omega_t^{h*}h_t^*|^{2r} = E|\lambda_t h_t^*|^{2r} = E|\sum_{i=1}^q \lambda_{t,i}h_{i,t}^*|^{2r} \leq \left[\sum_{i=1}^q |\lambda_{t,i}|(|E|h_{i,t}^*|^{2r})^{1/2r}\right]^{2r} < \infty$$

  where the first inequality uses Minkowski's inequality and the second inequality uses Assumption P4.

This implies that we can apply [Wooldridge and White, 1988] Central Limit Theorem for mixing processes on the zero-mean vector sequence $\left\{[V_T^{*-1/2}\omega_t^S \tilde{S}_t^*, V_T^{*-1/2}\omega_t^{h*}h_t^*]'\right\}_{t=1}^T$, and get

$$V_T^{*-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^T \omega_t^S \tilde{S}_t^* \\ \sum_{t=1}^T \omega_t^{h*}h_t^*\end{bmatrix} \xrightarrow{d} \mathcal{N}(O, I_2), \qquad V_T^* = var\left(T^{-1/2}\begin{bmatrix}\sum_{t=1}^T \omega_t^S \tilde{S}_t^* \\ \sum_{t=1}^T \omega_t^{h*}h_t^*\end{bmatrix}\right)$$

The result then follows from the fact that $V_T - V_T^* \xrightarrow{p} 0$, owing to consistency of $\hat{\theta}_t$ for θ^* under H_0.

1.8.5 Proof of Corollary 1

By using LS and QMLE, it is trivially satisfied that $\mathrm{E}[\frac{\partial S_t^*}{\partial \theta}] = 0$ for all t, as $S(\cdot)$ takes the form of a logarithm of the density estimator/forecasts

evaluated at the realization. Thus, it holds that

$$\mathrm{E}[D^*_{t+h}] = \mathrm{E}[\frac{\partial SS^*_{t+h}}{\partial\theta}] = \mathrm{E}[\frac{\partial S^*_{t+h}}{\partial\theta}] - \mathrm{E}[\frac{\partial \overline{S}^*_t}{\partial\theta}] = 0$$

The result then follows from reasonings analogous to those in the proof of Theorem 1.

1.8.6 Proof of Theorem 2

$$\begin{aligned}
&\sqrt{P}\left\{\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\phi}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h}(\phi^*)]\right\} \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h}(SS_{t+h}(\phi^*) - \mathrm{E}[SS_{t+h}(\phi^*)]) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h}\left(SS_{t+h}(\hat{\phi}_t) - SS_{t+h}(\phi^*)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h}(SS_{t+h}(\phi^*) - \mathrm{E}[SS_{t+h}(\phi^*)]) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h}\left(S_{t+h}(\hat{\phi}_t) - S_{t+h}(\phi^*) + \frac{1}{R}\sum_{j=t-R+h+1}^{t} S_j\right. \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h}(SS_{t+h}(\phi^*) - \mathrm{E}[SS_{t+h}(\phi^*)]) \\
&\quad + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} D^*_\phi(y_{t+h})\left(\hat{\phi}_t(y_t) - \phi^*(y_t)\right) + \frac{1}{R}\sum_{j=t-R+h+1}^{t} D^*_\phi(y_j)\left(\hat{\phi}_t(y_j) - \phi^*(y_j)\right)
\end{aligned} \tag{1.22}$$

The proof of the asymptotic distribution of the first component follows from reasonings analogous to those in the proof of Theorem 1. As for the second component, here we show that the second component is asymptotically irrelevant in the case of kernel density estimation.

Proof of Example (KDE)

Given Assumption NP-KDE1 to NP-KDE3, according to Theorem 1.4 in [Li and Racine, 2007], it holds that

$$\sup_y |\hat{\phi}_t(y) - \phi(y)| = O(\sqrt{\frac{\ln R}{R h_R}} + h_R^2) \qquad \forall t$$

See [Li and Racine, 2007] for detailed proof.

Given Assumption NP-KDE5, $D^*_\phi(.)$ is bounded.

Thus, it then holds that

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} D^*_\phi(y_{t+h})\left(\hat{\phi}_t(y_t)-\phi^*(y_t)\right)+\frac{1}{R}\sum_{j=t-R+h+1}^{t} D^*_\phi(y_j)\left(\hat{\phi}_t(y_j)-\phi^*(y_j)\right)\\
&\leq \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} |D^*_\phi(y_{t+h})|\left|\left(\hat{\phi}_t(y_t)-\phi^*(y_t)\right)\right|+\frac{1}{R}\sum_{j=t-R+h+1}^{t} |D^*_\phi(y_j)|\left|\left(\hat{\phi}_t(y_j)-\phi^*(y_j)\right)\right|\\
&\leq \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} \sup_y|D^*_\phi(y)|\sup_y|\hat{\phi}_t(y)-\phi(y)|+\frac{1}{R}\sum_{j=t-R+h+1}^{t} \sup_y|D^*_\phi(y)|\sup_y|\hat{\phi}_t(y)-\phi(y)|\\
&\leq \frac{1}{\sqrt{P}}P\sup_y|D^*_\phi(y)|\max_t\sup_y|\hat{\phi}_t(y)-\phi(y)|2\\
&=\sqrt{P}O(\sqrt{\frac{\ln R}{Rh_R}}+h_R^2)
\end{aligned}$$

Therefore, given Assumption NP-KDE4, $\frac{P\ln R}{Rh_R}\to 0$ and $Ph_R^4\to 0$ imply $\sqrt{P}O(\sqrt{\frac{\ln R}{Rh_R}}+h_R^2)=op(1)$.

1.8.7 Proof of Proposition 1 and 2

A mean value expansion of $P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}$ around θ_t^* is

$$\begin{aligned}
P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_t) &= P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\theta_t^*)+P^{-1/2}\sum_{t=R}^{T-h}\frac{\partial SS_{t+h}(\theta_t^*)}{\partial\theta}\left(\hat{\theta}_t-\theta_t^*\right)\\
&\quad+\frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}\left(\hat{\theta}_t-\theta_t^*\right)'\left(\frac{\partial^2 SS_{t+h}(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right)\left(\hat{\theta}_t-\theta_t^*\right)
\end{aligned}$$

where $\bar{\hat{\theta}}_t$ denotes some intermediate point between the interval $(\hat{\theta}_t,\theta_t^*)$.

The first term can be expressed as

$$P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\theta_t^*)=P^{-1/2}\sum_{t=R}^{T-h} S_{t+h}(\theta_t^*)-\overline{S}_t(\theta_t^*)=P^{-1/2}\sum_{t=R}^{T-h} S_{t+h}(\theta_t^*)-\overline{\sum}_j S_j(\theta_t^*)$$

The mean value expansions of $S_{t+h}(\theta_t^*)$ and $S_j(\theta_t^*)$ around θ_{t+h}^* and θ_j^* respectively are

$$S_{t+h}(\theta_t^*) = S_{t+h}(\theta_{t+h}^*) + \frac{\partial S_{t+h}(\theta_{t+h}^*)}{\partial \theta}\left(\theta_t^* - \theta_{t+h}^*\right) + \frac{1}{2}\left(\theta_t^* - \theta_{t+h}^*\right)' \left(\frac{\partial^2 S_{t+h}(\bar{\theta}_{t+h}^*)}{\partial \theta \partial \theta'}\right)\left(\theta_t^* - \theta_t^*\right)$$

$$S_j(\theta_t^*) = S_j(\theta_j^*) + \frac{\partial S_j(\theta_j^*)}{\partial \theta}\left(\theta_t^* - \theta_j^*\right) + \frac{1}{2}\left(\theta_t^* - \theta_j^*\right)' \left(\frac{\partial^2 S_j(\bar{\theta}_j^*)}{\partial \theta \partial \theta'}\right)\left(\theta_t^* - \theta_j^*\right)$$

where $(\bar{\theta}_{t+h}^*, \bar{\theta}_j^*)$ denote some intermediate points between the interval $(\theta_t^*, \theta_{t+h}^*), (\theta_t^*, \theta_j^*)$respectively.

Thus, it follows that

$$\begin{aligned} SS_{t+h}(\theta_t^*) &= S_{t+h}(\theta_t^*) - \overline{\sum}_j S_j(\theta_t^*) \\ &+ \frac{\partial S_{t+h}(\theta_{t+h}^*)}{\partial \theta}\left(\theta_t^* - \theta_{t+h}^*\right) - \overline{\sum}_j \frac{\partial S_j(\theta_j^*)}{\partial \theta}\left(\theta_t^* - \theta_j^*\right) \\ &+ \frac{1}{2}\left(\theta_t^* - \theta_{t+h}^*\right)' \left(\frac{\partial^2 S_{t+h}(\bar{\theta}_{t+h}^*)}{\partial \theta \partial \theta'}\right)\left(\theta_t^* - \theta_{t+h}^*\right) - \overline{\sum}_j \frac{1}{2}\left(\theta_t^* - \theta_j^*\right)' \left(\frac{\partial^2 S_j(\bar{\theta}_j^*)}{\partial \theta \partial \theta'}\right)\left(\theta_t^*\right. \end{aligned}$$

Then, it holds that

$$\begin{aligned} P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_t) &= P^{-1/2}\sum_{t=R}^{T-h} S_{t+h}(\theta_t^*) - \overline{\sum}_j S_j(\theta_t^*) \\ &+ P^{-1/2}\sum_{t=R}^{T-h} \frac{\partial S_{t+h}(\theta_{t+h}^*)}{\partial \theta}\left(\theta_t^* - \theta_{t+h}^*\right) - \overline{\sum}_j \frac{\partial S_j(\theta_j^*)}{\partial \theta}\left(\theta_t^* - \theta_j^*\right) \\ &+ P^{-1/2}\sum_{t=R}^{T-h} \frac{1}{2}\left(\theta_t^* - \theta_{t+h}^*\right)' \left(\frac{\partial^2 S_{t+h}(\bar{\theta}_{t+h}^*)}{\partial \theta \partial \theta'}\right)\left(\theta_t^* - \theta_{t+h}^*\right) \\ &- \overline{\sum}_j \frac{1}{2}\left(\theta_t^* - \theta_j^*\right)' \left(\frac{\partial^2 S_j(\bar{\theta}_j^*)}{\partial \theta \partial \theta'}\right)\left(\theta_t^* - \theta_j^*\right) \\ &+ P^{-1/2}\sum_{t=R}^{T-h} \frac{\partial SS_{t+h}(\theta_t^*)}{\partial \theta}\left(\hat{\theta}_t - \theta_t^*\right) + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T-h}\left(\hat{\theta}_t - \theta_t^*\right)' \left(\frac{\partial^2 SS_{t+h}(}{\partial \theta \partial \theta'}\right. \end{aligned}$$

By taking expectation, Proposition 1 is obtained. Besides, "Estimation Uncertainty II" component is obtained by adding in the mean square expansion of the term $\frac{\partial \overline{L}_t(\hat{\theta})}{\partial \theta} = 0$ around θ_t^*, that is, $\frac{\partial \overline{L}_t(\hat{\theta})}{\partial \theta} = \frac{\partial \overline{L}_t(\theta_t^*)}{\partial \theta} + \frac{\partial^2 \overline{L}_t(\bar{\hat{\theta}}_t)}{\partial \theta \partial \theta'} \left(\hat{\theta}_t - \theta_t^*\right) = 0$.

The proof of Proposition 2 follows from reasonings analogous to those in the proof of Proposition 1.

Corollary 2. ***(Special case: DFB test statistic with overfitting correction under LS and (Q)MLE)***
In this case, the DFB test statistic is modified as:

$$t^c_{R,P,h} = \frac{P^{1/2}\overline{SS}_{R,P} - c}{\hat{\sigma}_{R,P}}, \qquad c = -\gamma k$$

where k is the number of parameters; $\gamma = \frac{\sqrt{P}}{R}$ for fixed and rolling window schemes, and $\gamma = P^{-1/2}\ln(1 + P/R)$ for recursive window scheme.

Here a proof for the rolling window scheme is provided:

Since QMLE and LS are adopted together, then it holds that $\overline{S}_t(\theta_t^*) = \overline{\sum}_j \log \phi(y_j|\theta_t^*) = \overline{L}_t(\theta_t^*)$. Suppose $\theta_t^* = \theta^*$, then "estimation uncertainty II" degenerates to $\sqrt{P}\overline{\sum}_t \mathrm{E}\left(\hat{\theta}_t - \theta_t^*\right)' \frac{\partial^2 \overline{L}_t(\bar{\hat{\theta}}_t)}{\partial \theta \partial \theta'} \left(\hat{\theta}_t - \theta_t^*\right)$.

Apply Mean Value Expansion on $\frac{\partial \overline{L}_t(\bar{\hat{\theta}}_t)}{\partial \theta}$:

$$0 = \frac{\partial \overline{L}_t(\bar{\hat{\theta}}_t)}{\partial \theta} = \frac{\partial \overline{L}_t(\bar{\theta}^*)}{\partial \theta} + \frac{\partial^2 \overline{L}_t(\bar{\hat{\theta}}_t)}{\partial \theta \partial \theta'} \left(\hat{\theta}_t - \theta_t^*\right)$$

where the first equality holds according to (Q)MLE.

Thus it holds that

$$\left(\hat{\theta}_t - \theta_t^*\right) = \left[-\frac{\partial^2 \overline{L}_t(\bar{\hat{\theta}}_t)}{\partial \theta \partial \theta'}\right]^{-1} \frac{\partial \overline{L}_t(\bar{\theta}^*)}{\partial \theta}$$

so

$$\sqrt{R}\left(\hat{\theta}_t - \theta_t^*\right) = \left[-\frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right]^{-1} \sqrt{R}\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}$$

Thus,

$$\begin{aligned}
&\sqrt{P}\overline{\sum}_t \mathrm{E}\left(\hat{\theta}_t - \theta_t^*\right)' \frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\left(\hat{\theta}_t - \theta_t^*\right) \\
=&\sqrt{P}\overline{\sum}_t \mathrm{E}\left(\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}\right)' \left[-\frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right]^{-1} \frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'} \left[-\frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right]^{-1} \left(\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}\right) \\
=&-\frac{\sqrt{P}}{R}\overline{\sum}_t \mathrm{E}\left(\sqrt{R}\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}\right)' \left[-\frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right]^{-1} \left(\sqrt{R}\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}\right) \\
\approx&-\frac{\sqrt{P}}{R}k
\end{aligned}$$

as $\overline{\sum}_t \mathrm{E}\left(\sqrt{R}\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}\right)' \left[-\frac{\partial^2 \bar{L}_t(\bar{\hat{\theta}}_t)}{\partial\theta\partial\theta'}\right]^{-1} \left(\sqrt{R}\frac{\partial \bar{L}_t(\bar{\theta}^*)}{\partial\theta}\right) \sim \chi^2_k$.

1.8.8 Proof of Proposition 3

$$\begin{aligned}
\sigma^{-1}M^{-1/2}\sum_{j=t-M/2}^{t+M/2-1} SS_j &= \sigma^{-1}M^{-1/2}\sum_{j=R+h}^{t+M/2-1} SS_j - \sigma^{-1}M^{-1/2}\sum_{j=R+h}^{t-M/2-1} SS_j \\
&= (M/P)^{-1/2}\left(\sigma^{-1}P^{-1/2}\sum_{j=R+h}^{t+M/2-1} SS_j - \sigma^{-1}P^{-1/2}\sum_{j=R+h}^{t-M/2-1} SS_j\right)
\end{aligned}$$

By the conditions (i)(ii) in Proposition 3, together with reasonings analogous to those in the proof of Theorem 1, it holds that

$$\mathcal{F}_{t,R,M} \Rightarrow [\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu}$$

The result then follows from the fact that $\hat{\sigma}^2_{R,P}$ is a consistent estimator of σ^2 under the null hypothesis.

1.8.9 Proof of Proposition 4

This high-level assumption of condition (1.14) can be shown to hold for the existing DFB test of interest under Assumptions 1-7. Because the existing test had already imposed assumptions for the FCLT to take into account recursive, rolling, and fixed estimation schemes and because weak convergence to stochastic integrals can hold for partial sums. Now we verify that condition (1.14) is satisfied by the DFB test, provided that we assume that the relevant variance estimate is uniformly consistent over all window sizes, that is:

$$\sup_{\underline{R}\leq R\leq \overline{R}} |\hat{\sigma}^2_{R,P} - \sigma^2| = op(1)$$

We follow the proofs of [Inoue and Rossi, 2012], [West, 1996], [West and McCracken, 1998], and [Clark and McCracken, 2001] very closely and extend their results to weak convergence in the space of functions on $[\underline{\mu}, \bar{\mu}]$. By replacing the notations f in [West, 1996] with the notations $\bar{SS}$ in our setting, as well as the new coefficient parameter θ and the new B, D, h in our setting, Assumptions we adopted are identical to Assumptions 1-4 in [West, 1996], with Assumption P7 holding for all values of $R = [\underline{\mu}T], [\underline{\mu}T] + 1, \ldots, [\bar{\mu}T]$. It is relatively straightforward to show that lemmas A1-A6 of [West, 1996] hold under these assumptions, with the $\sup_t$ replaced by $\sup_{[\underline{\mu}T]\leq R\leq [\bar{\mu}T]} \sup_{R\leq t\leq T} \equiv \sup_{[\underline{\mu}T]\leq t\leq T}$.

Thus, the key is to prove our version of lemma 4.1 in [West, 1996], and it suffices to verify that assumptions C1, C2, and A5 in corollary 3.1 in Wooldridge and White (1988) are satisfied by the counterpart of the statistics in our setting.

Step 1: We prove our version of lemma 4.1 in [West, 1996]. We have

shown in proof of Theorem 1 that

$$\begin{aligned}
&\sqrt{P}\left(\frac{1}{P}\sum_{t=R}^{T} SS_{t+h}(\hat{\theta}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T} SS_{t+h}(\theta^*)]\right) \\
&= P^{-1/2}\sum_{t=R}^{T}\left(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]\right) + P^{-1/2}\sum_{t=R}^{T}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + op(1) \\
&= \begin{bmatrix}1 & 1\end{bmatrix} P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T} SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)] \\ \sum_{t=R}^{T}\mathrm{E}[D^*_{t+h}]B^*_t H^*_t\end{bmatrix} + op(1) \\
&\overset{H_0}{=} \begin{bmatrix}1 & 1\end{bmatrix} P^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega^S_t \tilde{S}^*_t \\ \sum_{t=1}^{T}\omega^{h*}_t h^*_t\end{bmatrix} + op(1)
\end{aligned}$$

and that

$$\left(\frac{T}{P}V^*_T\right)^{-1/2} P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T} SS_{t+h}(\theta^*) \\ \sum_{t=R}^{T} E[D^*_{t+h}]B^*_t H^*_t\end{bmatrix} = V^{*-1/2}_T T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega^S_t \tilde{S}^*_t \\ \sum_{t=1}^{T}\omega^{h*}_t h^*_t\end{bmatrix} \xrightarrow{d} \mathcal{N}(O, I_2)$$

$$V^*_T = \mathrm{Var}\left(T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega^S_t \tilde{S}^*_t \\ \sum_{t=1}^{T}\omega^{h*}_t h^*_t\end{bmatrix}\right) = \begin{bmatrix}V^{*SS}_T & V^{*Sh}_T \\ V^{*Sh}_T & V^{*hh}_T\end{bmatrix}$$

This result corresponds to lemma 4.1 in [West, 1996].

Step 2: Now we verify that assumptions C1, C2, and A5 in corollary 3.1 in Wooldridge and White (1988) are satisfied. We focus on $\sum_{t=R}^{T}\mathrm{E}[D^*_{t+h}]B^*_t H^*_t$, or say $\sum_{t=1}^{T}\omega^{h*}_t h^*_t$, because we have already assumed that SS_t satisfies these assumptions.

1. Assumption C1 is a nominal assumption and is trivially satisfied for the test statistics with rolling, recursive, and split-sample estimation techniques.

2. We show that assumption C2 is satisfied for $\omega^{h*}_t h^*_t$.

- Assumption C2(i) is satisfied as ω^{h*}_t and h^*_t are bounded in all schemes. ω^{h*}_t is bounded because our Assumption P3 and P4 bound B^* and D^*. h^*_t is bounded in Assumption P4.

- Because $\{\omega_t^{h*}\}$ is a sequence of deterministic bounded values as B^* and D^* are deterministic and bounded, and h_t^* is strong mixing by Assumption P1, $\omega_t^{h*}h_t^*$ is near-epoch dependent, as required by Assumptions C2(ii)(iii).
- Assumptions C2(iii) is satisfied according to our Assumption P1.
- Assumption C2(iv) is satisfied because of our Assumption on h_t^* and the boundedness of ω_t^{h*}.

3. Finally, assumption A5 is satisfied since

$$\mathrm{Var}\left(T^{-1/2}\sum_{t=R}^{[sT]}\mathrm{E}[D_{t+h}^*]B_t^*H_t^*\right) = \mathrm{Var}\left(s^{1/2}[sT]^{-1/2}\sum_{t=R}^{[sT]}\mathrm{E}[D_{t+h}^*]B_t^*H_t^*\right) \to (s-\mu)V_T^{*hh}$$

$$\mathrm{Var}\left(T^{-1/2}\sum_{t=R}^{[sT]}\mathrm{E}[D_{t+h}^*]B_t^*H_t^*\right) = \mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{[sT]}\omega_t^{h*}h_t^*\right) = \mathrm{Var}\left(s^{1/2}[sT]^{-1/2}\sum_{t=1}^{[sT]}\omega_t^{h*}h_t^*\right) \to s$$

Thus according to corollary 3.1 in Wooldridge and White (1988),

$$\begin{aligned}
&\frac{1}{\sigma}T^{-1/2}\sum_{t=1}^{[sT]}\left(SS_{t+h}(\hat{\theta}_t) - \mathrm{E}[SS_{t+h}(\theta^*)]\right) \\
&= \frac{1}{\sigma}\left(T^{-1/2}\sum_{t=1}^{[sT]}\left(SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)]\right) + T^{-1/2}\sum_{t=1}^{[sT]}\mathrm{E}\left[D_{t+h}^*\right]B_t^*H_t^*\right) + op(1) \\
&= \frac{1}{\sigma}\begin{bmatrix}1 & 1\end{bmatrix}T^{-1/2}\begin{bmatrix}\sum_{t=1}^{[sT]}SS_{t+h}(\theta^*) - \mathrm{E}[SS_{t+h}(\theta^*)] \\ \sum_{t=1}^{[sT]}\mathrm{E}[D_{t+h}^*]B_t^*H_t^*\end{bmatrix} + op(1) \\
&\xrightarrow{d} B(s)
\end{aligned}$$

converge to $B(s)$ uniformly in s, where σ^2 is the variance derived in Theorem 1.

Thus,

$$\begin{aligned}\mathcal{T}_T(R) &= \frac{P^{-1/2}\sum_{t=R}^{T-h} SS_{t+h}(\hat{\theta}_{t,R})}{\hat{\sigma}_R} \\ &= (\frac{P}{T})^{-1/2}\frac{1}{\hat{\sigma}_R}T^{-1/2}\sum_{t=1}^{T-h} SS_{t+h}(\hat{\theta}_{t,R}) - (\frac{P}{T})^{-1/2}\frac{1}{\hat{\sigma}_R}T^{-1/2}\sum_{t=1}^{R-1} SS_{t+h}(\hat{\theta}_{t,R}) \\ &\xrightarrow{d} (1-\mu)^{-1/2}[B(1)-B(\mu)]\end{aligned}$$

Combined with the assumption of uniform convergence $\sup_{\underline{R}\leq R\leq \overline{R}} |\hat{\sigma}^2_{R,P} - \sigma^2| = op(1)$, the sketch of the proof is completed.

1.9 Table and Figure Appendix

Table 1.8: P-values of Density Forecast Breakdown test, GDP+NFCI, recursive

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	0.961	0.884	0.248	0.003	0.637	0.954
$h = 4$	0.688	0.803	0.124	0.005	0.341	0.986

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a recursive estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 188$, $R = 89$. The HAC truncation is $T^{1/3}$.

Table 1.9: P-values of Density Forecast Breakdown test, GDP+NFCI, rolling

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	0.881	0.867	0.082	0.005	0.740	0.985
$h = 4$	0.070	0.636	0.045	0.008	0.220	0.991

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a rolling estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 188$, $R = 89$. The HAC truncation is $T^{1/3}$.

Table 1.10: P-values of Density Forecast Breakdown test, GDPonly, recursive

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	1.000	1.000	0.925	0.000	0.002	0.997
$h = 4$	1.000	1.000	0.472	0.002	0.003	1.000

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a recursive estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 280$, $R = 144$. The HAC truncation is $T^{1/3}$.

Table 1.11: P-values of Density Forecast Breakdown test, GDPonly, rolling

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	1.000	1.000	0.866	0.000	0.004	1.000
$h = 4$	0.999	1.000	0.409	0.003	0.007	0.999

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a rolling estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 280$, $R = 144$. The HAC truncation is $T^{1/3}$.

Table 1.12: P-values of Density Forecast Breakdown test, GDPonly, short R, recursive

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	1.000	1.000	0.821	0.001	0.026	0.989
$h = 4$	0.998	0.914	0.208	0.037	0.007	0.984

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a recursive estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 280$, $R = 92$. The HAC truncation is $T^{1/3}$.

Table 1.13: P-values of Density Forecast Breakdown test, GDPonly, short R, rolling

	LS	CRPS	CI			
			Q0:Q0.25	Q0.25:Q0.5	Q0.5:Q0.75	Q0.75:Q1
$h = 1$	0.998	1.000	0.474	0.007	0.306	0.998
$h = 4$	0.999	0.999	0.084	0.122	0.293	0.999

Note: The table reports p-values of the one-sided DFB test, considering different evaluation functions (LS, CRPS, and CI), different forecasting horizons ($h = 1, 4$ quarters), and a rolling estimation scheme. A small p-value implies a DFB, and a large p-value implies a DFI. For CI, four different intervals are considered. For example, $Q0.25 : Q0.5$ refers to that $(\underline{u}, \overline{u}) = (quantile(y_t, 0.25), quantile(y_t, 0.5))$. $T = 280$, $R = 92$. The HAC truncation is $T^{1/3}$.

Table 1.14: P-values of alternative tests, GDP+NFCI, recursive

	$h = 1$					
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.228	0.068	0.994	0.970	-	-
out-of-sample	0.052	0.104	0.000	0.000	0.000	0.000
	$h = 4$					
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.128	0.056	0.998	0.998	-	-
out-of-sample	0.198	0.098	0.026	0.000	0.018	0.000

Note: The table reports p-values of the [Inoue, 2001] distributional change tests based on Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics (denoted as Inoue-KS and Inoue-CvM), the [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests based on with Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics respectively (denoted as CS-KS, CS-CvM, RS-KS, and RS-CvM). The in-sample density estimators are constructed using the whole sample (1971Q1 to 2017Q4, $T = 188$), and the out-of-sample predictive densities are constructed using a recursive estimation scheme with $R = 89$.

Table 1.15: P-values of alternative tests, GDP+NFCI, rolling

	$h = 1$					
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.228	0.068	0.994	0.970	-	-
out-of-sample	0.128	0.110	0.002	0.000	0.002	0.000
	$h = 4$					
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.128	0.056	0.998	0.998	-	-
out-of-sample	0.180	0.090	0.032	0.008	0.002	0.006

Note: The table reports p-values of the [Inoue, 2001] distributional change tests based on Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics (denoted as Inoue-KS and Inoue-CvM), the [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests based on with Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics respectively (denoted as CS-KS, CS-CvM, RS-KS, and RS-CvM). The in-sample density estimators are constructed using the whole sample (1971Q1 to 2017Q4, $T = 188$), and the out-of-sample predictive densities are constructed using a rolling estimation scheme with $R = 89$.

Table 1.16: P-values of alternative tests, GDP only, recursive

	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
			$h = 1$			
in-sample	0.170	0.066	0.996	0.976	-	-
out-of-sample	0.118	0.092	0.000	0.000	0.000	0.000
			$h = 4$			
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.202	0.056	0.998	0.998	-	-
out-of-sample	0.100	0.056	0.012	0.010	0.012	0.008

Note: The table reports p-values of the [Inoue, 2001] distributional change tests based on Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics (denoted as Inoue-KS and Inoue-CvM), the [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests based on with Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics respectively (denoted as CS-KS, CS-CvM, RS-KS, and RS-CvM). The in-sample density estimators are constructed using the whole sample (1948Q1 to 2017Q4, $T = 280$), and the out-of-sample predictive densities are constructed using a recursive estimation scheme with $R = 144$.

Table 1.17: P-values of alternative tests, GDP only, rolling

	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
			$h = 1$			
in-sample	0.170	0.066	0.996	0.976	-	-
out-of-sample	0.194	0.106	0.006	0.004	0.004	0.004
			$h = 4$			
	Inoue-KS	Inoue-CvM	CS-KS	CS-CvM	RS-KS	RS-CvM
in-sample	0.202	0.056	0.998	0.998	-	-
out-of-sample	0.130	0.007	0.002	0.002	0.002	0.002

Note: The table reports p-values of the [Inoue, 2001] distributional change tests based on Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics (denoted as Inoue-KS and Inoue-CvM), the [Corradi and Swanson, 2006] and the [Rossi and Sekhposyan, 2013] correct specification tests based on with Kolmogorov-Smirnov-type and Cramer-von Mises-type statistics respectively (denoted as CS-KS, CS-CvM, RS-KS, and RS-CvM). The in-sample density estimators are constructed using the whole sample (1948Q1 to 2017Q4, $T = 280$), and the out-of-sample predictive densities are constructed using a rolling estimation scheme with $R = 144$.

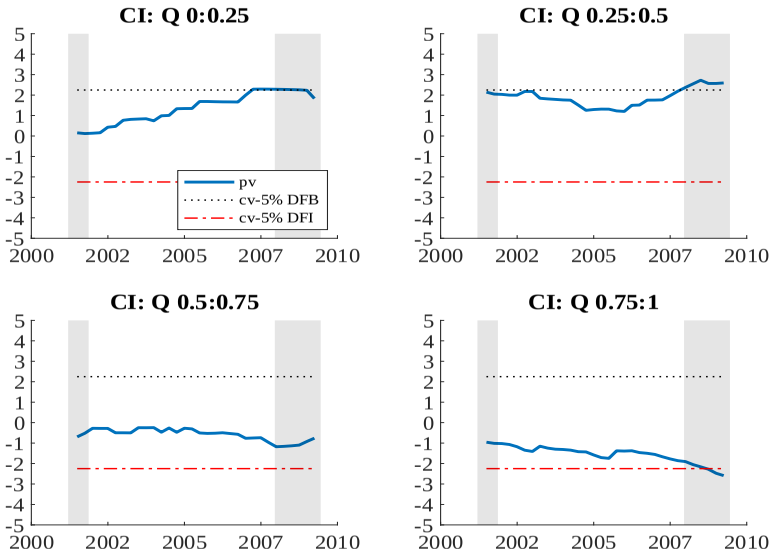


Figure 1.9: $F_{R,M,t}$ statistics, GDP+NFCI, $h = 1$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 188$, $R = 89$, $\frac{M}{P} = 0.7$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

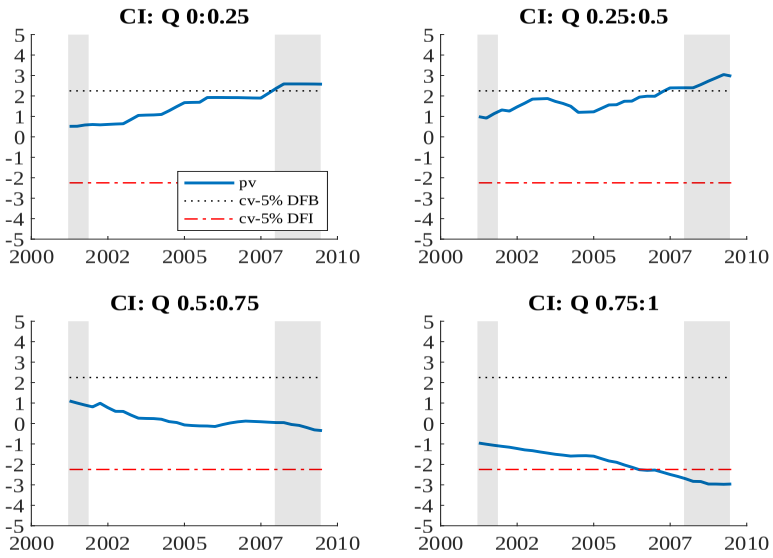


Figure 1.10: $F_{R,M,t}$ statistics, GDP+NFCI, $h = 4$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 188$, $R = 89$, $\frac{M}{P} = 0.7$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

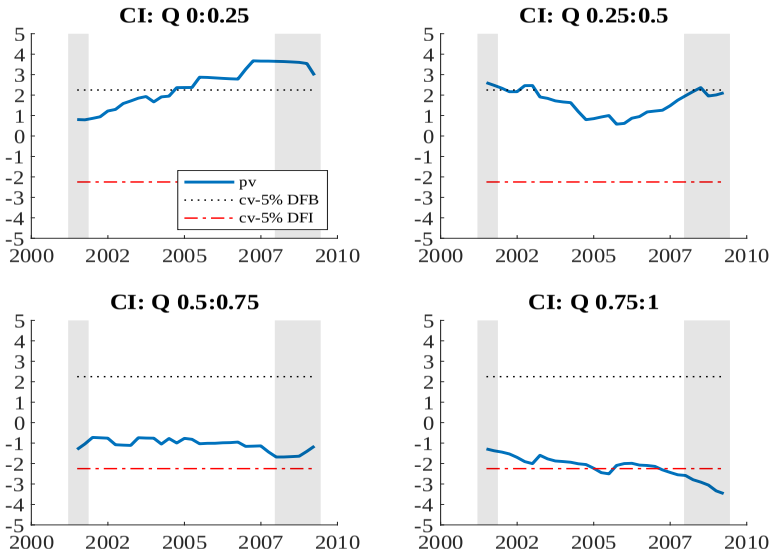


Figure 1.11: $F_{R,M,t}$ statistics, GDP+NFCI, $h = 1$, rolling

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 188$, $R = 89$, $\frac{M}{P} = 0.7$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

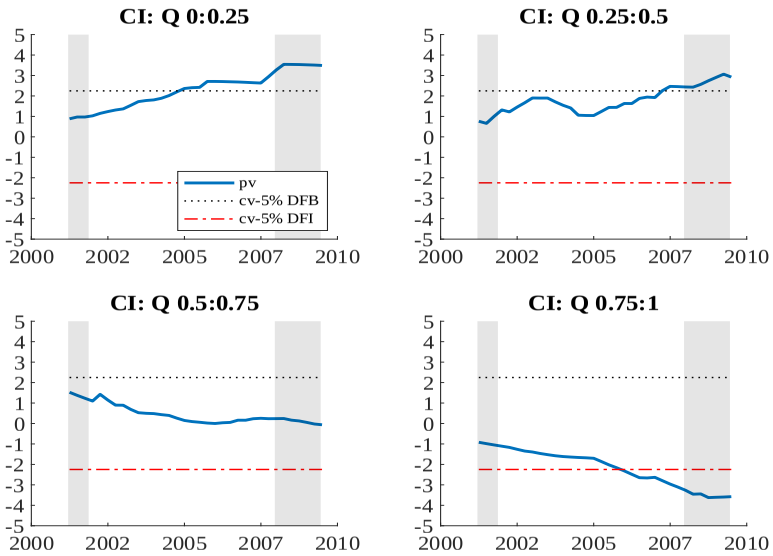


Figure 1.12: $F_{R,M,t}$ statistics, GDP+NFCI, $h = 4$, rolling

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 188$, $R = 89$, $\frac{M}{P} = 0.7$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

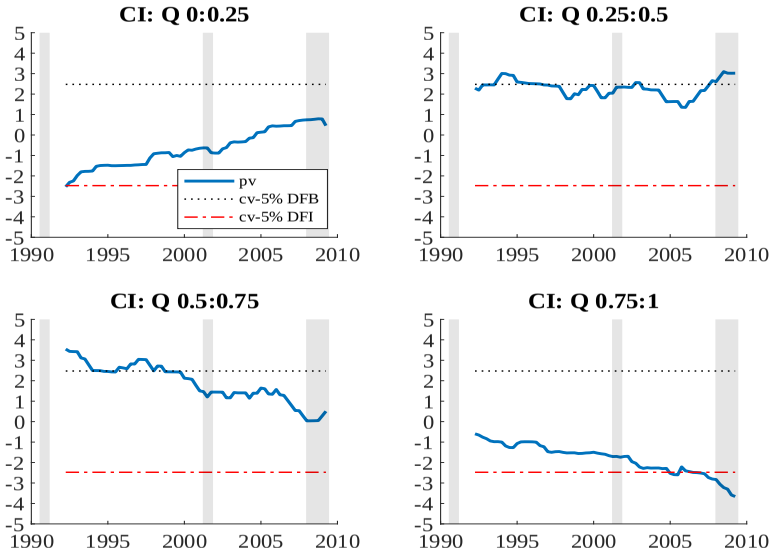


Figure 1.13: $F_{R,M,t}$ statistics, GDPonly, $h = 1$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 144$, $\frac{M}{P} = 0.5$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

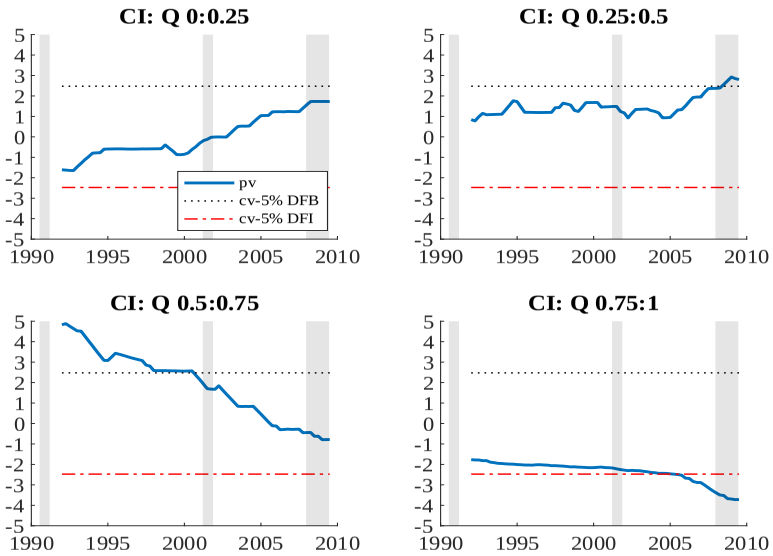


Figure 1.14: $F_{R,M,t}$ statistics, GDPonly, $h = 4$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 144$, $\frac{M}{P} = 0.5$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

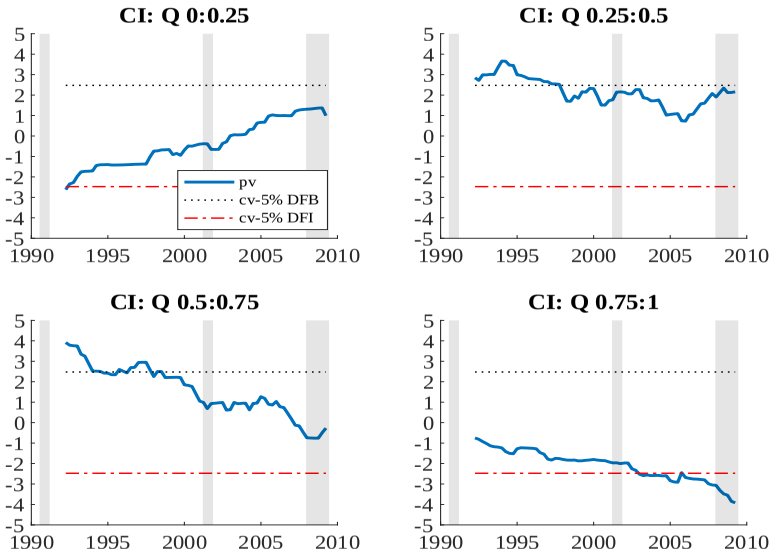


Figure 1.15: $F_{R,M,t}$ statistics, GDPonly, $h = 1$, rolling

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 144$, $\frac{M}{P} = 0.5$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

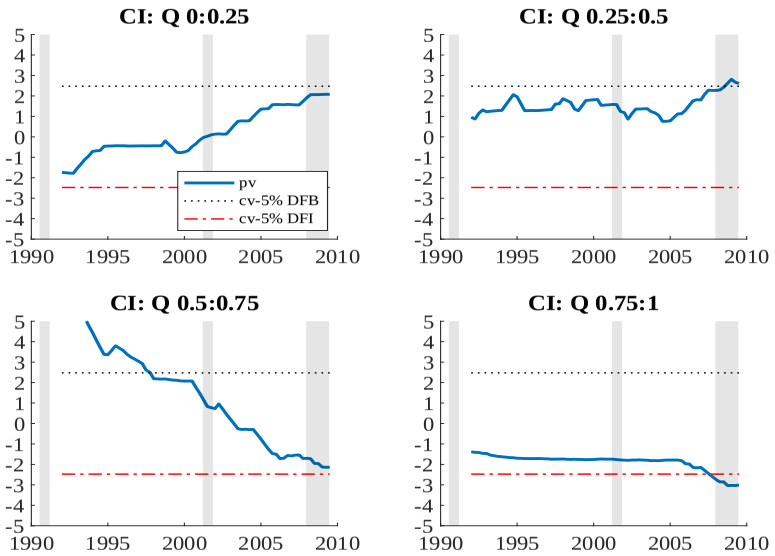


Figure 1.16: $F_{R,M,t}$ statistics, GDPonly, $h = 4$, rolling

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 144$, $\frac{M}{P} = 0.5$. *The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.*

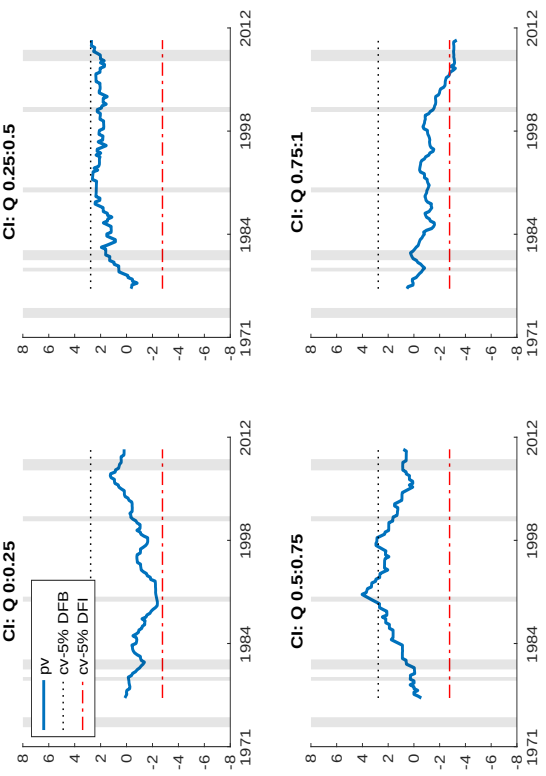


Figure 1.17: $F_{R,M,t}$ statistics, GDPonly, short R, $h = 1$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 92$, $\frac{M}{P} = 0.3$. The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.

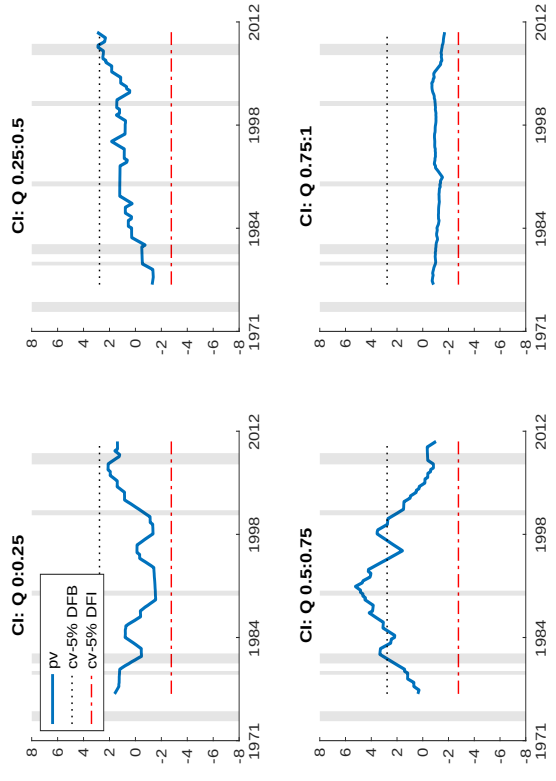


Figure 1.18: $F_{R,M,t}$ statistics, GDPonly, short R, $h = 4$, recursive

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 92$, $\frac{M}{P} = 0.3$. The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.

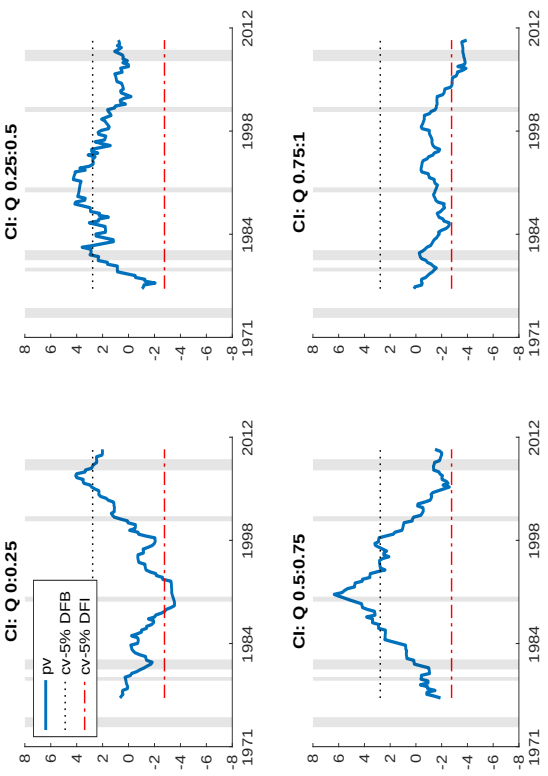


Figure 1.19: $F_{R,M,t}$ statistics, GDPonly, short R, $h = 1$, rolling

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 92$, $\frac{M}{P} = 0.3$. The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.

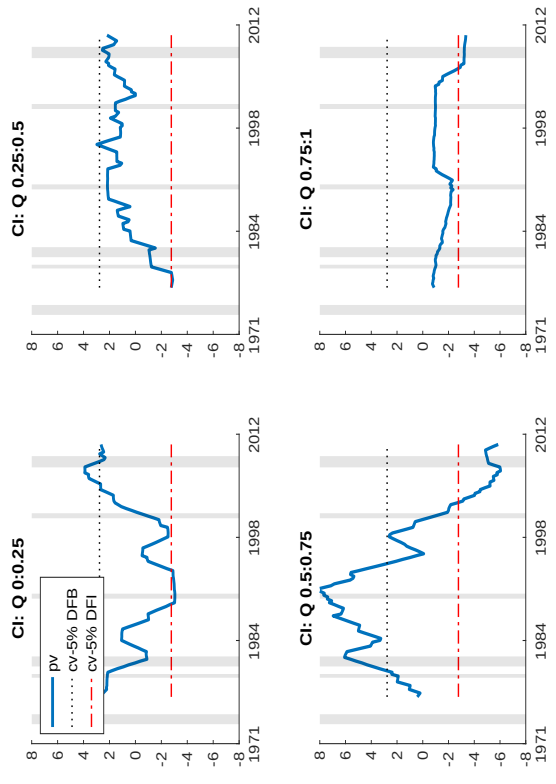


Figure 1.20: $F_{R,M,t}$ statistics, GDPonly, short R, $h = 4$, rolling

Note: The figure shows Fluctuation DFB test statistics considering specific quantiles of the distribution and the critical values of the Fluctuation DFB/DFI test. 'cv-5% DFB' is the critical value for the one-sided Fluctuation DFB test, and 'cv-5% DFI' is the critical value for the one-sided Fluctuation DFI test. Larger statistics imply worse out-of-sample performance relative to in-sample fit. $T = 280$, $R = 92$, $\frac{M}{P} = 0.3$. The shaded bands correspond to US recessions reported by the National Bureau of Economic Research.

Chapter 2

IDENTIFICATION AND ESTIMATION OF PARAMETER INSTABILITY IN A HIGH DIMENSIONAL APPROXIMATE FACTOR MODEL

2.1 Introduction

I propose a novel methodology for identifying and estimating structural breaks in the factor loadings of a high dimensional approximate factor model with an unknown number of latent factors. The approach is robust to structural changes in the volatility of the factors (the second moment of the factors), applicable to multiple structural breaks, and easy to implement for practitioners. Factor models are widely adopted in structural analyses (e.g., [Bernanke et al., 2005]) and forecasting (e.g., [Stock and Watson, 2002b]) in macroeconomics and finance. Empirical studies that consider a factor model generally assume that the parameters of the fac-

tor model are time-invariant, see [Connor and Korajczyk, 1986, Connor and Korajczyk, 1988, Connor and Korajczyk, 1993], [Forni and Reichlin, 2000, Forni and Reichlin, 2004], [Forni and Lippi, 2001], [Bai and Ng, 2002], [Bai, 2003a], and [Forni and Zaffaroni, 2015, Forni and Zaffaroni, 2017]. However, this assumption is in contrast to a strand of literature where structural breaks are deemed important features of macroeconomic panels, see [Stock and Watson, 1996], [Bates et al., 2013], [Corradi and Swanson, 2014], [Ma and Su, 2016], [Massacci, 2017], and [Su and Wang, 2017].

The consequences of ignoring large structural breaks in factor models are severe. First, standard estimators for the number of common factors become inconsistent ([Bates et al., 2013]). Second, forecasting performance, one of the major selling points of the factor model, deteriorates ([Banerjee et al., 2008]). Third, in structural analyses, incorrect conclusions regarding the mapping of the common factors to the observations can lead to misguided policy decisions. An example for the monetary transmission mechanism is provided in [Korobilis, 2013]. It is thus indispensable to develop reliable techniques to test and estimate structural breaks in factor models.

A number of studies have tested structural breaks in factor loadings, see [Stock and Watson, 2009], [Breitung and Eickmeier, 2011], [Yamamoto and Tanaka, 2015], [Chen et al., 2014], [Han and Inoue, 2015], [Cheng and Schorfheide, 2016], and [Barigozzi and Trapani, 2017]. In addition, a smaller body of recent work has focused on the estimation of structural breaks, see [Chen, 2015] and [Baltagi et al., 2017]. Yet the methods employed have predominantly relied on estimated factors (or loadings), assuming the second moment of the factors as constant. This makes it difficult to distinguish between a structural break in the factor loadings and a structural break in the volatility of the factors themselves, which in turn complicates understanding the relative importance of common factors and idiosyncratic terms.

In this article, I introduce a new methodology based on the observation that the sum of the numbers of pseudo factors (i.e., the factors needed to describe the samples) in the pre- and post-split subsamples is mini-

mized if the sample is split at the correct structural break. In other words, if the split point is incorrectly chosen, the number of factors in either the pre-split subsample or the post-split subsample will be overestimated ([Bai and Ng, 2002], [Breitung and Eickmeier, 2011]). This identification scheme is applicable regardless of the number of structural breaks. I operationalize the methodology by estimating the number of factors at different potential break points using the eigenvalue ratio criterion developed by [Ahn and Horenstein, 2013] and then adopting an appropriate criterion based on the eigenvalue ratios of the covariance matrices in the pre- and post-split subsamples. Each potential break yields the eigenvalue ratios of the pre- and post-break subsamples. By appropriately transforming these statistics, the true structural break ratio is consistently detected based on the well-known property that the r largest eigenvalues of the variance matrix of N response variables grow unboundedly as N increases, whereas the other eigenvalues remain bounded.

This study builds on previous work, including [Han and Inoue, 2015], who developed tests to detect structural breaks in factor loadings so as to determine whether all factor loadings are constant over time. They considered three different types of structural breaks, and used the second moments of the estimated factors to reduce the infinite-dimensional problem into a finite-dimensional problem. Relatedly, [Baltagi et al., 2017] discussed estimations of factor loading instability points based on changes in the second moments of the estimated factors. In this article, I adopt the same definitions and types of breaks as [Han and Inoue, 2015] and [Baltagi et al., 2017], but propose a different criterion. The new methodology does not, in fact, incorrectly attribute a structural break in the volatility of the common factor to a structural break in the factor loadings as the criterion proposed in this article is a function of eigenvalue ratios, which does not require the identification or estimation of the factors. In contrast, the approaches developed by [Han and Inoue, 2015] and [Baltagi et al., 2017] suffer from this problem, as they assume a constant second moment of the factors, and their statistics are functions of estimated factors. Moreover, most studies testing structural breaks in factor loadings are subject to the same limitation if their criteria are functions of estimated factors.

A combination of the methodologies of [Han and Inoue, 2015] and [Baltagi et al., 2017], together with this new methodology, allows instead to distinguish between a structural change in the factor loadings and a structural change in the volatility of the factors. This distinction is potentially quite important because of the need to distinguish between changes in the coefficient (i.e., the relative importance of the factors) and changes in the variance (i.e., the volatility of the factors) in macroeconomic and financial studies.

In a related paper, [Barigozzi and Trapani, 2017] tested the null hypothesis that the factor structure does not change based on the well-known property of the $(r+1)$-th eigenvalue of the sample covariance matrix of the data. Under the null hypothesis, the $(r+1)$-th eigenvalue is bounded, under the alternative of a change it becomes spiked. The methodology in this article differs in that the criterion in [Barigozzi and Trapani, 2017] is based on randomized eigenvalues, while the criterion I adopt is a transformation of eigenvalue ratios, and is simple to implement. Furthermore, our approach is robust to structural changes in the volatility of factors, and offers a procedure to distinguish between structural changes in the factor loadings and structural changes in the volatility of factors, as well as to estimate both instability points.

I carry out a Monte Carlo simulation study to evaluate the proposed methodology. The panels are simulated in different data-generating processes that involve different types of structural breaks in the factor loadings, allowing for different degrees of heteroskedasticity, and serial autocorrelation in idiosyncratic shocks as well as in the factors. The results show that the application of the new methodology is satisfactory in moderately large samples. Moreover, it compares favorably to the approach in [Baltagi et al., 2017].

Finally, I demonstrate the usefulness of this novel methodology in an empirical analysis. The analysis investigates whether there are structural breaks in a large panel of disaggregated US inflation series, which were studied in [Clark, 2006], [Boivin et al., 2009], [Reis and Watson, 2010], and [Stock and Watson, 2016]. The empirical results imply two structural breaks in the factor loadings, which correspond to the 1973 oil price shock

and the 2008 financial crisis, respectively. Furthermore, in combining this new approach and the existing methodologies in [Han and Inoue, 2015] and [Baltagi et al., 2017], I find empirical evidence in favor of a structural change in the volatility of the factors in January 1991. The results also provide empirical evidence that no structural breaks occurred either in the factor loadings or in the volatility of the factors from November 2008 to July 2017. The stability of the inflation panel during this period suggests that an unconventional monetary policy is as effective as a conventional one in stabilizing the economy, consistent with the findings in [Debortoli et al., 2018].

The remainder of this article is organized as follows. Section 3.2 introduces the factor model with one structural break in the factor loadings. Sections 2.3 and 2.4 discuss the identification and estimation of the structural breaks. Section 2.5 extends the methodology to allow for multiple structural breaks. Section 2.6 presents the results of the Monte Carlo simulation study and Section 2.7 explains the results of the empirical study. Section 3.5 concludes.

Notation: For any real number z, $\lfloor z \rfloor$ denotes the integer part of z. For a positive semidefinite matrix A, I use $\psi_k(A)$ to denote the k^{th} largest eigenvalue of a positive semidefinite matrix A, and $\hat{\cdot}$ to denote the estimated counterpart. I denote the norm of a matrix A as $\|A\| = [trace(A'A)]^{1/2}$.

2.2 Statistical model

In this section, I outline the theoretical framework and the types of structural breaks considered. For exposition purposes, the methodology is developed for the case of one structural break. In Section 2.5, I extend the framework to allow for multiple structural breaks. The extensions are conceptually straightforward, but notationally somewhat cumbersome.

Consider N time series that are each observed for T time periods. Let x_{it} denote the observation of series i at period t for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. At each point in time, the observations are modeled by a

factor model with r common factors, e.g. [Stock and Watson, 2002a]. But it deviates from the standard set-up by assuming that there is a break in the factor loadings at period $T^* = \lfloor \pi^* T \rfloor$, with $\pi^* \in (0, 1)$. The loadings of $r_0 < r$ factors do not change across T^*, while the loadings of the remaining r_1 factors change after period $\lfloor \pi^* T \rfloor$. Let us write the model with a structural break in the factor loadings as

$$x_{it} = \begin{cases} f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{1,i} + e_{it} & \text{if} \quad 1 \leqslant t \leqslant \lfloor \pi^* T \rfloor \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{2,i} + e_{it} & \text{if} \quad \lfloor \pi^* T \rfloor + 1 \leqslant t \leqslant T \end{cases} \tag{2.1}$$

where $f_{0,t}$ is the $r_0 \times 1$ vector of common factors at period t with time-invariant factor loadings $\lambda_{0,i}$, $f_{1,t}$ is the $r_1 \times 1$ vector of common factors at period t with pre-break factor loadings $\lambda_{1,i}$ and post-break factor loadings $\lambda_{2,i}$, and e_{it} is the idiosyncratic shock for series i at period t. In matrix notation, model (2.1) can be written as

$$X = \begin{bmatrix} F_{0,1}\Lambda'_0 + F_{1,1}\Lambda'_1 \\ F_{0,2}\Lambda'_0 + F_{1,2}\Lambda'_2 \end{bmatrix} + E \tag{2.2}$$

where X is the $T \times N$ matrix of observations with element x_{it} in the i^{th} column and t^{th} row. The common factors are collected in $F_{0,1} = (f_{0,1}, \ldots, f_{0,\lfloor \pi^* T \rfloor})'$, $F_{0,2} = (f_{0,\lfloor \pi^* T \rfloor + 1}, \ldots, f_{0,T})'$, $F_{1,1} = (f_{1,1}, \ldots, f_{1,\lfloor \pi^* T \rfloor})'$ and $F_{1,2} = (f_{1,\lfloor \pi^* T \rfloor + 1}, \ldots, f_{1,T})'$. The corresponding loading matrices are $\Lambda_0 = (\lambda_{0,1}, \ldots, \lambda_{0,N})'$, $\Lambda_1 = (\lambda_{1,1}, \ldots, \lambda_{1,N})'$ and $\Lambda_2 = (\lambda_{2,1}, \ldots, \lambda_{2,N})'$, respectively. Finally, E is the $T \times N$ idiosyncratic matrix with elements e_{it}.

Following [Han and Inoue, 2015], different types of breaks in the factor loadings around T^* can be defined in terms of the following pseudo model.

$$X = \begin{bmatrix} F_{0,1} & F_{1,1} & O \\ F_{0,2} & O & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda'_0 \\ \Lambda'_1 \\ \Lambda'_2 \end{bmatrix} + E = G\Theta' + E \tag{2.3}$$

In particular, the rank of $\Theta'\Theta/N$ characterizes the types of breaks in the following way. $\mathrm{Rank}(\Theta'\Theta/N) = r_0 + r_1 + l$, where $0 \leq l \leq r_1$. When

$l = r_1$, it follows that Λ_1 and Λ_2 are linearly independent, i.e., the loadings change in a completely different direction. In contrast, when $l = 0$, it follows that $\Lambda_2 = \Lambda_1 Z$, and two cases can be distinguished: if Z is singular, some common factors disappear after the break, for example, if $\Lambda_2 = O$, then r_1 factors disappear after the break; if Z is non-singular, it can be treated as a break in the volatility of the factors, for example, if $\Lambda_2 = 2\Lambda_1$, then the factors are less volatile while the relative strength between factors still remain the same. When $0 < l < r_1$, it is a combination of the aforementioned breaks. Note that by construction $[\Lambda_0, \Lambda_1, \Lambda_2]$ may not be of full column rank, but at least one of Λ_1 and Λ_2 is set to be of full rank, that is, the case when there are some factors appearing and some factors disappearing at the same time is not considered in this context. Summarizing, based on $\mathrm{Rank}(\Theta'\Theta/N)$, it is able to distinguish between different types of structural breaks.

2.3 Identification of structural breaks

In this section, I show that the sum of the numbers of pseudo factors in the pre- and post-split subsamples can be used to identify the structural break. In particular, when viewing the break as a parameter of choice that selects T^*, the structural break minimizes the numbers of pseudo factors needed to describe both the pre- and the post-split panel.

Let us split the whole sample into two subsamples at period $\lfloor \pi T \rfloor$ with $\pi \in (0, 1)$, and denote the pre- and post-split subsamples as $X_\pi^{pre} = (x_1^{'}, \ldots, x_{\lfloor \pi T \rfloor}^{'})'$ and $X_\pi^{post} = (x_{\lfloor \pi T \rfloor + 1}^{'}, \ldots, x_T^{'})'$. $G_\pi^{pre/post}$, $\Theta_\pi^{pre/post}$ and $E_\pi^{pre/post}$ are similarly defined. Three cases can be considered based on the relationship between π and π^*: $\pi < \pi^*$, $\pi = \pi^*$ and $\pi > \pi^*$. Let us denote the number of pseudo factors[1] of the pre-split subsample as $r^{pre}(\pi)$ and the number of pseudo factors of the post-split subsample as $r^{post}(\pi)$. Both $r^{pre}(\pi)$ and $r^{post}(\pi)$ are functions of π. Let us denote them as r^{pre}

[1]The number of "pseudo" factors means the number of factors needed to describe the corresponding sample, i.e. G in (2.3), which are not necessarily the true factors.

and r^{post} for simplicity. Next, let us discuss the three cases separately.

Case 1: $\pi < \pi^*$
In this case, the break in the loadings lies in the post-split subsample. The model for this case can be written as

$$\begin{aligned} X_\pi^{pre} &= G_\pi^{pre}\Theta_\pi^{pre'} + E_\pi^{pre} = \begin{bmatrix} F_{0,1}^{pre} & F_{1,1}^{pre} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_1' \end{bmatrix} + E_\pi^{pre} \\ X_\pi^{post} &= G_\pi^{post}\Theta_\pi^{post'} + E_\pi^{post} = \begin{bmatrix} F_{0,1}^{post} & F_{1,1}^{post} & O \\ F_{0,2} & O & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_1' \\ \Lambda_2' \end{bmatrix} + E_\pi^{post} \end{aligned} \tag{2.4}$$

where $F_{0,1}^{pre}$ and $F_{1,1}^{pre}$ denote the the pre-split part of $F_{0,1}$ and $F_{1,1}$; while $F_{0,1}^{post}$ and $F_{1,1}^{post}$ denote the post-split part of $F_{0,1}$ and $F_{1,1}$. Equations (2.4) imply that when $\pi < \pi^*$, the number of factors of the pre-split subsample is $r_0 + r_1$ and the number of factors of the post-split subsample depends on the rank of $\Theta'\Theta/N$. The $\pi < \pi^*$ column of Table 2.1 summarizes r^{pre} and r^{post} for each type of breaks when $\pi < \pi^*$.

Case 2: $\pi = \pi^*$
When $\pi = \pi^*$, both the pre-split subsample $X_{\pi^*}^{pre}$ and the post-split subsample $X_{\pi^*}^{post}$ do not have any breaks in the factor loadings. The following expression is obtained:

$$\begin{aligned} X_{\pi^*}^{pre} &= G_{\pi^*}^{pre}\Theta_{\pi^*}^{pre'} + E_{\pi^*}^{pre} = \begin{bmatrix} F_{0,1} & F_{1,1} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_1' \end{bmatrix} + E_{\pi^*}^{pre} \\ X_{\pi^*}^{post} &= G_{\pi^*}^{post}\Theta_{\pi^*}^{post'} + E_{\pi^*}^{post} = \begin{bmatrix} F_{0,2} & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_2' \end{bmatrix} + E_{\pi^*}^{post} \end{aligned} \tag{2.5}$$

In the same spirit as above, the number of factors for the pre- and post-break subsamples can be deduced. The number of factors of the pre-split subsample is $r_0 + r_1$ and the number of factors of post-split subsample depends on $[\Lambda_0\ \Lambda_2]$, which in its turn depends on the type of breaks. The $\pi = \pi^*$ column in Table 2.1 lists the different outcomes. The integer c in the table depends on $[\Lambda_0\ \Lambda_2]$, which could have columns of zeros. Especially, if there is a break in the volatility of the factor, i.e., there is a

non-singular Z such that $\Lambda_2 = \Lambda_1 Z$, then $c = r_1$. Otherwise, $c = 0$, i.e., some factors disappear.

Case 3: $\pi > \pi^*$
When $\pi > \pi^*$, the pre-split subsample X_π^{pre} has a break in the factor loadings while the post-split subsample X_π^{post} does not have any break in the factor loadings. The pre- and post-split subsamples have the expression as equation (2.6).

$$\begin{aligned} X_\pi^{pre} &= G_\pi^{pre}\Theta_\pi^{pre\prime} + E_\pi^{pre} = \begin{bmatrix} F_{0,1} & F_{1,1} & O \\ F_{0,2}^{pre} & O & F_{1,2}^{pre} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_1' \\ \Lambda_2' \end{bmatrix} + E_\pi^{pre} \\ X_\pi^{post} &= G_\pi^{post}\Theta_\pi^{post\prime} + E_\pi^{post} = \begin{bmatrix} F_{0,2}^{post} & F_{1,2}^{post} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_2' \end{bmatrix} + E_\pi^{post} \end{aligned} \tag{2.6}$$

where $F_{0,2}^{pre}$ and $F_{1,2}^{pre}$ denote the the pre-split part of $F_{0,2}$ and $F_{1,2}$; while $F_{0,2}^{post}$ and $F_{1,2}^{post}$ denote the post-split part of $F_{0,2}$ and $F_{1,2}$.

The number of factors of the pre-split subsample depends on the pseudo factor loading matrix, that is, the rank of $\Theta'\Theta/N$; while the number of factors of the post-split subsample depends on $[\Lambda_0\ \Lambda_2]$, which in turn depends on the type of breaks. The $\pi > \pi^*$ column in Table 2.1 summarizes r^{pre} and r^{post} for each type of breaks when $\pi > \pi^*$.

Adding up the numbers of pesudo factors of the implied pre- and post-split subsamples, I find that:

- For the case of a structural break in the factor loadings at period $T^* = \lfloor \pi^* T \rfloor$, i.e. Type 1 and 3 in Table 2.1, it holds that

$$r^{pre}(\pi^*) + r^{post}(\pi^*) < r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi \neq \pi^* \tag{2.7}$$

- For the case of factors disappearing at period $T^* = \lfloor \pi^* T \rfloor$, i.e. Type 2 in Table 2.1 with $c = 0$, it holds that

$$\begin{aligned} r^{pre}(\pi^*) + r^{post}(\pi^*) &< r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi < \pi^* \\ r^{pre}(\pi^*) + r^{post}(\pi^*) &= r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi \geq \pi^* \end{aligned} \tag{2.8}$$

- For the case of a structural break in the volatility of the factors at period $T^* = \lfloor \pi^* T \rfloor$, i.e. Type 2 in Table 2.1 with $c = r_1$, it holds that

$$r^{pre}(\pi^*) + r^{post}(\pi^*) \quad = \quad r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi \tag{2.9}$$

Summarizing, $r^{pre}(\pi) + r^{post}(\pi)$ depends on where the sample is split, and at the same time, it characterizes the type of breaks. Table 2.1 shows that this information is sufficient to identify structural breaks.

Table 2.1: SUMMARY OF r^{pre}, r^{post} AND $r^{pre} + r^{post}$

Type of Breaks	-	$\pi < \pi^*$	$\pi = \pi^*$	$\pi > \pi^*$
Type 1: $\text{rank}(\Theta'\Theta/N) = r_0 + r_1 + r_1$	$r^{pre}(\pi)$	$r_0 + r_1$	$r_0 + r_1$	$r_0 + r_1 + r_1$
	$r^{post}(\pi)$	$r_0 + r_1 + r_1$	$r_0 + r_1$	$r_0 + r_1$
	$r^{pre}(\pi) + r^{post}(\pi)$	$2r_0 + 3r_1$	$\mathbf{2r_0 + 2r_1}$	$2r_0 + 3r_1$
Type 2: $\text{rank}(\Theta'\Theta/N) = r_0 + r_1$	$r^{pre}(\pi)$	$r_0 + r_1$	$r_0 + r_1$	$r_0 + r_1$
	$r^{post}(\pi)$	$r_0 + r_1$	$r_0 + c$	$r_0 + c$
	$r^{pre}(\pi) + r^{post}(\pi)$	$2r_0 + 2r_1$	$\mathbf{2r_0 + r_1 + c}$	$\mathbf{2r_0 + r_1 + c}$
Type 3: $\text{rank}(\Theta'\Theta/N) = r_0 + r_1 + l$	$r^{pre}(\pi)$	$r_0 + r_1$	$r_0 + r_1$	$r_0 + r_1 + l$
	$r^{post}(\pi)$	$r_0 + r_1 + l$	$r_0 + l$	$r_0 + l$
	$r^{pre}(\pi) + r^{post}(\pi)$	$2r_0 + 2r_1 + l$	$\mathbf{2r_0 + r_1 + l}$	$2r_0 + r_1 + 2l$

Note that l, c are integers that satisfy $0 < l < r_1$, $c \in \{0, r_1\}$.

2.4 Estimation and inference

In this section, I operationalize the identification strategy of the previous section. The sum of the pre- and post-split subsamples' numbers of common factors of different choices of split point π can be used to detect π^*.

Given the identification results described in equations (2.7), a natural estimator for the structural break in the factor loadings is given by

$$\tilde{\pi} = \arg\min_{\pi \in \Pi} \ \tilde{r}^{pre}(\pi) + \tilde{r}^{post}(\pi) \tag{2.10}$$

where $\tilde{r}^{pre}(\pi)$ is the estimated number of factors of the pre-split subsample and $\tilde{r}^{post}(\pi)$ is the estimated number of factors of the post-split subsample. Different estimators can be considered for the estimation of the number of factors, see for example [Bai and Ng, 2002], [Onatski, 2010] and [Ahn and Horenstein, 2013].

While estimator defined in eq (2.10) is intuitively appealing, it has some considerable drawbacks. First, the case of factors disappearing, where the sum of pre- and post-split numbers of factors remains constant after the break, see eq (2.8), will not be detected by this estimator. Second, the objective function defined in eq (2.10) is an integer-valued function, which has one single drop at the break. It is difficult to construct the convergence theory based on such objective function.

Instead, I modify the objective function by (a) looking at the values of the eigenvalue ratios instead of their index and (b) differencing the criterion to be able to detect breaks of factors (dis)appearing, i.e., the case in eq (2.8).

I first introduce some notations. I define, as [Ahn and Horenstein, 2013], that

$$\hat{\mu}^{pre}_{NT,k}(\pi) = \psi_k\left(\frac{X^{pre\prime}X^{pre}}{N\lfloor \pi T \rfloor}\right), \qquad \hat{\mu}^{pre}_{NT,0}(\pi) = \frac{\sum_{k=1}^{\min(T,N)} \hat{\mu}^{pre}_{NT,k}(\pi)}{\ln(\min(\lfloor \pi T \rfloor, N))}$$

$$\hat{\mu}^{post}_{NT,k}(\pi) = \psi_k\left(\frac{X^{post\prime}X^{post}}{N(T - \lfloor \pi T \rfloor)}\right), \qquad \hat{\mu}^{post}_{NT,0}(\pi) = \frac{\sum_{k=1}^{\min(T,N)} \hat{\mu}^{post}_{NT,k}(\pi)}{\ln(\min(T - \lfloor \pi T \rfloor, N))}$$

With these notations introduced, I consider the following criterion

$$\Delta\hat{Q}(\pi) = \hat{Q}(\pi) - \hat{Q}(\pi - \Delta\pi) \tag{2.11}$$

with

$$\hat{Q}(\pi) = \frac{1}{N} \max_{k_1,k_2,\ s.t.\ k_1+k_2=\hat{R}} \left\{ \frac{\hat{\mu}^{pre}_{NT,k_1}(\pi)}{\hat{\mu}^{pre}_{NT,k_1+1}(\pi)} + \frac{\hat{\mu}^{post}_{NT,k_2}(\pi)}{\hat{\mu}^{post}_{NT,k_2+1}(\pi)} \right\} \tag{2.12}$$

and

$$\hat{R} = \min_{\tilde{\pi}} \left\{ \hat{r}^{pre}(\tilde{\pi}) + \hat{r}^{post}(\tilde{\pi}) \right\} \tag{2.13}$$

where $\Delta\pi$ is a small measurement of distance.[2] $\hat{r}^{pre/post}(\cdot)$ is the estimated number of factors of the corresponding pre- or post-split subsamples' covariance matrices using the eigenvalue ratio estimator proposed in Ahn and Horenstein(2013) [3].

The estimator for the structural break based on eq (2.11) is given by

$$\hat{\pi} = \arg\max_{\pi \in \Pi} \Delta\hat{Q}(\pi). \tag{2.14}$$

As shown in [Ahn and Horenstein, 2013], the k^{th} eigenvalue ratio, if k is the correct number of factors of the corresponding covariance matrix, will explode; while the other eigenvalue ratios will be bounded. The alternative objective function adopted in (2.14) will capture the structural break and the intuition is the following: in population, with the restriction $k_1 + k_2 = \min_{\tilde{\pi}} \{r^{pre}(\tilde{\pi}) + r^{post}(\tilde{\pi})\}$, at the correct split point (i.e., split at the break), the objective function will attain the largest eigenvalue ratios of both pre- and post-break subsamples' covariance matrices; while at

[2] In theory, the convergence of $\hat{\pi}$ requires $\Delta\pi = \frac{T^\epsilon}{T} \to 0$ as $T \to \infty$. In practice, I suggest choosing $\Delta\pi$ based on the sample size and economic meaning. For example, if a researcher deals with a quarterly panel with $T = 100$ and he/she may search through $\pi \in \Pi = \{0.15, 0.16, \ldots, 0.84, 0.85\}$, then he/she can choose $\Delta\pi = 0.01$ with $\Delta\pi T$ refers a quarter or choose $\Delta\pi = 0.04$ with $\Delta\pi T$ refers a year.

[3] Ahn and Horenstein (2013) Eigenvalue Ratio estimator(ER): $\hat{r}_{ER} = \max_{r \leqslant r_{max}} \frac{\tilde{\mu}_{NT,r}}{\tilde{\mu}_{NT,r+1}}$, $\tilde{\mu}_{NT,r} \equiv \psi_r[X'X/(NT)]$, where $\psi_k(.)$ denotes the k^{th} largest eigenvalue of the corresponding matrix.

the incorrect split point, the objective function cannot attain both largest eigenvalue ratios of the two subsamples' covariance matrices simultaneously, which makes $Q(.)$ maximized at the break. Besides, taking difference of the criterions $Q(.)$ help achieve a general form of the estimator in cases of different types of breaks, as in the case of factors disappearing, $Q(.)$ is supposed to remain at a high level after the break while $\Delta Q(.)$ will be maximized at the break. Therefore, the alternative objective function adopted in (2.14) captures the structural break and in the same time help circumvent the problems in estimation equation (2.10).

Next, I set up the conditions under which $\hat{\pi}$ is a consistent estimator of π^*. Let X be generated by model (2.2), which has a representation (2.3). Let r^{whole} denote the true pseudo number of factors of the whole sample. For any $\pi \in \Pi \subset (0,1)$, let $r^{pre/post}(\pi)$ denote the true pseudo pre- and post-split numbers of factors as described in Table 2.1, and let $G_\pi^{pre/post}$,$\Theta_\pi^{pre/post}$ denote the true pseudo pre- and post-split factors and loadings as described in eq (2.4)-(2.6). Finally, let $m = \min(N,T)$, $M = \max(N,T)$. I assume that

Assumption 1. $T^* = \lfloor \pi^* T \rfloor$, *where* $\eta < \pi^* < 1-\eta$, $\eta > 0$.

Assumption 2. *(i) Let* $\mu_{NT,k}^{whole} = \psi_k \left[\frac{\Theta'\Theta}{N} \frac{G'G}{T} \right]$ *for* $k = 1, \ldots, r^{whole}$. *Then for each* $k = 1, \ldots, r^{whole}$, $\operatorname{plim}_{m\to\infty} \mu_{NT,k}^{whole} = \mu_k^{whole}$, *and* $0 < \mu_k^{whole} < \infty$. *(ii) Let* $\mu_{NT,k}^{pre}(\pi^*) = \psi_k \left[\frac{\Theta_{\pi^*}^{pre'} \Theta_{\pi^*}^{pre'}}{N} \frac{G_{\pi^*}^{pre'} G_{\pi^*}^{pre}}{\lfloor \pi^* T \rfloor} \right]$ *for* $k = 1, \ldots, r^{pre}(\pi^*)$, *and* $\mu_{NT,k}^{post}(\pi^*) = \psi_k \left[\frac{\Theta_{\pi^*}^{post'} \Theta_{\pi^*}^{post}}{N} \frac{G_{\pi^*}^{post'} G_{\pi^*}^{post}}{T - \lfloor \pi^* T \rfloor} \right]$ *for* $k = 1, \ldots, r^{post}(\pi^*)$. *Then for each* $k = 1, \ldots, r^{pre/post}(\pi^*)$, $\operatorname{plim}_{m\to\infty} \mu_{NT,k}^{pre/post}(\pi^*) = \mu_k^{pre/post}$, *and* $0 < \mu_k^{pre/post} < \infty$. *(iii)* r^{whole}, $r^{pre/post}(\pi^*)$ *are finite.*

Assumption 3. *(i)* $E \left\| f_t \right\|^4 < c_1$, $\left\| \lambda_i \right\| < c_1$, *for all* i *and* t. *(ii)* $E \left\| N^{-1/2} \sum_i \epsilon_{it} \lambda_i \right\|^2 < c_1$, *for all* t. *(iii)* $E \left\| T^{-1/2} \sum_i \epsilon_{it} f_t \right\|^2 < c_1$.

Assumption 4. *(i)* $0 < y \equiv \lim_{m\to\infty} m/M \leqslant 1$. *(ii) Let the idiosyncratic shocks* E *take the structure* $E = R_T^{1/2} U G_N^{1/2}$ *where* $R_T^{1/2}$ *and*

$G_N^{1/2}$ are symmetric square roots of positive semidefinite matrices R_T and G_N, $U = [u_{it}]_{T\times N}$. (iii) u_{it} are $i.i.d$ random variables, $E(u_{it}^4) < \infty$. (iv) $\psi_1(R_T) < c_1$, and $\psi_1(G_N) < c_1$, uniformly in T and N respectively.

Assumption 5. *(i) $\psi_T(R_T) > c_2$ for all T. (ii) Let $y^* = \lim_{m\to\infty} m/N = \min(y, 1)$. Then there exists a real number $d^* \in (1-\eta y^*, 1]$ such that $\psi_{\lfloor d^* N\rfloor}(G_N) > c_2, \forall N$.*

Assumption 1 describes the structural break. It indicates that the structural break is not at two extremes of the whole sample and each segment increases proportionately as the sample size increases. Later this assumption will be extended to allow multiple breaks. Assumption 2 naturally extends the Assumption A of [Ahn and Horenstein, 2013]. Assumption 2(i), following many factor model literatures (see [Ahn and Horenstein, 2013], [Connor and Korajczyk, 1993], [Bai and Ng, 2002, Bai and Ng, 2006], [Bai, 2003a]), assumes that the cumulative effect of the "least influential factor" diverges in probability to infinity as the sample size grows. In Assumption 2(ii), I extend this assumption such that it applies for pre- and post-break subsamples, i.e., the cumulative effect of the "least influential factor" of the pre- and post-break subsample also diverges in probability to infinity as the sample size grows. Note that each segment increases proportionately as the sample size increases is ensured in Assumption 1. Assumptions 3-4 are exactly the same assumptions of [Ahn and Horenstein, 2013]. Assumption 3 allows weak dependence between factors and idiosyncratic shocks, as well as loadings and idiosyncratic shocks. Assumption 4 restricts the covariance structure of the idiosyncratic shocks, it allows both autocorrelation, determined by R_T, and cross-sectional correlation, governed by G_N, in the idiosyncratic shocks. Assumption 5 states that none of the idiosyncratic shocks and their linear functions can be perfectly predicted by their past values, and it permits multicollinearity among variables (see examples in [Ahn and Horenstein, 2013]) so long as an asymptotically nonnegligible portion of the eigenvalues of G_N is bounded below by a positive number. I restrict $\exists d^* \in (1-\eta y^*, 1]$ to ensure that it holds for subsamples. Assumptions 2-5 ensure that the eigenvalue ratio estimator for the number of common factors in [Ahn and Horenstein, 2013] can consistently identify the number of common factors for

different pre- and post-split subsamples.

With the assumptions introduced above, I achieve the following result.

Theorem 3. *Let $x_{i,t}$ be generated by model* (2.1)*. Suppose Assumptions 1-5 hold, $\hat{\pi}$ is defined in equation* (2.14)*, then $\hat{\pi} = \pi^* + O_p(T^{-1})$.*

The criterion function (2.12) can shed some insights into which type of breaks is being detected. If $\hat{Q}(.)$ is smooth across time, then there are neither structural breaks in the factor loadings nor factors (dis)appearing, but there still might be structural breaks in the volatility of the factors; if $\hat{Q}(.)$ has a peak, then there is a structural break in the factor loadings; if $\hat{Q}(.)$ jumps to a high value and then remains (or vice versa), then there are factors disappearing (appearing). Moreover, combining this new method with the literature (see [Han and Inoue, 2015], [Baltagi et al., 2017]), researchers can distinguish between a structural break in the factor loadings and a structural break in the volatility of the factors. If the existing methods in the literature capture a structural break while the new method does not, i.e., $\hat{Q}(.)$ is smooth, then there is a structural break in the volatility of the factors.

2.5 Detecting multiple structural breaks

In this section, I extend the results from the previous sections to the case of multiple structural breaks. In general, I assume that there are q structural breaks and denote the structural breaks by π_j^* for $j = 1, \ldots, q$. The detection procedure is based on the observation that at each structural break, the true data generating process can be rewritten in a representation that is conceptually equivalent to model (2.3).

2.5.1 Identification of multiple structural breaks

A factor model with q structural breaks is given by eq (2.15).

$$x_{it} = \begin{cases} f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{1,i} + f'_{2,t}\lambda^2_{1,i} + \cdots + f'_{q,t}\lambda^q_{1,i} + e_{it} & \text{if} \quad 1 \leqslant t \leqslant \lfloor \pi^*_1 T \rfloor \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{2,i} + f'_{2,t}\lambda^2_{1,i} + \cdots + f'_{q,t}\lambda^q_{1,i} + e_{it} & \text{if} \quad \lfloor \pi^*_1 T \rfloor + 1 \leqslant t \leqslant \lfloor \pi^*_2 T \rfloor \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{2,i} + f'_{2,t}\lambda^2_{2,i} + \cdots + f'_{q,t}\lambda^q_{1,i} + e_{it} & \text{if} \quad \lfloor \pi^*_2 T \rfloor + 1 \leqslant t \leqslant \lfloor \pi^*_3 T \rfloor \\ \cdots & \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{2,i} + f'_{2,t}\lambda^2_{2,i} + \cdots + f'_{q,t}\lambda^q_{2,i} + e_{it} & \text{if} \quad \lfloor \pi^*_{q-1} T \rfloor + 1 \leqslant t \leqslant T \end{cases} \tag{2.15}$$

where $f_{0,t}$ is a $r_0 \times 1$ vector prepresenting factors at period t with time-invariant factor loadings $\lambda_{0,i}$. For $j = 1, \ldots, q$, $f_{j,t}$ is a $r_j \times 1$ vector representing factors whose loadings change at j^{th} break ($T^*_j = \lfloor \pi^*_j T \rfloor$), with pre-$j^{th}$break factor loadings $\lambda^j_{1,i}$ and post-j^{th}break factor loadings $\lambda^j_{2,i}$. e_{it} is the idiosyncratic shock for series i at period t. In matrix notation, model (2.15) can be written as eq (2.16).

$$X = \begin{bmatrix} F_{0,1} & F_{1,1} & F_{2,1} & \ldots & F_{q,1} & \vdots & O & O & \ldots & O \\ F_{0,2} & O & F_{2,2} & \ldots & F_{q,2} & \vdots & F_{1,2} & O & \ldots & O \\ \vdots & & & & & & & & & \vdots \\ F_{0,q+1} & O & O & \ldots & O & \vdots & F_{1,q+1} & F_{2,q+1} & \ldots & F_{q,q+1} \end{bmatrix} \begin{bmatrix} \Lambda'_0 \\ \Lambda^{1'}_1 \\ \Lambda^{2'}_1 \\ \vdots \\ \Lambda^{q'}_1 \\ \Lambda^{1'}_2 \\ \Lambda^{2'}_2 \\ \vdots \\ \Lambda^{q'}_2 \end{bmatrix} + E \tag{2.16}$$

Consider structural break $T^*_j = \lfloor \pi^*_j T \rfloor$. The panel includes $j-1$ breaks before T^*_j and $q-j$ breaks after T^*_j. Both the pre- and post-jth break subsample can be written as an alternative pseudo factor model with constant loading matrices as in (2.16). Given that r_j factor loadings

change at T_j^*, the data generating process can be written as

$$X = G\Theta' + E = \begin{bmatrix} \tilde{F}_{0,1}\tilde{\Lambda}_0' + \tilde{F}_{1,1}\tilde{\Lambda}_1' \\ \tilde{F}_{0,2}\tilde{\Lambda}_0' + \tilde{F}_{1,2}\tilde{\Lambda}_2' \end{bmatrix} + E = \begin{bmatrix} \tilde{F}_{0,1} & \tilde{F}_{1,1} & O \\ \tilde{F}_{0,2} & O & \tilde{F}_{1,2} \end{bmatrix} \begin{bmatrix} \tilde{\Lambda}_0' \\ \tilde{\Lambda}_1' \\ \tilde{\Lambda}_2' \end{bmatrix} + E \tag{2.17}$$

where $\tilde{F}_{0,1}$ and $\tilde{F}_{1,1}$ are $T_j^* \times \tilde{r}_0$ and $T_j^* \times \tilde{r}_1$ pseudo pre-j^{th}break factor matrices, i.e., the upper-left part and upper-right part of big factor matrix in eq (2.16) respectively. $\tilde{F}_{0,2}$ and $\tilde{F}_{0,2}$ are $(T - T_j^*) \times \tilde{r}_0$ and $(T - T_j^*) \times \tilde{r}_1$ pseudo post-j^{th}break factor matrices, i.e., the bottom-left part and bottom-right part of big factor matrix in eq (2.16) respectively. The number of pseudo factors before T_j^* is $\tilde{r}_0$. $\tilde{\Lambda}_0$ is an $N \times \tilde{r}_0$ pseudo factor loading matrix that denotes loadings that remain constant at T_j^*, $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ are $N \times \tilde{r}_1$ matrices that denote pre- and post-j^{th}break loadings that change at T_j^*.

Based on the representation in eq (2.17), I deduce that different types of breaks at point T_j^* can be described by the rank of $\tilde{\Theta}$. Similar to the results in Table 2.1, after adding up the numbers of factors of the implied pre- and post-split subsamples, it follows that:

- For the case of a structural break in the loadings at period $T_j^* = \lfloor \pi_j^* T \rfloor$, it holds that

$$r^{pre}(\pi_j^*) + r^{post}(\pi_j^*) < r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi \neq \pi_j^*, \pi_{j-1}^* < \pi < \pi_{j+1}^* \tag{2.18}$$

- For the case of factors disappearing at period $T_j^* = \lfloor \pi_j^* T \rfloor$, it holds that

$$\begin{aligned} r^{pre}(\pi_j^*) + r^{post}(\pi_j^*) &< r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi_{j-1}^* < \pi < \pi_j^* \\ r^{pre}(\pi_j^*) + r^{post}(\pi_j^*) &= r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi_j^* \leq \pi < \pi_{j+1}^* \end{aligned} \tag{2.19}$$

- For the case of a structural break in the volatility of the common factors at period $T_j^* = \lfloor \pi_j^* T \rfloor$, it holds that

$$r^{pre}(\pi_j^*) + r^{post}(\pi_j^*) = r^{pre}(\pi) + r^{post}(\pi) \quad \forall \pi, \pi_{j-1}^* < \pi < \pi_{j+1}^* \tag{2.20}$$

2.5.2 Estimation and inference for multiple structural breaks

Given the identification results described in equations (2.18), a natural estimator for the structural break π_j^* is given by

$$\tilde{\pi}_j = \arg \min_{\pi_{j-1}^* < \pi < \pi_{j+1}^*} \tilde{r}^{pre}(\pi) + \tilde{r}^{post}(\pi) \tag{2.21}$$

Similarly, I modify the objective function and consider

$$\Delta \hat{Q}_j(\pi) = \hat{Q}_j(\pi) - \hat{Q}_j(\pi - \Delta\pi) \tag{2.22}$$

with

$$\hat{Q}_j(\pi) = \frac{1}{N} \max_{k_1,k_2,\ s.t.\ k_1+k_2=\hat{R}_j} \left\{ \frac{\hat{\mu}_{NT,k_1}^{pre}(\pi)}{\hat{\mu}_{NT,k_1+1}^{pre}(\pi)} + \frac{\hat{\mu}_{NT,k_2}^{post}(\pi)}{\hat{\mu}_{NT,k_2+1}^{post}(\pi)} \right\} \tag{2.23}$$

and

$$\hat{R}_j = \min_{\pi_{j-1}^* < \tilde{\pi} < \pi_{j+1}^*} \left\{ \hat{r}^{pre}(\tilde{\pi}) + \hat{r}^{post}(\tilde{\pi}) \right\} \tag{2.24}$$

The estimator for the instability point based on (2.22) is given by

$$\hat{\pi}_j = \arg \max_{\pi_{j-1}^* < \pi < \pi_{j+1}^*} \Delta \hat{Q}_j(\pi) \qquad \forall j = 1, 2, \ldots, q \tag{2.25}$$

Note that in multiple structural breaks case, when identifying and estimating the j^{th} break, I use the trimming Π_j that satisfies $\pi_j \in \Pi_j \subseteq (\pi_{j-1}^*, \pi_{j+1}^*)$. In practice, π_{j-1}^* and π_{j+1}^* are unknown, so I suggest replacing $\hat{R}_j$ in eq (2.23) and (2.24) with

$$\hat{R}_j(\pi, h) = \min_{\max(\pi_0, \pi-h) < \tilde{\pi} < \min(\pi+h, 1-\pi_0)} \left\{ \hat{r}^{pre}(\tilde{\pi}) + \hat{r}^{post}(\tilde{\pi}) \right\} \tag{2.26}$$

π_0 and $1-\pi_0$ are the starting and ending points of the trimming set of $\pi \in \Pi = (\pi_0, \pi_0)$, which is usually set that $\pi \in (0.15, 0.85)$ in practice. Idealy, $h \in (0, 1)$ should be set no less then half the duration between adjacent

breaks to ensure that the interval $(\max(\pi_0, \pi - h), \min(\pi + h, 1 - \pi_0))$ contains a structural break if there are any through the whole sample. For example, if the duration between adjacent breaks is assumed to be no longer than $0.4T$ periods, say $\pi_1^* = 0.3, \pi_2^* = 0.7$, then the ideal h satisfies that $h \in (0.2, 1)$, for example, $h = 0.5$. Then at any $\pi \in \Pi$, $(\max(0.15, \pi - h), \min(\pi + h, 0.85))$ will contain one break, which makes the restriction (2.26) valid across the sample. In practice, researchers may choose h based on information of the economy. But, if a small h is chosen, there could be a chance that the interval $(\max(\pi_0, \pi - h), \min(\pi + h, 1 - \pi_0))$ contains no break; if a large h is chosen, the interval $(\max(\pi_0, \pi - h), \min(\pi + h, 1 - \pi_0))$ may contain more than one breaks. I'll suggest a large h, as in the Monte Carlo simulations and empirical studies, it is shown that the results are robust to large h.

Next, I set up the conditions under which $\hat{\pi}_j$ is a consistent estimator of π_j^*. Now let X be generated by model (2.15), which has a representation (2.17). And accordingly, I denote $r^{pre/post}(\pi)$, $G_\pi^{pre/post}$ and $\Theta_\pi^{pre/post}$, for any $\pi \in \Pi \subset (0, 1)$. Finally, let $m = \min(N, T)$, $M = \max(N, T)$. I replace Assumptions 1 and 2 the following assumptions.

Assumption 1′. *For $j = 1, 2, \ldots, q$ $T_j^* = \lfloor \pi_j^* T \rfloor$, where $0 < \pi_1^* < \pi_2^* < \cdots < \pi_q^* < 1$.*

Assumption 2′. *(i) Let $\mu_{NT,k}^{whole} = \psi_k \left[\frac{\Theta'\Theta}{N} \frac{G'G}{T} \right]$ for $k = 1, \ldots, r^{whole}$. Then for each $k = 1, \ldots, r^{whole}$, $\operatorname{plim}_{m\to\infty} \mu_{NT,k}^{whole} = \mu_k^{whole}$, and $0 < \mu_k^{whole} < \infty$. (ii) For each π_j, let $\mu_{NT,k}^{pre}(\pi_j^*) = \psi_k \left[\frac{\Theta_{\pi_j^*}^{pre\prime} \Theta_{\pi_j^*}^{pre}}{N} \frac{G_{\pi_j^*}^{pre\prime} G_{\pi_j^*}^{pre}}{\lfloor \pi_j^* T \rfloor} \right]$ for $k = 1, \ldots, r^{pre}(\pi_j^*)$, and $\mu_{NT,k}^{post}(\pi_j^*) = \psi_k \left[\frac{\Theta_{\pi_j^*}^{post\prime} \Theta_{\pi_j^*}^{post}}{N} \frac{G_{\pi_j^*}^{post\prime} G_{\pi_j^*}^{post}}{T - \lfloor \pi_j^* T \rfloor} \right]$ for $k = 1, \ldots, r^{post}(\pi_j^*)$. Then for each $k = 1, \ldots, r^{pre/post}(\pi_j^*)$, $\operatorname{plim}_{m\to\infty} \mu_{NT,k}^{pre/post}(\pi_j^*) = \mu_{k,j}^{pre/post}$, and $0 < \mu_{k,j}^{pre/post} < \infty$, for $j = 1, \ldots, q$. (iii) $r^{pre/post}(\pi_j^*)$ is finite for all j.*

Assumption $1'$ is a standard requirement to permit the development of an asymptotic theory and allows the structural breaks to be asymptotically distinct. It considers the asymptotic experiments under the assumption that each segment increases proportionately as the sample size increases. I refer to the quantities $\pi^* = (\pi_1^*, \ldots, \pi_q^*)$ as the break fractions and let $\pi_0^* = 0$ and $\pi_{q+1}^* = 1$. Assumption $2'$ is an extension of Assumption 2. In Assumption $2'$(ii), I extend Assumption 2(ii) such that at each structural break, the cumulative effect of the "least influential factor" of the pre- and post-break subsamples diverges in probability to infinity as the sample size grows.

Theorem 4. *Let $x_{i,t}$ be generated by model* (2.15)*. Suppose Assumptions 1′, 2′ and 3-5 hold, $\hat{\pi}_j$ is defined in equation* (2.25)*, then $\hat{\pi}_j = \pi_j^* + O_p(T^{-1})$ for $j = 1, 2, \ldots, q$.*

2.6 Monte Carlo simulation

In this section, a Monte Carlo simulation study is conducted to verify the small sample properties of this new methodology of detecting structural breaks. I evaluate for different simulation designs the frequency by which the true structural break can be recovered. I compare this new methodology to the method of [Baltagi et al., 2017].

2.6.1 Simulation design

For the single break case, the data is generated from

$$x_{it} = \begin{cases} f_{0,t}'\lambda_{0,i} + f_{1,t}'\lambda_{1,i} + e_{it} & \text{if} \quad 1 \leqslant t \leqslant \lfloor \pi^* T \rfloor \\ f_{0,t}'\lambda_{0,i} + f_{1,t}'\lambda_{2,i} + e_{it} & \text{if} \quad \lfloor \pi^* T \rfloor + 1 \leqslant t \leqslant T \end{cases}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. I consider different combinations of N and T given by $(N, T) \in \{(300, 400), (100, 100)\}$. The structural break is chosen as $T^* = \pi^* T$ with $\pi^* = 0.5$. The number of factors is equal to $r_0 = 3$ and $r_1 = 2$, corresponding to the dimensions of $f_{0,t}$ and $f_{1,t}$ respectively.

For the multiple breaks case, I generate the data from

$$x_{it} = \begin{cases} f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{1,i} + f'_{2,t}\lambda^2_{1,i} + e_{it} & \text{if} \quad 1 \leqslant t \leqslant \lfloor \pi^*_1 T \rfloor \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{2,i} + f'_{2,t}\lambda^2_{1,i} + e_{it} & \text{if} \quad \lfloor \pi^*_1 T \rfloor + 1 \leqslant t \leqslant \lfloor \pi^*_2 T \rfloor + 1 \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda^1_{2,i} + f'_{2,t}\lambda^2_{2,i} + e_{it} & \text{if} \quad \lfloor \pi^*_2 T \rfloor + 1 \leqslant t \leqslant T \end{cases}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. I consider combinations of N and T given by $(N, T) \in \{(300, 400)\}$. Two structural breaks are considered and the structural breaks are chosen as $T^*_1 = \pi^*_1 T$ with $\pi^*_1 = 0.35$ and $T^*_2 = \pi^*_2 T$ with $\pi^*_2 = 0.65$. The number of factors is equal to $r_0 = 3$, $r_1 = 2$, and $r_2 = 2$, corresponding to the dimensions of $f_{0,t}$, $f_{1,t}$ and $f_{2,t}$ respectively.

The individual factors are all generated by the autoregressive process

$$f_{j,t,p} = \rho f_{j,t-1,p} + u_{j,t,p}, \quad \text{for} \quad t = 1, \ldots, T, \quad j = 0, \ldots, q, \quad p = 1, \ldots, r_j$$

where q is the number of the breaks, i.e., $q = 1$ for single break case, $q = 2$ for multiple breaks case; ρ is the autoregressive parameter and $u_{j,t,p}$ is the disturbance. I consider different values of ρ given by $\rho \in \{0, 0.5\}$.

The idiosyncratic shocks $e_{i,t}$ are generated from

$$e_{i,t} = \alpha e_{i,t-1} + v_{i,t}, \quad \text{for} \quad t = 1, \ldots, T, \quad i = 1, \ldots, N$$

where $v_t = (v_{1,t}, \ldots, v_{N,t})'$ and $v_t \sim NID(0, \Omega)$ for $t = 2, \ldots, T$. The elements of the matrix Ω are given by $\Omega_{ij} = \beta^{|i-j|}$. The parameters α and β capture the serial correlation and cross-sectional dependence in the idiosyncratic shocks, respectively. I consider different combinations of α and β given by $\alpha \in \{0, 0.2\}$, $\beta \in \{0, 0.2\}$.

To investigate the different types of structural breaks, I consider the following scenarios.

(i) **Scenario 1: A structural break in the factor loadings:** I assume there is a single break and generate data based on the single break model. I generate $\lambda_{0,i}$, $\lambda_{1,i}$ and $\lambda_{2,i}$ from normal distribution respectively, i.e. after the break, the loadings of r_1 factors($f_{1,t,p}$) change from $\lambda_{1,i}$ to $\lambda_{2,i}$. And for all p, I generate $u_{j,t,p} \sim NID(0, 1)$ for $t = 1, \ldots, T$, i.e. there is no break in the volatility of the factors.

(ii) **Scenario 2: Factors (dis)appearing:** I assume there is a single break and generate data based on the single break model. I generate $\lambda_{0,i}$ and $\lambda_{1,i}$ from normal distribution respectively, and set $\lambda_{2,i} = 0$, i.e. after the break, r_2 factors($f_{2,t,p}$) disappear. And for all p, I generate $u_{j,t,p} \sim NID(0,1)$ for $t = 1, \ldots, T$, i.e. there is no break in the volatility of the factors.

(iii) **Scenario 3: Multiple structural breaks:** I assume there are two breaks and generate data based on the multiple breaks model. I generate $\lambda_{0,i}$, $\lambda^1_{1,i}$ and $\lambda^1_{2,i}$ from normal distribution respectively, i.e. after the first break, the loadings of $f_{1,t,p}$ change from $\lambda^1_{1,i}$ to $\lambda^2_{2,i}$. I generate $\lambda^2_{1,i}$ from normal distribution, and set $\lambda^2_{2,i} = 0$, i.e. after the second break, r_2 factors($f_{2,t,p}$) disappear. And for all p, I generate $u_{j,t,p} \sim NID(0,1)$ for $t = 1, \ldots, T$, i.e. there is no break in the volatility of the factors.

(iv) **Scenario 4: A structural break in the volatility of the factors:** Data are generated based on the single break model. I set $\lambda_{2,i} = \lambda_{1,i} = 0$, so that there is no break in the factor loadings and the total number of factors is $r = r_0$. But, for all $p = 1, \ldots, r$ I generate $u_{j,t,p} \sim NID(0,1)$ for $t = 1, \ldots, \tau^*$ and $u_{j,t,p} \sim NID(0,\theta)$ for $t = \tau^* + 1, \ldots, T$. I set $\tau^* = 0.5$ and $\theta = 4$.

For each simulation design, I generate $J = 1000$ panels, set the trimming $\Pi = 0.15 : \frac{1}{T} : 0.85$ and $\Delta\pi = \frac{1}{T}$. I calculate $\hat{Q}(.)$ at each possible period and get estimated $\hat{\pi}$ based on $\Delta\hat{Q} = \hat{Q}(\frac{t}{T}) - \hat{Q}(\frac{t-1}{T})$, for $t = \lfloor 0.15T \rfloor, \lfloor 0.15T \rfloor + 1, \ldots, \lfloor 0.85T \rfloor$.

2.6.2 Monte Carlo results

The main outcome criterion is the frequency by which the estimators (2.14) and (2.25) are within certain periods around the true structural breaks. Table 2.2 displays the main results of Scenario 1-3. Table 2.2(a) and 2.2(b) display the results of the single break case, where I compare this method (denoted as ER and displayed in the first big column) with

[Baltagi et al., 2017] (denoted as BKW and displayed in the second big column). Table 2.2(c) displays the results of the multiple breaks case, where I display the results of each break in each column. I find that this new method performs well in moderately large samples, for all combinations of parameters. In moderately large samples, it performs better than [Baltagi et al., 2017] most of the time, especially in the case of factors disappearing. Small sample results for Scenario 1-2 are provided in Table **??** in Appendix 1.9.

Besides, I provide the average criterions $\hat{Q}$ and $|\Delta\hat{Q}|$ as well as the histogram of the estimated structural break across 1000 iterations for each scenario with different combinations of (ρ, α, β) in Appendix 1.9. It confirms that the criterion has a peak around the structural breaks as discussed in the theory, the histogram of the estimated structural break are distributed around the true structural breaks, and that this criterion will not treat a structural break in the volatility of factors as a break in the factor loadings.

2.7 Structural breaks in US inflation

To illustrate the new methodology, I investigate whether there are structural breaks in a large panel of disaggregated US prices that was previously studied by [Reis and Watson, 2010]. The price data, ranging from 1959:M1-2017:M7[4], comprise monthly chained price indices for US personal consumption expenditures by major type of product and expenditure. Inflation is measured in percentage points at an annual rate prorated each month: $\pi_{i,t} = 100 \times ln(P_{i,t}/P_{i,t-12})$, where $P_{i,t}$ are monthly prices. And I end up with a panel consisting of $N = 167$ monthly time series, which refer to price series for goods and services at the highest available level of disaggregation and covering $T = 690$ periods, after taking first difference to ensure stationarity. I aim to answer the following questions. First, I question whether there are structural breaks in the factor loadings in this large panel of disaggregated US prices. Second, I adopt

[4]Jan 1959 to Jul 2017

Table 2.2: Main Results

(a) A structural break in the factor loadings, $T = 400, N = 300, \pi^* = 0.5$

	ER		BKW	
(ρ, α, β)	$P(\lvert\hat{\pi}^{ER} - \pi^*\rvert < 0.01)$	$P(\lvert\hat{\pi}^{ER} - \pi^*\rvert < 0.025)$	$P(\lvert\hat{\pi}^{BKW} - \pi^*\rvert < 0.01)$	$P(\lvert\hat{\pi}^{BKW} - \pi^*\rvert <$
$(0, 0, 0)$	0.853	0.998	0.87	0.988
$(0, 0, 0.2)$	0.839	0.996	0.844	0.978
$(0, 0.2, 0)$	0.803	0.998	0.848	0.982
$(0, 0.2, 0.2)$	0.771	0.996	0.854	0.974
$(0.5, 0, 0)$	0.897	1	0.834	0.968
$(0.5, 0, 0.2)$	0.879	0.999	0.83	0.968
$(0.5, 0.2, 0)$	0.872	0.999	0.836	0.964
$(0.5, 0.2, 0.2)$	0.852	0.996	0.816	0.966

(b) Factors disappearing, $T = 400, N = 300, \pi^* = 0.5$

	ER		BKW	
(ρ, α, β)	$P(\lvert\hat{\pi}^{ER} - \pi^*\rvert < 0.01)$	$P(\lvert\hat{\pi}^{ER} - \pi^*\rvert < 0.025)$	$P(\lvert\hat{\pi}^{BKW} - \pi^*\rvert < 0.01)$	$P(\lvert\hat{\pi}^{BKW} - \pi^*\rvert <$
$(0, 0, 0)$	0.84	0.995	0.744	0.9
$(0, 0, 0.2)$	0.831	0.993	0.72	0.926
$(0, 0.2, 0)$	0.794	0.982	0.752	0.926
$(0, 0.2, 0.2)$	0.789	0.992	0.726	0.908
$(0.5, 0, 0)$	0.884	0.998	0.726	0.89
$(0.5, 0, 0.2)$	0.868	0.993	0.702	0.902
$(0.5, 0.2, 0)$	0.856	0.995	0.712	0.91
$(0.5, 0.2, 0.2)$	0.841	0.994	0.71	0.892

(c) Multiple structural breaks, $T = 400, N = 300, \pi_1^* = 0.3, \pi_2^* = 0.7$

	Break 1		Break 2	
(ρ, α, β)	$P(\lvert\hat{\pi}_1^{ER} - \pi_1^*\rvert < 0.01)$	$P(\lvert\hat{\pi}_1^{ER} - \pi_1^*\rvert < 0.025)$	$P(\lvert\hat{\pi}_2^{ER} - \pi_2^*\rvert < 0.01)$	$P(\lvert\hat{\pi}_2^{ER} - \pi_2^*\rvert < 0.02$
$(0, 0, 0)$	0.841	0.998	0.818	0.989
$(0, 0, 0.2)$	0.834	0.995	0.783	0.994
$(0, 0.2, 0)$	0.817	0.997	0.752	0.991
$(0, 0.2, 0.2)$	0.757	0.984	0.715	0.986
$(0.5, 0, 0)$	0.877	1	0.843	0.996
$(0.5, 0, 0.2)$	0.869	0.999	0.854	0.998
$(0.5, 0.2, 0)$	0.843	0.996	0.802	0.996
$(0.5, 0.2, 0.2)$	0.859	0.993	0.794	0.993

Note: In the single break case, I use $\hat{\pi}^{ER}$ and $\hat{\pi}^{BKW}$ to denote the estimators using the new method and using [Baltagi et al., 2017] respectively; in the multiple breaks case, I use $\hat{\pi}_j^{ER}, j = 1, 2$ to denote the estimators using the new method. This table displays the frequency by which each estimator is within δT ($\delta \in \{0.01, 0.025\}$) periods around the true structural breaks.

conventional macro-econometric methods to determine how the underlying structure changes before and after a break. Third, I question whether there are structural breaks in the volatility of the factors in this panel.

2.7.1 Detecting structural breaks in the factor loadings

I apply the estimator for multiple breaks to the inflation panel with the trimming $\pi_j^* \in \Pi = (0.15, 0.85), \forall j = 0, 1, \ldots$. The starting and ending points correspond to the time period from Oct 1965 to Oct 2011. The whole period covers events such as oil price shocks (1973, 1979), the Great Moderation period of the 1980s, and the 2008 financial crisis.

Figure 2.1 displays the $|\Delta\hat{Q}|$ of the inflation panel from 1960 to 2017, with h set to equal 0.5 and $\lfloor \Delta\pi T \rfloor = 12$.[5] I identify two big breaks in the panel, which correspond to the two peaks in Figure 2.1.[6] The estimated structural breaks are around May 1974 and Nov 2008, which are identified by picking the period corresponding to the highest value among each cluster of peaks. I find that each estimated structural break corresponds to some economic/financial event. The break of May 1974 is close to the 1973 oil price shock, while the Nov 2008 break coincides with the 2008 financial crisis. Some other economic/financial events seem not to point to any structural breaks here, however, for example, the 1979 oil price shock and the 1984 Great Moderation. This might lead researchers to consider what could influence the underlying structure of disaggregated inflations.

[5]This criterion is robust to the choice of big h as well as different choices of $\Delta\pi$. Appendix **??** provides criterions measured with different values of h and $\Delta\pi$ ($\lfloor \Delta\pi T \rfloor = 1, 4, 6, 12$ which correspond to a month, a quarter, half a year and a year respectively).

[6]When $\lfloor \Delta\pi T \rfloor$ is small, there might be clusters of peaks. I identify each cluster of peaks as one break, as I assume the adjacant breaks are not too close.

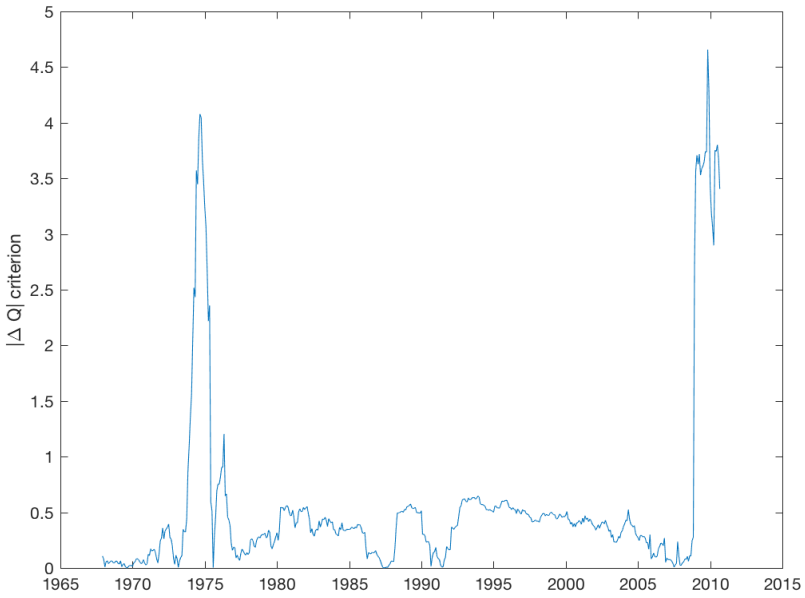


Figure 2.1: Criterions $|\Delta\hat{Q}|$ from Inflation Panel, 1960 to 2017

2.7.2 Conventional analysis of subsamples

I split the whole sample into three subsamples: 1960:M2-1973:M12, 1974:M9-2008:M6, and 2009:M3-2017:M7, excluding several periods around each estimated break point. Figure 2.2 displays the first 10 largest eigenvalues of each subsample, confirming that the structure of the eigensystem of each subsample does differ.

To get a clearer view of how the co-movement of the disaggregated inflation panel changes at each break, I regress each series on the first common factor for each subsample, and display the R^2 value of each in Figure 2.3.[7] Figure 2.3 implies that the structure of the co-movement

[7]In Figure 2.3, the x-axis refers to the ID of each disaggregated series, and the dotted vertical lines divide the series into good inflation series and service inflation series. There

of these inflation index series changes over time. For example: (1) Series 116 (Fuel Oil in Gasoline and Other Energy Goods) contributes only a small part to the first factor in subsamples 1 and 2, but it contributes greatly to the first factor in subsample 3; (2) Almost 90% of the volatility of series 180-182 (Hospitals in Health Care) can be explained by the first factor in subsample 1, while, in subsamples 2 and 3, this percentage crashes greatly; (3) Series 221-223 (Gambling in Recreation Service) have an increasing contribution to the first factor.

2.7.3 Are there structural breaks in the volatility of the factors?

It is very surprising that some economic/financial events are not accompanied by structural breaks, especially the 1984 Great Moderation, which is identified as a common break in the literature (e.g., see [Stock and Watson, 2009]). However, even though the empirical result implies no structural break in the factor loadings around 1984, there could still be a structural break in the volatility of the factors. As mentioned at the end of Section 2.4, by combining this new method with the literature ([Han and Inoue, 2015], [Baltagi et al., 2017]), which adopts a similar model framework to eq (2.2)(2.3), a structural break in the factor loadings and a structural break in the volatility of the factors can be distinguished. If the methods in [Han and Inoue, 2015] and [Baltagi et al., 2017] capture a break when this new method does not, then there is a structural break in the volatility of the factors. Below, I conduct a three-step analysis on periods 1974:M5-2008:M11 and 2008:M11-2017:M7, which researchers may be interested in, but I find no structural breaks.

- Step 1: Apply test in [Han and Inoue, 2015] on the targeted period. If their null hypothesis is not rejected, skip Step 2-3 and conclude that there is neither structural break in the factor loading nor structural break in the volatility of the factors during this period. If their null hypothesis is rejected, go to Step 2.

are 167 series. Note that the IDs are distinct from the series numbers.

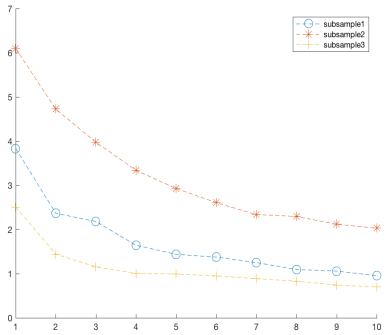
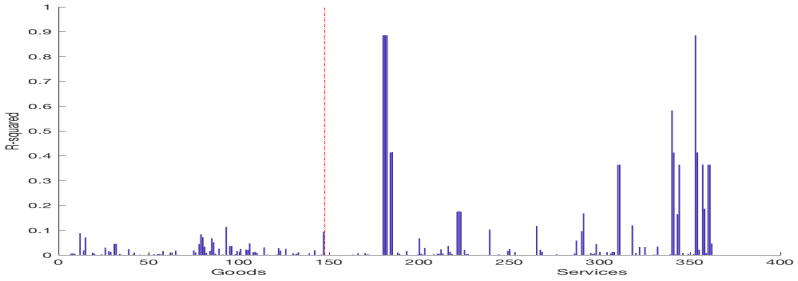
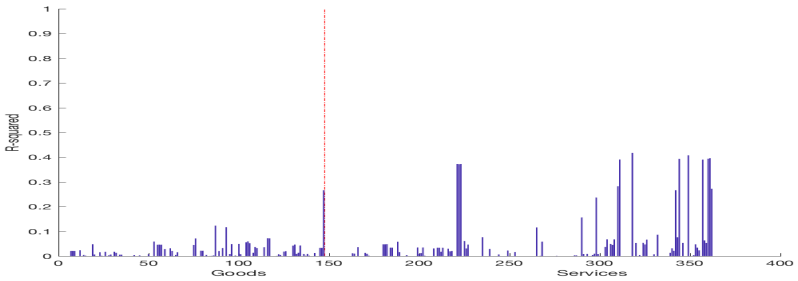


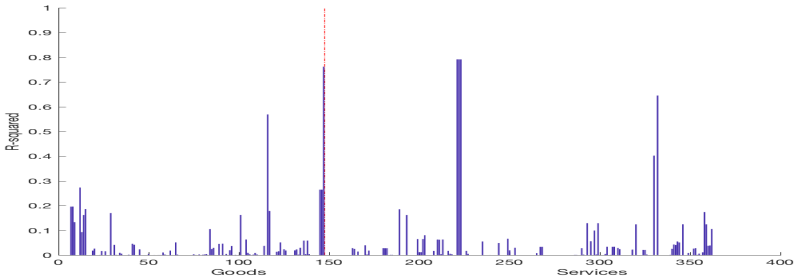
Figure 2.2: First 10 Eigenvalues of 3 Subsamples



(a) R^2 on Subsample 1 (1960M2-1973M12), First Factor



(b) R^2 on Subsample 2 (1974M9-2008M6), First Factor



(c) R^2 on Subsample 3 (2009M3-2017M7), First Factor

Figure 2.3: R^2 on Subsamples

- Step 2: Apply the estimation method in [Baltagi et al., 2017] on the targeted period and conclude that there is a structural break in the volatility of the factors around the estimated structural break.

- Step 3: Apply test in [Han and Inoue, 2015] on subsamples divided at the estimated structural break in Step 2 to check whether there are multiple breaks. If yes, repeat Step 2-3 until there are no more breaks in any subsamples.[8]

Table 2.3: Rejection results of testing for unknown break date

(a) 1974M5-2008M11

	sup-Wald	exp-Wald	mean-Wald	sup-LM	exp-LM	mean-LM
White	1	1	1	1	1	1
HAC	0	1	1	1	1	1

(b) 1974M5-1991M1

	sup-Wald	exp-Wald	mean-Wald	sup-LM	exp-LM	mean-LM
White	0	0	0	1	1	0
HAC	0	0	0	0	0	0

(c) 1991M1-2008M11

	sup-Wald	exp-Wald	mean-Wald	sup-LM	exp-LM	mean-LM
White	0	0	0	0	0	0
HAC	0	0	0	0	0	0

(d) 2008M11-2017M7

	sup-Wald	exp-Wald	mean-Wald	sup-LM	exp-LM	mean-LM
White	0	0	0	0	0	0
HAC	0	0	0	0	0	0

Note: The result is at 5% significance level. 1 indicates rejection; 0 otherwise. White refers to White's heteroskedastic robust variance estimator and HAC refers to using HAC Bartlett kernel.

Period 1974M5-2008M11

Period 1974:M5-2008:M11 experienced several economic/financial events, such as the 1979 oil shock, the 1984 Great Moderation and the early 1990s

[8] Step 3 is conducted as both [Han and Inoue, 2015] and [Baltagi et al., 2017] assume there is only one break during the targeted period.

recession, but there is no consensus in the literature on exactly when there are structural breaks in the macro panel data. [Stock and Watson, 2009] identify 1984 as a common break in the factor model, while [Chen et al., 2014] find a structural break around the 1979 oil shock but no structural break around 1984. The result in Table 2.3(a), 2.3(b), 2.3(c) can shed some light on these discrepancies. Table 2.3(a) gives the rejection results for period 1974:M5-2008:M11 at the 5% significance level. For period 1974:M5-2008:M11, all the tests except the $sup - Wald$ test using an HAC Bartlett kernel reject the null hypothesis that there is no break at the 5% significance level. So, for this period, I apply the estimation method of [Baltagi et al., 2017] and get the estimated structural break at 1991:M1. Additionally, Table 2.3(b) and 2.3(c) give the rejection results for periods 1974:M5-1991:M1 and 1991:M1-2008:M11 at the 5% significance level. For period 1974:M5-1991:M1, all tests except the $sup/exp - LM$ tests using White's heteroskedastic robust variance estimator do not reject the null hypothesis that there is no break at the 5% significance level. For period 1991:M1-2008:M11, all the tests do not reject the null hypothesis that there is no break at the 5% significance level. Therefore, I conclude that there is only one structural break in the volatility of the factors around 1991:M1, which corresponds to the early 1990s recession.

Period 2008M11-2017M7

Period 2008:M11-2017:M7 experienced the zero lower bound (ZLB) period, from January 2009 through December 2015, during which the federal funds rate reached its ZLB. Recently, a growing body of literature has aimed to empirically evaluate the effectiveness of various unconventional monetary policies. For example, [Debortoli et al., 2018] finds no significant changes in the estimated responses over the period when the federal funds rate reached the ZLB. Considering that inflation plays an important role in the Taylor rule, which can be greatly influenced by monetary policy, the result in Table 2.3(d) sheds light on empirically evaluating the effectiveness of various unconventional monetary policies. Table 2.3(d) gives the rejection results for period 2008:M11-2017:M7 at the 5% signif-

icance level. All the tests do not reject the null hypothesis that there is no break at the 5% significance level. Thus, I conclude that there is neither a structural break in the factor loading nor a structural break in the volatility of the factors during this period. The empirical result implies that the inflation panel is stable during 2008:M11-2017:M7, which is consistent with the hypothesis of "perfect substitutability" between conventional and unconventional monetary policies, suggesting that unconventional monetary policy works as well as conventional monetary policy in terms of stabilizing the economy.

2.8 Conclusions

This article proposes a novel methodology for identifying and estimating structural breaks, which occurr at unknown common dates, in the factor loadings of a high dimensional approximate factor model with an unknown number of latent factors. The approach is robust to structural changes in the volatility of factors, uncomplicated to implement, and can easily be extended to multiple structural breaks. Furthermore, combining this new methodology with the existing methodology allows to distinguish between a structural change in the factor loadings and a structural change in the volatility of factors. In an empirical study of disaggregated US inflation series, I find two breaks in the factor loadings around the 1973 oil price shock and the 2008 financial crisis, one break in the volatility of the factors in January 1991, and no break from November 2008 to July 2017.

Chapter 3

IMPULSE RESPONSES ESTIMATION UNDER UNSTABLE LOCAL PROJECTIONS

(joint with Atsushi Inoue and Barbara Rossi)

3.1 Introduction

This paper introduces time variation into the *local projections* framework and proposes an impulse responses estimation methodology under unstable *local projections*.

Impulse responses have been important tools to estimate the effect on the economy of unanticipated structural disturbances, i.e., structural shocks, in empirical macroeconomics analysis. Conventional impulse response estimators are obtained by recursively iterating vector autoregressions (VARs) to characterize the structure of successive observations, see [Sims, 1980] and [Stock and Watson, 2016] for detailed identification

strategies. As VARs may indeed be a significantly misspecified representation of the underlying DGP,[1] an increasingly widespread alternative to estimating impulse responses from VARs is the method of *local projections* in [Jorda, 2005]. The idea of *local projections* is to projects future outcomes on current covariates for each forecast horizon, see, for instance, [Ramey, 2016], [Angrist et al., 2018], [Miranda-Agrippino and Ricco, 2018], and [Stock and Watson, 2018]. *local projections* are favored due to its flexible specifications and easy accommodation of nonlinearities that may be impractical in a multivariate context, as well as its simple estimation and inference.

As there is substantial evidence of instabilities both in the systematic part of monetary policy[2] and in the transmission mechanism, i.e., how macroeconomics variables respond to structural shocks, it is natural to model and estimate time variation in multivariate linear structures, in both the coefficients and the variances. Much work has been done following the line of time-varying parameter vector autoregressions (TVP-VAR), see, for instance, [Primiceri, 2005] and [Koop and Korobilis, 2010] for time-varying parameter Bayesian vector autoregressions.[3]

However, fewer works have studied the time variation in impulse responses under unstable *local projections*. [Auerbach and Gorodnichenko, 2012] and [Ramey and Zubairy, 2018] find evidence that impulse responses under *local projections* vary under different economic conditions. But their researches focus on time variation in *local projections* with discrete breaks, i.e., a small number of switching regimes, rather than a smooth stochastic evolution of the parameters.

This paper, from an estimation standpoint, introduces local time variation into the *local projections* framework, in the sense that the instability in *local projections* will be detected with a probability smaller than 1 even in the limit, and proposes an impulse responses estimation methodology

[1] See, for instance, [Zellner and Palm, 1974] and [Wallis, 1977].

[2] See, for instance, [Cogley and Sargent, 2005] and [Boivin and Giannoni, 2006].

[3] Other research includes [Canova, 1993], [Stock and Watson, 1996], [Uhlig, 1997], [Boivin, 2005], [Cogley and Sargent, 2005] who model and estimate VARs with drifting coefficients and/or time varying variances.

under unstable *local projections*. Importantly, the local time variation is considered in both the coefficients and the variances, thus modeling and estimating changes both in structural shocks and in the transmission mechanism. Documenting the time variation in impulse responses, i.e., the path estimators under unstable *local projections*, is useful for several purposes. First, the path estimators themselves are helpful in describing the potential sources of instability. Second, the path estimators under *local projections* have insights on how macroeconomics variables respond to structural shocks at different periods, implying different transmission mechanisms under different economic conditions. Third, the endpoint of the parameter path provides useful and the latest information from the perspective of forecasting.

The impulse responses estimation methodology introduced in this paper builds upon [Muller and Petalas, 2010] path estimators in a multivariate system. [Muller and Petalas, 2010] consider an unstable time series model with a log-likelihood function of the form $\sum_{t=1}^{T} l_t(\theta_t) = \sum_{t=1}^{T} l_t(\theta + \delta_t)$, while its corresponding stationary and stable model has the same likelihood with time invariant parameters $\theta \in \Theta \subset \mathbb{R}^k$. [Muller and Petalas, 2010] derive asymptotically weighted average risk (WAR) minimizing path estimators for $\{\theta_t\}_{t=1}^{T} = \{\theta + \delta_t\}_{t=1}^{T}$ and weighted average power (WAP) maximizing parameter stability test statistics, assuming an approximately stationary model and a weighting function for the variability $\{\delta_t\}_{t=1}^{T}$ that is a (demeaned) multivariate Gaussian random walk. The covariance matrix of the approximate posterior of the path estimators is further provided as a WAR minimizing interval estimator to gain some sense of the accuracy of the path estimators. In the context of unstable *local projections*, proper elements in the path estimators coincide with the impulse responses of the corresponding structural shocks, while proper elements in the covariance matrix of the approximate posterior of the path estimators help gain some sense of the accuracy of the impulse responses.

The Monte Carlo scetion contain evidence illustrating that [Muller and Petalas, 2010] asymptotically WAR minimizing path estimators and WAP maximizing parameter stability test statistics perform well in the unstable

local projections framework with flexible specifications.

To illustrate the estimation methodology, in this paper, we revisit the small quarterly time-varying SVAR model of the U.S. economy studied in [Primiceri, 2005]. The findings can be summarized as follows. First, the time-varying standard deviation of the identified monetary policy shocks has a similar pattern with that in [Primiceri, 2005], confirming the path of the relative importance and changes of the monetary policy. However, the path estimators indicate that the estimated coefficients experience much time variation under these different economic conditions, especially around 1975:Q1, 1981:Q3, and 2008:Q3, corresponding to the NBER business cycle trough date, the NBER business cycle peak date, and the financial crisis respectively, while [Primiceri, 2005] time-varying SVAR indicates little time variation in the estimated coefficients. Furthermore, compared with impulse responses computed based on the time-varying SVAR model in [Primiceri, 2005], impulse responses under unstable *local projections* share similar trends most of the time, but are generally less smooth and imply that the monetary policy shocks have effects of larger magnitude and that last longer.

The remainder of this paper is organized as follows. Section 3.2 discusses the unstable *local projections* framework and the impulse responses estimation methodology. Section 3.3 presents the Monte Carlo simulation studies. Section 3.4 revisits the small quarterly time-varying SVAR model of the U.S. economy. Section 3.5 concludes.

3.2 Impulse Responses under Unstable Local Projections

As described in [Hamilton, 1994] and [Koop et al., 1996], an impulse response can be defined as the difference between two forecasts and statistically obtains the best, mean-squared, h-step ahead predictions:

$$IR(t,h,d_i) = \mathrm{E}[Y_{t+h}|v_t = d_i; X_t] - \mathrm{E}[Y_{t+h}|v_t = 0; X_t], \qquad h = 0, 1, \ldots \tag{3.1}$$

where Y_t is a $K \times 1$ vector that $Y_t = [y_{t,1}, y_{t,2}, \ldots, y_{t,K}]'$, $X_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$, v_t is the $K \times 1$ vector of reduced-form disturbance, and d_i represents the structural shock to the i^{th} element in Y_t. Note that the identification strategy is not the focus of this paper.

An impulse response can be calculated by recursively iterating a model to characterize the structure of successive observations, e.g., recursively iterating VARs, or based on direct linear regressions of future outcomes on current covariates for each forecast horizon, e.g., local projections in [Jorda, 2005] considering projecting Y_{t+h} onto the linear space generated by $(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$:

$$Y_{t+h} = C + \Theta_1^{(h)} Y_{t-1} + \cdots + \Theta_p^{(h)} Y_{t-p} + U_t^{(h)}, \tag{3.2}$$

where C is a $K \times 1$ vector of constants, and $\Theta_j, j = 1, \ldots, p$ are matrices of coefficients for each lag j and horizon h, and the residuals $U_t^{(h)}$ are moving averages of the past forecast errors and therefore serially correlated that $U_t^{(h)} \sim \mathcal{N}(O, \Sigma_u^{(h)})$.

The collection of regressions in eq (3.2) for $h = 0, 1, \ldots, H$ is denoted as *local projections*. Thus, the impulse responses computed based on *local projections* are:

$$\hat{IR}(t, h, d_i) = \hat{\Theta}_1^{(h)} d_i, \qquad h = 0, 1, \ldots, H \tag{3.3}$$

with the normalization on $\Theta_1^{(0)} = I$. The *local projections* framework does not require correct specification and estimation of the unknown true multivariate data generating process.

Let θ, a $q \times 1$ vector, be the vectorized verion of the coefficient matrice $\Theta_j^{(h)}, j = 1, \ldots, p$ and the variance matrix of $U_t^{(h)}$ in eq (3.2) for a certain $h = 0, 1, \ldots, H$. Then eq (3.2) is a stationary and stable system of *local projections* with a known log-likelihood function of the form $\sum_{t=1}^{T} l_t(\theta)$, with parameter $\theta \in \Theta \subset \mathbb{R}^k$.

Consider the corresponding unstable *local projections* which has the same likelihood with time-varying parameters $\{\theta_t\}_{t=1}^T = \{\theta + \delta_t\}_{t=1}^T$:

$$Y_{t+h} = C + \Theta_{1,t+h}^{(h)} Y_{t-1} + \cdots + \Theta_{p,t+h}^{(h)} Y_{t-p} + U_t^{(h)}, \tag{3.4}$$

where $\Theta_{j,t+h}, j = 1, \ldots, p$ are time-varying matrices of coefficients for each lag j and horizon h.

As shown in [Muller and Petalas, 2010], the sample information about θ and $\{\delta_t\}_{t=1}^T$ is approximately independent and described by the pseudo model

$$\begin{aligned} \hat{\theta} &= \theta + T^{-1/2}\hat{H}^{-1}\nu_0 \\ s_t(\hat{\theta}) &= \hat{H}\delta_t + \nu_t, \quad \nu_t \sim \text{i.i.d.}\mathcal{N}(0, \hat{H}), \quad t = 1, \ldots, T \end{aligned} \tag{3.5}$$

where $s_t(\theta) = \partial l_t(\theta)/\partial\theta, t = 1, \ldots, T$ is the sequence of $q \times 1$ score vectors, and matrix $\hat{H}$ is defined as $\hat{H} = \frac{1}{T}\sum_{t=1}^T h_t(\hat{\theta})$ with $h_t(\theta) = -\partial s_t(\theta)/\partial\theta, t = 1, \ldots, T$ the sequence of $q \times q$ Hessians.

[Muller and Petalas, 2010] derive asymptotically weighted average risk (WAR) minimizing path estimators $\{\hat{\theta}_t\}_{t=1}^T$ and weighted average power (WAP) maximizing parameter stability test statistics $qLL(10)$ assuming an approximately stationary model and a weighting function for $\{\delta_t\}_{t=1}^T$ that is a (demeaned) multivariate Gaussian random walk as follows:

1. For $t = 1, \ldots, T$, let x_t and $\tilde{y}_t$ be all the elements of $\hat{H}^{-1}s_t(\hat{\theta})$ and $\hat{H}\hat{V}^{-1}s_t(\hat{\theta})$, respectively.

2. For $c_i \in C = \{0, 5, 10, \ldots, 50\}$,[4] $i = 1, \ldots, 10$, compute

   (a) $r_i = 1 - \frac{c_i}{T}$, $z_{i,1} = x_1$, and $z_{i,t} = r_i z_{i,t-1} + x_t - x_{t-1}, t = 2, \ldots, T$;

   (b) the residuals of $\{\tilde{z}_{i,t}\}_{t=1}^T$ of a linear regression of $\{z_{i,t}\}_{t=1}^T$ on $\{r_i^{t-1}I_q\}_{t=1}^T$;

   (c) $\bar{z}_{i,T} = \tilde{z}_{i,T}$, and $\bar{z}_{i,t} = r_i\bar{z}_{i,t+1} + \tilde{z}_{i,t} - \tilde{z}_{i,t+1}, t = 1, \ldots, T-1$;

[4] For the factor of proportionality $\frac{c^2}{T^2}$, [Muller and Petalas, 2010] suggest a default choice of minimizing WAR relative to an equal-probability mixture of $n_G = 11$ values $c \in \{0, 5, 10, \ldots, 50\}$, which represents the standard deviation of the end point of the random walk weighting function and covers a wide range of magnitudes for the time variation.

(d) $\{\hat{\theta}_{i,t}\}_{t=1}^{T} = \{\hat{\theta} + x_t - r_i \bar{z}_{i,t}\}_{t=1}^{T}$;

(e) $qLL(c_i) = \sum_{t=1}^{T} (r_i \bar{z}_{i,t} - x_t)' \tilde{y}_t$ and $\tilde{w}_i = \sqrt{T(1-r_i^2) r_i^{T-1} / ((1-r_i^{2T}))} exp[-\frac{1}{2} qLL(c_i)]$ (set $\tilde{w}_0 = 1$).

3. Compute $w_i = \tilde{w}_i / \sum_{j=0}^{10} \tilde{w}_j$.

4. The parameter path estimator is given by $\{\hat{\theta}_t\}_{t=1}^{T} = \{\sum_{i=0}^{10} w_i \hat{\theta}_{i,t}\}_{t=1}^{T}$.

5. The statistic $qLL(10)$ tests the null hypothesis of stability of θ and rejects for small values. Critical values depend on q and are tabulated in Table 1 of [Elliott and Muller, 2006].

The covariance matrix of the approximate posterior for θ_t, with weighting function for $\{\delta_t\}_{t=1}^{T}$ and θ interpreted as priors from a Bayesian perspective, is further provided as follows:

$$\Omega_t = \sum_{i=0}^{10} w_i \left(T^{-1} \hat{S}_\theta \kappa_t(c_i) + (\hat{\theta}_{i,t} - \hat{\theta}_t)(\hat{\theta}_{i,t} - \hat{\theta}_t)' \right),$$

where $\hat{S}_\theta = \hat{H}^{-1} \hat{V} \hat{H}^{-1}$, $\kappa_t(c) = \frac{c(1+e^{2c}+e^{2ct/T}+e^{2c(1-t/T)})}{2e^{2c}-2}$, and $\kappa_t(0) = 1$. This approximate posterior distribution is a mixture of multivariate normals $\mathcal{N}(\hat{\theta}_{i,t}, T^{-1} \hat{S}_\theta \kappa_t(c_i)), i = 1, \ldots, 10$, with mixing probabilities w_i. Thus, the confidence interval $\left[\hat{\theta}_{t,j} - 1.96\sqrt{\Omega_{t,jj}}, \hat{\theta}_{t,j} + 1.96\sqrt{\Omega_{t,jj}} \right]$ with $\hat{\theta}_{t,j}$ the j-th element of $\hat{\theta}_t$ and $\Omega_{t,jj}$ the (j,j) element of Ω_t is approximately the 95% equal-tailed posterior probability interval for $\theta_{t,j}$, the j-th element of θ at time t.[5]

Thus, the impulse responses computed based on eq (3.4) are:

$$\hat{IR}(t, h, d_i) = \hat{\Theta}_{1,t+h}^{(h)} d_i, \qquad h = 0, 1, \ldots, H \tag{3.6}$$

[5]This interval can be justified without explicit Bayesian reasoning as a WAR minimizing interval estimator – see Chapter 5.2.5 of Schervish (1995).

where $\hat{\Theta}^{(h)}_{1,t+h}$ are the proper elements in the [Muller and Petalas, 2010] path estimators $\hat{\theta}_{t+h}$. Furthermore, to get some sense of the accuracy of the impulse responses, the confidence bands can be constructed based on Ω_t, the covariance matrix of the approximate posterior for θ_t.

3.3 Monte Carlo Simulation

This section introduces Monte Carlo simulations analyzing performances of qLL statistics testing the null hypothesis of stability of the corresponding parameters in VARs, Muller-Petalas path estimators in VARs, as well as Muller-Petalas path estimators under unstable *local projections*.

3.3.1 Data generating process

Consider the following time-varying parameter VAR (TVP-VAR) as the data generating process (DGP):

$$Y_t = A_t(L)Y_t + u_t = A_{1,t}Y_{t-1} + \cdots + A_{p,t}Y_{t-p} + U_t, \qquad U_t \sim \mathcal{N}(O, \Sigma_u) \tag{3.7}$$

where Y_t is a $K \times 1$ vector that $Y_t = [y_{t,1}, y_{t,2}, \ldots, y_{t,K}]'$, $A_{j,t}, j = 1, \ldots, p$ are $(K \times K)$ time-varying parameter matrices.

Let α_t be the vectorized version of the time-varying coefficient matrices $A_{j,t}, j = 1, \ldots, p$. Consider the following time-varying process:

$$\alpha_t = \alpha_{t-1} + \eta_t, \qquad \eta_t \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(O, \Sigma_\eta \frac{c^2}{T^2}\right) \tag{3.8}$$

where η_t is a $(K^2p \times 1)$ vector. We set $A_{1,0} = A_{2,0} = O$, $\Sigma_\eta = I_{K^2p}$, and c is interpreted as the standard deviation of the end point of the random walk weighting function.

Thus, equation (3.7) is a TVP-VAR(p) process, which degenerates to (i) a univariate process when $K = 1$ and (ii) a time-invariant VAR(p)/AR(p) process when $c = 0$. For simplicity, we exclude constant terms from the model.

3.3.2 QLL statistics in a multivariate setting

Now let's consider a time-invariant model with $c = 0$. This subsection introduces the exercise examining the performance of qLL statistic[6] testing the null hypothesis of stability of the corresponding parameters.

We consider both univariate and multivariate cases. For the univariate case, we consider an AR(p) process, i.e., $K = 1$ and consider the number of lags $p \in \{2, 3, 4\}$. Especially, we set the parameters $A_j = 0.2, j = 1, \ldots, p$ and $U_t \sim \mathcal{N}(0, 1)$. For the multivariate case, we consider a VAR(1) with two variables, i.e., $K = 2$ and $p = 1$. We set the coefficient matrix to be lower triangular $A_1 = \begin{bmatrix} 0.2 & 0 \\ 0.1 & 0.3 \end{bmatrix}$. The error term U_t is drawn from $U_t \sim \mathcal{N}(O, \Sigma_u)$, and we consider two cases: $\Sigma_u = I_K$ and $\Sigma_u = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}$.

For each grouping (K, p), 2000 Monte Carlo replications are conducted, each generating T data points as in Equation (3.7) with $T \in \{2000, 4000, 10000\}$.[7] In each replication, and large sample weighted average power (WAP) maximizing parameter stability test statistic, denoted as $qLL(10)$, are obtained following procedures in [Muller and Petalas, 2010].[8] We focus on the first q elements in all the parameters.[9] For example, if we focus on all the elements in parameter matrices $A_j, j = 1, \ldots, p$, then $q = K^2 \times p$. Besides, critical values for the WAP maximizing parameter stability test, tabulated in Table 1 of [Elliott and Muller, 2006], also depend on q.

Table 3.1-3.2 display the frequency we reject the null hypothesis of stability of the first q parameters in the parameter matrices $A_j, j = 1, \ldots, p$

[6] See details of $qLL(10)$ in [Muller and Petalas, 2010].

[7] We generate $T + 100$ data points and discard the first 100 data points. The very first p data points of y are drawn from Normal distribution, i.e., $y_t \sim \mathcal{N}(O, I_K)$.

[8] The approximately weighted average risk (WAR) minimizing path estimator under truncated quadratic loss with large truncation point is obtained in step 4, and large sample weighted average power (WAP) maximizing parameter stability test statistic is obtained in step 5. See [Muller and Petalas, 2010] for more details.

[9] This notation q corresponds to the notation p in [Muller and Petalas, 2010]. We change the notation because p refers to the lags in this setting.

according to the test statistic $qLL(10)$ for different scenarios. Table 3.1 displays the results for the univariate scenario AR(p). Table 3.2 displays the results for the multivariate scenario VAR(1) with $\Sigma_u = I_K$. Table 3.3 displays the results for the multivariate scenario VAR(1) with $\Sigma_u = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}$. Table 3.2 and 3.3 show that the large sample WAP maximizing parameter stability test statistic $qLL(10)$ performs well in the multivariate setting.

3.3.3 Muller-Petalas path estimators in unstable VARs as well as by LPs

This subsection examines the performance of Muller-Petalas path estimators in unstable VARs, as well as Muller-Petalas path estimators based on *local projections*. Let's consider a TVP-VAR process with different choices of c, and without losing generality, we consider a TVP-VAR(2) from now on, i.e., equation (3.7) with $p = 2$.

Local Projections

Suppose we are interested in h-step ahead impulse response function with $h = 2$. By iterating eq (3.7), we get the following

$$\begin{aligned} Y_{t+2} &= A_{1,t+2}Y_{t+1} + A_{2,t+2}Y_t + U_{t+2} \\ &= A_{1,t+2}\Big(A_{1,t+1}Y_t + A_{2,t+1}Y_{t-1} + U_{t+1}\Big) + A_{2,t+2}Y_t + U_{t+2} \\ &= \Big(A_{1,t+2}A_{1,t+1} + A_{2,t+2}\Big)Y_t + \Big(A_{1,t+2}A_{2,t+1}\Big)Y_{t-1} + \Big(U_{t+2} + A_{1,t+2}U_{t+1}\Big) \\ &= \Theta^{(2)}_{1,t+2}Y_t + \Theta^{(2)}_{2,t+2}Y_{t-1} + U^{(2)}_{t+2} \end{aligned} \tag{3.9}$$

where $\Theta^{(2)}_{1,t+2} = A_{1,t+2}A_{1,t+1} + A_{2,t+2}$, $\Theta^{(2)}_{2,t+2} = A_{1,t+2}A_{2,t+1}$, and $U^{(2)}_{t+2} = U_{t+2} + A_{1,t+2}U_{t+1}$. Thus, $\{U^{(2)}_t\}_{t=1}^T$ is serially correlated.

Similarly, if $h = 3$, we get

$$\begin{aligned}
Y_{t+3} &= A_{1,t+3}Y_{t+2} + A_{2,t+3}Y_{t+1} + U_{t+3} \\
&= A_{1,t+3}\Big(A_{1,t+2}Y_{t+1} + A_{2,t+2}Y_t + U_{t+2}\Big) + A_{2,t+3}Y_{t+1} + U_{t+3} \\
&= \Big(A_{1,t+3}A_{1,t+2} + A_{2,t+3}\Big)\Big(A_{1,t+1}Y_t + A_{2,t+1}Y_{t-1} + U_{t+1}\Big) \\
&\quad + \Big(A_{1,t+3}A_{2,t+2}\Big)Y_t + \Big(U_{t+3} + A_{1,t+3}U_{t+2}\Big) \\
&= \Big(A_{1,t+3}A_{1,t+2}A_{1,t+1} + A_{2,t+3}A_{1,t+1} + A_{1,t+3}A_{2,t+2}\Big)Y_t \\
&\quad + \Big(A_{1,t+3}A_{1,t+2}A_{2,t+1} + A_{2,t+3}A_{2,t+1}\Big)Y_{t-1} \\
&\quad + \Big(U_{t+3} + A_{1,t+3}U_{t+2} + \Big(A_{1,t+3}A_{1,t+2} + A_{2,t+3}\Big)U_{t+1}\Big) \\
&= \Theta^{(3)}_{1,t+3}Y_t + \Theta^{(3)}_{2,t+3}Y_{t-1} + U^{(3)}_{t+3}
\end{aligned} \tag{3.10}$$

where $\Theta^{(3)}_{1,t+3} = A_{1,t+3}A_{1,t+2}A_{1,t+1} + A_{2,t+3}A_{1,t+1} + A_{1,t+3}A_{2,t+2}$, $\Theta^{(3)}_{2,t+3} = A_{1,t+3}A_{1,t+2}A_{2,t+1} + A_{2,t+3}A_{2,t+1}$, and $U^{(3)}_{t+2} = U_{t+3} + A_{1,t+3}U_{t+2} + \Big(A_{1,t+3}A_{1,t+2} + A_{2,t+3}\Big)U_{t+1}$ is serially correlated.

Recursively, we will get

$$Y_{t+h} = \Theta^{(h)}_{1,t+h}Y_t + \Theta^{(h)}_{2,t+h}Y_{t-1} + U^{(h)}_{t+h} \tag{3.11}$$

where $\Theta^{(h)}_{j,t+h}, j = 1, 2$ are functions of $\{A_{j,t}\}_{t=1}^{T}, j = 1, 2$. According to [Jorda, 2005], the h-step ahead impulse response functions by LP depend on $\Theta^{(h)}_{1,t+h}$.

Two estimation procedures are considered to estimate $\Theta^{(h)}_{j,t}, j = 1, 2, h = 1, 2, \ldots$.

- Procedure 1: Apply Muller-Petalas estimation pocedures on eq (3.7) and generate path estimators of parameters $A_{j,t}, j = 1, 2$, denoted as $\hat{A}^{MP-VAR}_{j,t}, j = 1, 2$. Then the path estimators of $\Theta^{(h)}_{j,t}, j = 1, 2$,

denoted as $\hat{\Theta}_{j,t}^{(h),MP-VAR}, j = 1, 2$, are computed accordingly as

$$\begin{aligned}
\hat{\Theta}_{1,t+2}^{(2),MP-VAR} &= \hat{A}_{1,t+2}^{MP-VAR}\hat{A}_{1,t+1}^{MP-VAR} + \hat{A}_{2,t+2}^{MP-VAR} \\
\hat{\Theta}_{2,t+2}^{(2),MP-VAR} &= \hat{A}_{1,t+2}^{MP-VAR}\hat{A}_{2,t+1}^{MP-VAR} \\
\hat{\Theta}_{1,t+3}^{(3),MP-VAR} &= \hat{A}_{1,t+3}^{MP-VAR}\hat{A}_{1,t+2}^{MP-VAR}\hat{A}_{1,t+1}^{MP-VAR} \\
&\quad + \hat{A}_{2,t+3}^{MP-VAR}\hat{A}_{1,t+1}^{MP-VAR} + \hat{A}_{1,t+3}^{MP-VAR}\hat{A}_{2,t+2}^{MP-VAR} \\
\hat{\Theta}_{2,t+3}^{(3),MP-VAR} &= \hat{A}_{1,t+3}^{MP-VAR}\hat{A}_{1,t+2}^{MP-VAR}\hat{A}_{2,t+1}^{MP-VAR} + \hat{A}_{2,t+3}^{MP-VAR}\hat{A}_{2,t+1}^{MP-VAR} \\
&\vdots
\end{aligned} \tag{3.12}$$

- Procedure 2: Apply the Muller-Petalas estimation procedures on eq (3.9) and generate path estimators of parameters $\Theta_{j,t}^{(h)}, j = 1, 2$, denoted as $\hat{\Theta}_{j,t}^{(h),MP-LP}, j = 1, 2$.

We consider the multivariate scenario VAR(2), i.e., $K = 2$, $p = 2$, and set $q = 8$, $T = 4000$. 5000 Monte Carlo replications are conducted. In each replication i, we obtain $\{\hat{A}_{j,t}^{MP-VAR,(i)}, j = 1, 2\}_{t=1}^{T}$, $\{\hat{\Theta}_{j,t}^{(h),MP-VAR,(i)}, , j = 1, 2\}_{t=1}^{T}$, and $\{\hat{\Theta}_{j,t}^{(h),MP-LP,(i)}, , j = 1, 2\}_{t=1}^{T}$. For each period t, we compute the average values of the path estimators across 5000 replications, that is, $\bar{\hat{A}}_{j,t}^{MP-VAR} = \frac{1}{5000}\sum_i \hat{A}_{j,t}^{MP-VAR,(i)}, j = 1, 2$, $\bar{\hat{\Theta}}_{j,t}^{(h),MP-VAR} = \frac{1}{5000}\sum_i \hat{\Theta}_{j,t}^{(h),MP-VAR,(i)}, j = 1, 2, h = 2, 3$, and $\bar{\hat{\Theta}}_{j,t}^{(h),MP-LP} = \frac{1}{5000}\sum_i \hat{\Theta}_{j,t}^{(h),MP-LP,(i)}, j = 1, 2, h = 2, 3$.

Figure 3.1 plot the average values of the path estimators $\{\bar{\hat{A}}_{j,t}^{MP-VAR}\}_{t=1}^{T}$ for $j = 1, 2$ in unstable VARs with different choices of $c = 4, 8, 12$. It shows that the average Muller-Petalas path estimators are close to the true time-varying parameters in unstable VARs.

Figure 3.3-3.5 plot the average values of the path estimators $\{\bar{\hat{\Theta}}_{j,t}^{(h),MP-VAR}\}_{t=1}^{T}$, $\{\bar{\hat{\Theta}}_{j,t}^{(h),MP-LP}\}_{t=1}^{T}$ together with the true $\{\Theta_{j,t}^{(h)}\}_{t=1}^{T}$ computed using true $\{A_{j,t}\}_{t=1}^{T}$, $j = 1, 2$ $h = 2, 3$, for model with time-varying process in eq (3.8), with different choices of c. Figure 3.3-3.5 imply that Muller-Petalas path estimators by LPs perform well with proper choices of $c = 4, 8, 12$,

that is, moderate variability gives good results. But if c is too small, say $c = 1$, Muller-Petalas path estimators by LPs perform badly for $h = 3$, see Figure 3.2. In the meanwhile, too large c causes nonstationarity.

3.3.4 Bayesian credible intervals of Muller-Petalas path estimators

This subsection further analyzes the accuracy of Muller-Petalas path estimators in unstable VARs, obtained in the context of eq (3.7), as well as Muller-Petalas path estimators by LPs for $h = 2, 3$, obtained in the context of eq (3.11), by checking the Bayesian credible intervals.

Let θ_t denotes the vector of the targeted parameters (the first q parameters) at period $t, t = 1, \ldots, T$. We estimate θ_t by applying the Muller-Petalas path estimator on eq (3.7) under VAR setting, and eq (3.11) based on LPs for $h = 2, 3$. Thus, under VAR setting, $\hat{\theta}_t = vec\left(\hat{A}_{j,t}^{MP-VAR}, j = 1, 2\right)$, and for $h = \{2, 3\}$, $\hat{\theta}_t = vec\left(\hat{\Theta}_{j,t}^{(h),MP-LP}, j = 1, 2\right)$ by LPs. Along with the Muller-Petalas path estimators $\hat{\theta}_t$, a $(q \times q)$ covariance matrix of the approximate posterior for θ_t is obtained, denoted as Ω_t. The interval $\left[\hat{\theta}_{t,j} - 1.96\sqrt{\Omega_{t,jj}}, \hat{\theta}_{t,j} + 1.96\sqrt{\Omega_{t,jj}}\right]$ with $\hat{\theta}_{t,j}$ the j-th element of $\hat{\theta}_t$ and $\Omega_{t,jj}$ the (j, j) element of Ω_t is thus approximately the 95% equal-tailed posterior probability interval for $\theta_{t,j}$, the j-th element of θ at time t.

With the group of parameters $(K = 2, p = 2, T = 4000)$, $n_D = 100$ datasets are generated from eq (3.7) with time-varying parameters following eq (3.8), each dataset denoted by $\mathcal{D}^d, d = 1, \ldots, n_D$. The path estimators and their Bayesian credible intervals obtained from dataset $\mathcal{D}^d$ are denoted by $\hat{\theta}_t^{(d)}$ and $BCI_{t,j}^{(d)} = \left[\hat{\theta}_{t,j}^{(d)} - 1.96\sqrt{\Omega_{t,jj}^{(d)}}, \hat{\theta}_{t,j}^{(d)} + 1.96\sqrt{\Omega_{t,jj}^{(d)}}\right]$ for $t = 1, \ldots, T, j = 1, \ldots, q, d = 1, \ldots, n_D$. For each dataset $\mathcal{D}^d, d = 1, \ldots, n_D$, we do the following:

- Step 1: For each scenario, either eq (3.7) or (3.11), we estimate them as a time-varying parameter Bayesian VAR (TVP-BVAR) and generate n_{sim} draws from the posterior distribution of θ_t, denoted

as $\{\theta_t^{(d,i)}\}_{i=1}^{n_{sim}}$, using Gibbs sampling (see [Carter and Kohn, 1994] and [Koop and Korobilis, 2010]). As for the priors, all parameters are initialized at their estimates at period 1, i.e., $\hat{\theta}_1$, with the natural conjugate priors, that is, we assume Normal distribution for coefficient parameter vectors and inverted Wishart distribution for variance parameter matrices.

- Step 2: We calculate the rejection rate sequence across these n_{sim} draws, denoted by $\{r_t^d\}_{t=1}^T$, with the j-th element of r^d at time t computed as $r_{t,j}^d = Pr(\theta_{t,j}^{(d)} \notin BCI_{t,j}^{(d)})$. That is, for each $i = 1, \ldots, n_{sim}$, we reject if $\theta_{t,j}^{(d,i)}$ falls out of the interval $BCI_{t,j}^{(d)} = \left[\hat{\theta}_{t,j}^{(d)} - 1.96\sqrt{\Omega_{t,jj}^{(d)}}, \hat{\theta}_{t,j}^{(d)} + 1.96\sqrt{\Omega_{t,jj}^{(d)}}\right]$, for $j = 1, \ldots, q$.

From step 1 and 2 above, we get n_D rejection rate sequences, $\{r_t^{(d)}\}_{t=1}^T, d = 1, \ldots, n_D$. We calculate the average of the rejection rate sequences across n_D datasets, denoted as $\bar{r}_t = \frac{1}{n_D} r_t^{(d)}, t = 1, \ldots, T$. Figure 3.6 - 3.11 display $\{\bar{r}_t\}_{t=1}^T$ of all the parameters for the scenario VAR ($h = 1$) and LP ($h = 2, 3$) with different choices of $c = 4, 8, 12$, respectively. The results show that $\{\bar{r}_t\}_{t=1}^T$ deviate a little bit from 5% at each period $t = 1, \ldots, T$.

3.4 Empirical Study

This section applies the Muller-Petalas path estimators for the estimation of the small quarterly time-varying SVAR model of the U.S. economy studied in [Primiceri, 2005].

Following [Primiceri, 2005], we consider a small VAR with three variables: inflation rate (the annual percentage change in a chain-weighted GDP price index), unemployment rate (seasonally adjusted civilian unemployment rate, all workers over age 16), and a short-term nominal interest rate (yield on the three month Treasury bill rate).[10] These three

[10]The data are obtained from the Federal Reserve Bank of St. Louis website.

variables are typically used in New Keynesian VARs. The same, or similar, dataset is studied in [Stock and Watson, 2001], [Cogley and Sargent, 2005], [Primiceri, 2005], and [Koop and Korobilis, 2010]. The alternative would be larger sets of variables, see for example [Bernanke and Mihov, 1998] and [Sims and Zha, 1998]. The sample ranges from 1953:Q1 to 2019:Q4. Two lags are used for the estimation. The same identification scheme as in [Primiceri, 2005] is adopted.

Figure 3.12 presents a plot of the Muller-Petalas path estimators in a VAR setting and the [Primiceri, 2005] posterior mean of the standard deviations of each equation. The time-varying standard deviation of the identified monetary policy shocks, i.e., the interest rate equation, measures the relative importance and changes of the monetary policy. The two measurements are very close. Similar to [Primiceri, 2005] time-varying measurement, the Muller-Petalas measurement of the monetary policy shocks also (i) exhibits a substantially higher variance during period 1979-1983, i.e., the Volcker period; and (ii) is less volatile and remains very low and substantially constant in the post-Volcker period compared with the pre-Volcker period, implying that Taylor rules have been good approximations of the U.S. monetary policy after the Volcker period. Compared with the [Primiceri, 2005] time-varying measurement, the Muller-Petalas measurement of the monetary policy shocks is lower around 1981 and slightly higher after the 2008 financial crisis.

Figure 3.13 presents a plot of the Muller-Petalas path estimators in a VAR setting and the [Primiceri, 2005] posterior mean of the coefficient matrix A_1 for the first lag. It shows that the Muller-Petalas path estimators are much more volatile compared with the [Primiceri, 2005] time-varying posterior mean estimators, especially around 1975:Q1, 1981:Q3, and 2008:Q3. These three dates correspond to the NBER business cycle trough date, the NBER business cycle peak date, and the financial crisis respectively, which are meant to capture very different economic conditions. The Muller-Petalas path estimators indicate that the estimated coefficients experience much time variation under these different economic conditions, while [Primiceri, 2005] time-varying SVAR indicates little time variation in the estimated coefficients.

Figure 3.14 presents the effects of the monetary policy shocks, that is, the impulse responses of inflation, unemployment rate, and interest rate, computed based on the Muller-Petalas path estimators under local projections. We explore the effects at various dates of the sample: 1975:Q1, 1981:Q3, 1996:Q1, 2002:Q1, 2008:Q3, and 2016:Q1, which are either representative of the typical economic conditions or arbitrarily chosen.[11] The plot implies that the effects of the monetary policy shocks differ across time, especially at period 1981:Q3. At period 1981:Q3, the effects of the monetary policy shocks are opposite to the other periods, which is not surprising due to the Volcker period. Among the other dates of the sample, the impulse responses of inflation and unemployment rate at 1975:Q1 and 2008:Q3 are slightly more volatile.

Figures 3.15-3.20 further compare the effects of the monetary policy shocks computed based on the Muller-Petalas path estimators under local projections with the effects computed based on [Primiceri, 2005] time-varying SVAR at each period. These graphs present several interesting features. Compared with [Primiceri, 2005] time-varying SVAR, the impulse responses based on Muller-Petalas path estimators under local projections (i) share similar trends most of the time, (ii) are generally less smooth and imply that the monetary policy shocks have effects of larger magnitude and that last longer. Compared with impulse responses computed based on the time-varying SVAR model in [Primiceri, 2005], which replies on linearity, the impulse responses based on Muller-Petalas path estimators under local projections could better uncover the true impulse response as they are robust to misspecification of the data generating process and they easily accommodate the situation when GDP follows a non-linear model which are often impractical or infeasible in a multivariate context. Besides, this new procedure is easy to implement as doesn't require the complicated MCMC algorithm.

[11] As mentioned in [Primiceri, 2005], 1975:Q1, 1981:Q3, and 1996:Q1 are representative of the typical economic conditions of the chairmanships of Burns, Volcker and Greenspan.

3.5 Conclusion

This paper introduces local time variation into the *local projections* framework, considering a smooth stochastic evolution of both the coefficients and the variances, and proposes the impulse responses estimation methodology under unstable *local projections* built on [Muller and Petalas, 2010] weighted average risk (WAR) minimizing path estimators in a multivariate system. Proper elements in the path estimators coincide with the impulse responses of the corresponding structural shocks, and proper elements in the covariance matrix of the approximate posterior of the path estimators help gain some sense of the accuracy of the impulse responses.

The Monte Carlo studies show that [Muller and Petalas, 2010] asymptotically WAR minimizing path estimators and WAP maximizing parameter stability test statistics perform well in the unstable *local projections* framework with flexible specifications.

In the empirical study, we revisit the small quarterly time-varying SVAR model of the U.S. economy studied in [Primiceri, 2005], and find that (i) the time-varying standard deviation of the identified monetary policy shocks has a similar pattern with that in [Primiceri, 2005], (ii) the path estimators experience much time variation under these different economic conditions, which is opposite to [Primiceri, 2005], and (iii) impulse responses under unstable *local projections* share similar trends most of the time, but are generally less smooth and imply that the monetary policy shocks have effects of larger magnitude and that last longer.

Table 3.1: Rejection Rates of $qLL(10)$, Normal Size 0.05, Univariate

AR(p): $K = 1, \Sigma_u = 1$		
		$p = 2$
q	T	Rejection Rate (%)
2	2000	5.55
2	4000	4.40
2	10000	4.65
		$p = 3$
q	T	Rejection Rate (%)
2	2000	5.50
2	4000	5.35
2	10000	5.10
3	2000	4.90
3	4000	5.40
3	10000	4.35
		$p = 4$
q	T	Rejection Rate (%)
2	2000	4.90
2	4000	5.75
2	10000	5.80
3	2000	5.60
3	4000	6.35
3	10000	6.05
4	2000	5.70
4	4000	5.25
4	10000	5.90

Note: This table reports rejection frequencies over 2000 Monte Carlo replications for the univariate scenario AR(p), i.e., $K = 1$, using the qLL(10) statistic in [Muller and Petalas, 2010] and critical values in Table 1 in [Elliott and Muller, 2006]. T and q denote the number of observations and the number of targeted parameters, respectively. The parameter matrices are set to be $A_j = 0.2, j = 1, \ldots, p$ and and $\Sigma_u = 1$.

Table 3.2: Rejection Rates of $qLL(10)$, Normal Size 0.05, Multivariate

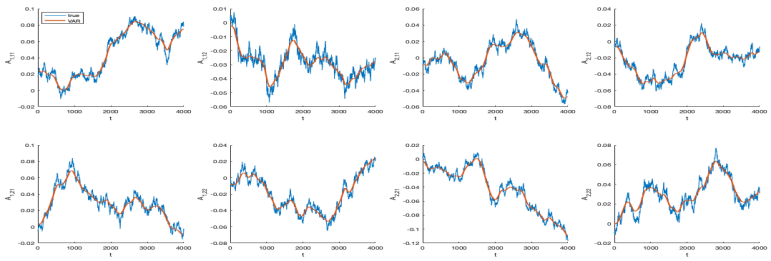
VAR(1): $K = 2, p = 1, \Sigma_u = I_K$		
q	T	Rejection Rate (%)
2	2000	4.40
2	4000	3.90
2	10000	4.60
4	2000	5.80
4	4000	5.00
4	10000	4.60

Note: This table reports rejection frequencies over 1000 Monte Carlo replications for the multivariate scenario VAR(1) with 2 variables, i.e., $K = 2, p = 1$, using the qLL(10) statistic in [Muller and Petalas, 2010] and critical values in Table 1 in [Elliott and Muller, 2006]. T and q denote the number of observations and the number of targeted parameters, respectively. The parameter matrices are set to be $A_1 = \begin{bmatrix} 0.2 & 0 \\ 0.1 & 0.3 \end{bmatrix}$ and $\Sigma_u = I_K$.

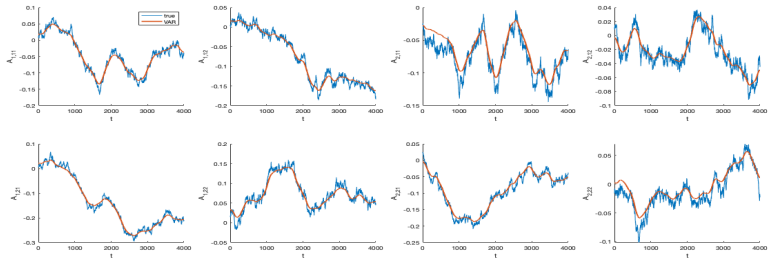
Table 3.3: Rejection Rates of $qLL(10)$, Normal Size 0.05, Multivariate

VAR(1): $K = 2, p = 1, \Sigma_u \neq I_K$		
q	T	Rejection Rate (%)
2	2000	5.00
2	4000	4.80
2	10000	4.80
4	2000	4.70
4	4000	5.10
4	10000	4.90

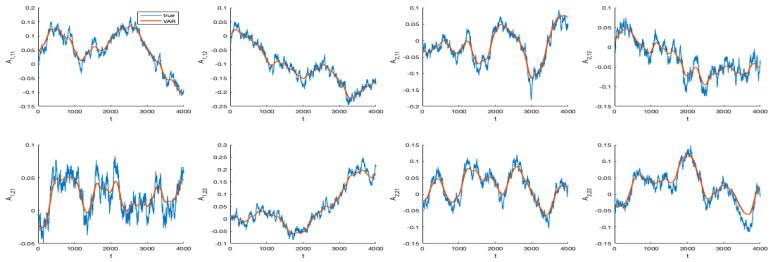
Note: This table reports rejection frequencies over 1000 Monte Carlo replications for the multivariate scenario VAR(1) with 2 variables, i.e., $K = 2, p = 1$, using the qLL(10) statistic in [Muller and Petalas, 2010] and critical values in Table 1 in [Elliott and Muller, 2006]. T and q denote the number of observations and the number of targeted parameters, respectively. The parameter matrices are set to be $A_1 = \begin{bmatrix} 0.2 & 0 \\ 0.1 & 0.3 \end{bmatrix}$ and $\Sigma_u = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}$.



(a) $c = 4$



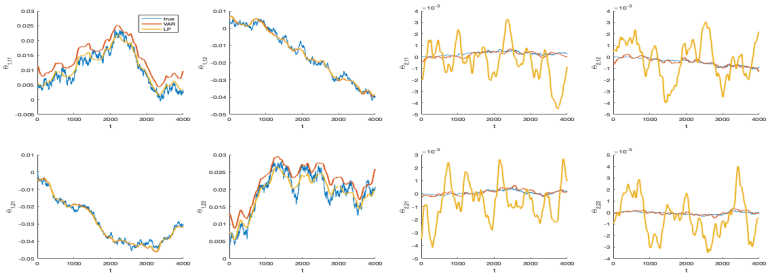
(b) $c = 8$



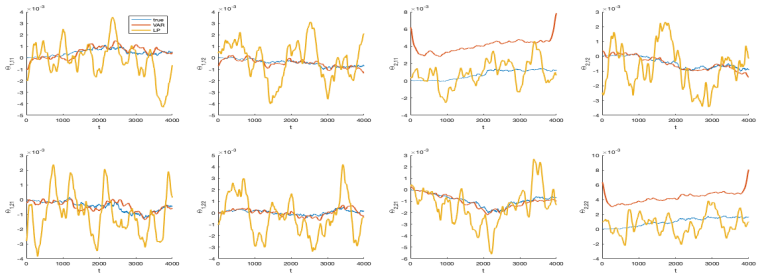
(c) $c = 12$

Figure 3.1: Path estimators of $\{A_{j,t}\}_{t=1}^{T}$, VAR

Note: This figure plots the average values of the parameter path estimators for $A_{j,t}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 8$, $T = 4000$, and different choices of $c = 4, 8, 12$ under VAR setting. The blue lines are the true parameter values of $\{A_{j,t}\}_{t=1}^{T}, j = 1, 2$, and the red lines are the average path estimators $\{\bar{\tilde{A}}_{j,t}^{MP-VAR}\}_{t=1}^{T}, j = 1, 2$. The x-axis refers to period, and y-axis refers to the parameters. For example, $A_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $A_{j,t}, j = 1, 2$.



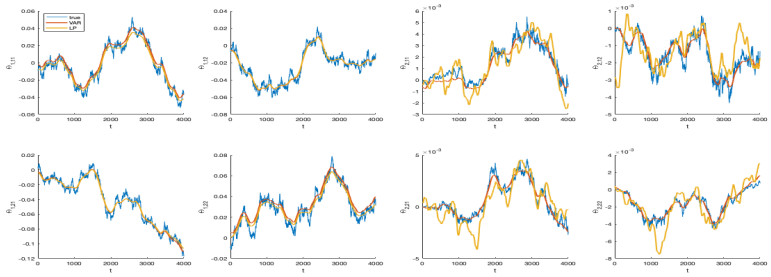
(a) $h = 2$



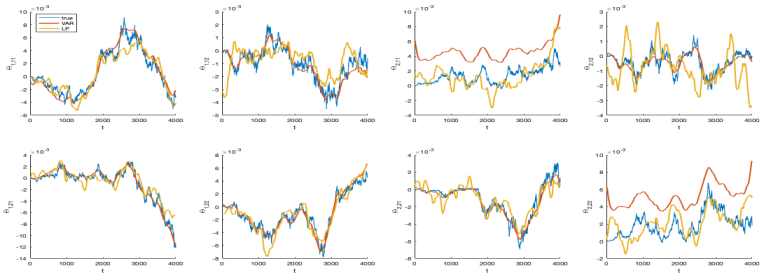
(b) $h = 3$

Figure 3.2: Path estimators of $\{\Theta_{j,t}^{(h)}\}_{t=1}^{T}$, $c = 1$

Note: This figure plots the average values of the parameter path estimators for $\Theta_{j,t}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 8$, $T = 4000$, and $c = 1$, using different estimation procedures. The blue lines are the true parameter values of $\{\Theta_{j,t}\}_{t=1}^{T}, j = 1, 2$, the red lines are the average path estimators $\{\bar{\Theta}_{j,t}^{MP-VAR}\}_{t=1}^{T}, j = 1, 2$, and the orange lines are the average path estimators $\{\bar{\Theta}_{j,t}^{MP-LP}\}_{t=1}^{T}, j = 1, 2$. The x-axis refers to period, and y-axis refers to the parameters. For example, $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.



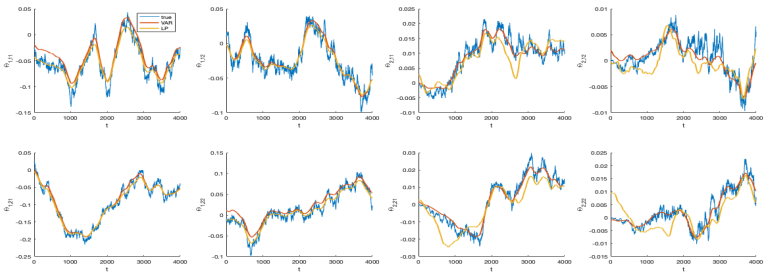
(a) $h = 2$



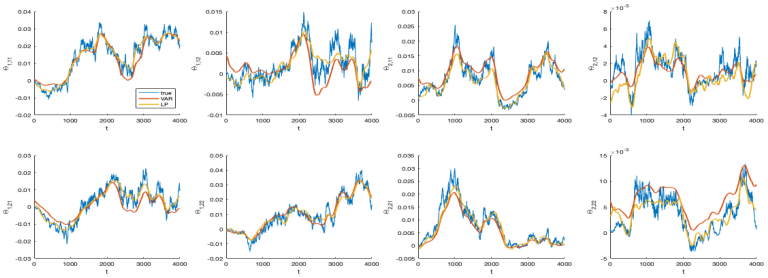
(b) $h = 3$

Figure 3.3: Path estimators of $\{\Theta_{j,t}^{(h)}\}_{t=1}^{T}$, $c = 4$

Note: This figure plots the average values of the parameter path estimators for $\Theta_{j,t}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 8$, $T = 4000$, and $c = 4$, using different estimation procedures. The blue lines are the true parameter values of $\{\Theta_{j,t}\}_{t=1}^{T}, j = 1, 2$, the red lines are the average path estimators $\{\bar{\tilde{\Theta}}_{j,t}^{MP-VAR}\}_{t=1}^{T}, j = 1, 2$, and the orange lines are the average path estimators $\{\bar{\tilde{\Theta}}_{j,t}^{MP-LP}\}_{t=1}^{T}, j = 1, 2$. The x-axis refers to period, and y-axis refers to the parameters. For example, $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.



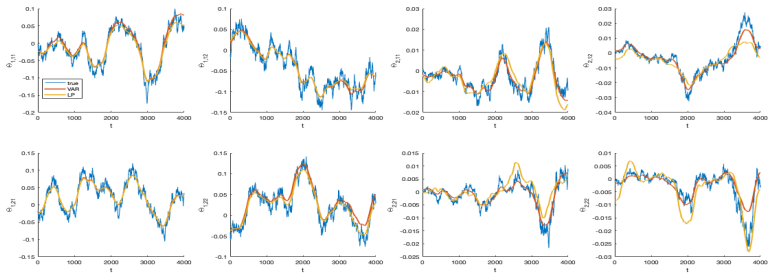
(a) $h = 2$



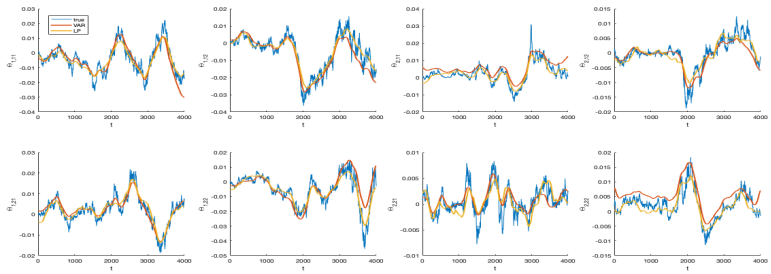
(b) $h = 3$

Figure 3.4: Path estimators of $\{\Theta_{j,t}^{(h)}\}_{t=1}^{T}$, $c = 8$

Note: This figure plots the average values of the parameter path estimators for $\Theta_{j,t}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 8$, $T = 4000$, and $c = 8$, using different estimation procedures. The blue lines are the true parameter values of $\{\Theta_{j,t}\}_{t=1}^{T}, j = 1, 2$, the red lines are the average path estimators $\{\bar{\tilde{\Theta}}_{j,t}^{MP-VAR}\}_{t=1}^{T}, j = 1, 2$, and the orange lines are the average path estimators $\{\bar{\tilde{\Theta}}_{j,t}^{MP-LP}\}_{t=1}^{T}, j = 1, 2$. The x-axis refers to period, and y-axis refers to the parameters. For example, $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.



(a) $h = 2$



(b) $h = 3$

Figure 3.5: Path estimators of $\{\Theta_{j,t}^{(h)}\}_{t=1}^{T}$, $c = 12$

Note: This figure plots the average values of the parameter path estimators for $\Theta_{j,t}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 8$, $T = 4000$, and $c = 12$, using different estimation procedures. The blue lines are the true parameter values of $\{\Theta_{j,t}\}_{t=1}^{T}, j = 1, 2$, the red lines are the average path estimators $\{\bar{\tilde{\Theta}}_{j,t}^{MP-VAR}\}_{t=1}^{T}, j = 1, 2$, and the orange lines are the average path estimators $\{\bar{\tilde{\Theta}}_{j,t}^{MP-LP}\}_{t=1}^{T}, j = 1, 2$. The x-axis refers to period, and y-axis refers to the parameters. For example, $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.

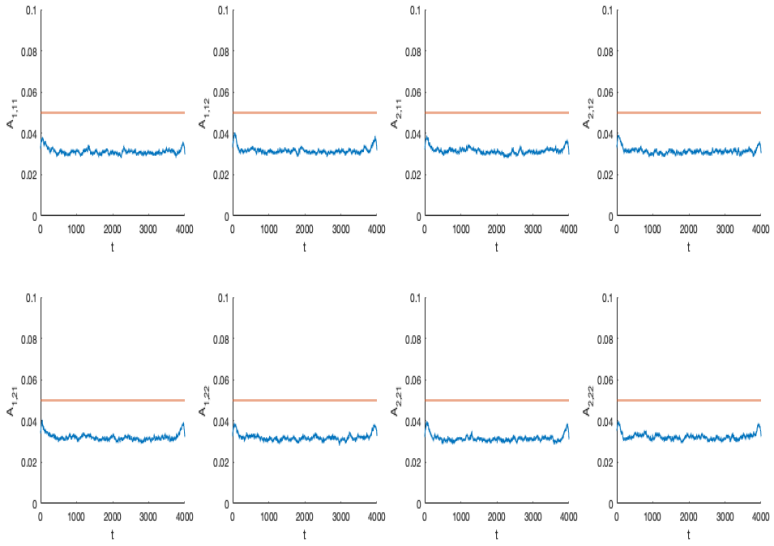
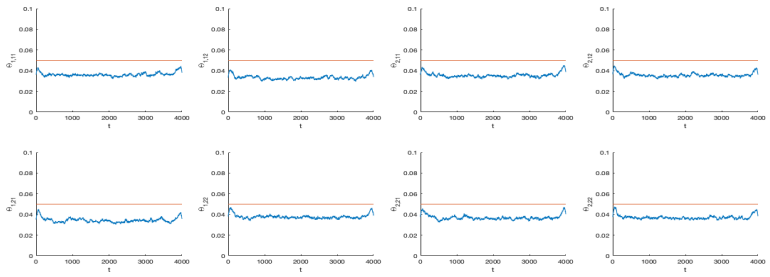
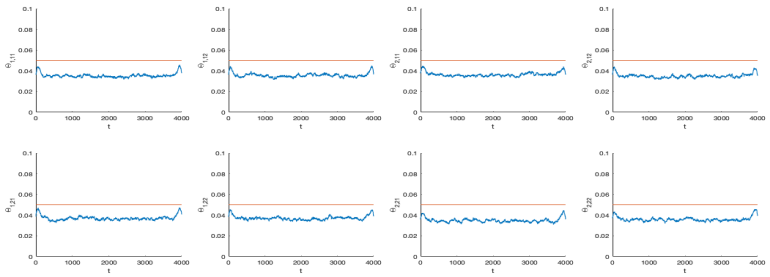


Figure 3.6: Bayesian credible interval check of $\{A_{j,t}\}_{t=1}^{T}$, $c = 4$, VAR

Note: This figure plots the coverage rates of the parameter path estimators for $A_{j,t}^{MP-VAR}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 10$, $T = 4000$, and $c = 4$, under VAR setting. The blue lines are the rejection rates, the red lines are 5%. The x-axis refers to period, and y-axis refers to the parameters. For example, $A_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $A_{j,t}, j = 1, 2$.



(a) $h = 2$



(b) $h = 3$

Figure 3.7: Bayesian credible interval check of $\{\Theta_{j,t}^{(h),MP-LP}\}_{t=1}^{T}$, $c = 4$, LP

Note: This figure plots the coverage rate of the parameter path estimators $\Theta_{j,t}^{(h),MP-LP}, j = 1, 2, h = 2, 3$ across 5000 replications, i.e., the frequency when $n_{sim} = 500$ draws from the posterior distribution of the parameters are within the 95% Bayesian credible interval across $n_D = 100$ datasets, for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 10$, $T = 4000$, and $c = 4$, by LPs. The Bayesian credible interval is constructed based on the covariance matrix of the approximate posterior for the path estimators discussed in [Muller and Petalas, 2010]. The blue lines are the rejection rates, the red lines are 5%. The x-axis refers to period, and y-axis refers to the coverage rate. And $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.

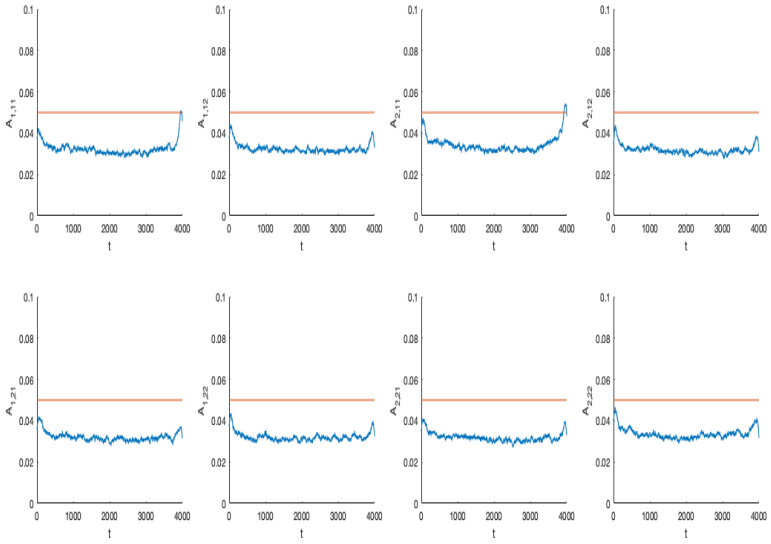
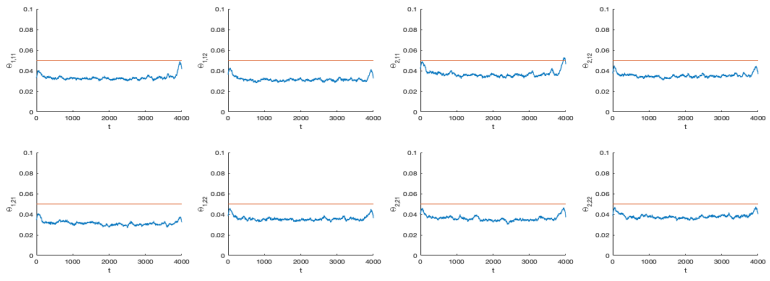
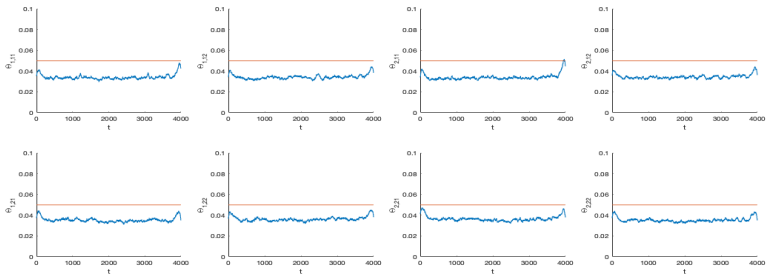


Figure 3.8: Bayesian credible interval check of $\{A_{j,t}\}_{t=1}^{T}$, $c = 8$, VAR

Note: This figure plots the coverage rates of the parameter path estimators for $A_{j,t}^{MP-VAR}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 10$, $T = 4000$, and $c = 8$, under VAR setting. The blue lines are the rejection rates, the red lines are 5%. The x-axis refers to period, and y-axis refers to the parameters. For example, $A_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $A_{j,t}, j = 1, 2$.



(a) $h = 2$



(b) $h = 3$

Figure 3.9: Bayesian credible interval check of $\{\Theta_{j,t}^{(h),MP-LP}\}_{t=1}^{T}$, $c = 8$, LP

Note: This figure plots the coverage rate of the parameter path estimators $\Theta_{j,t}^{(h),MP-LP}, j = 1, 2, h = 2, 3$ across 5000 replications, i.e., the frequency when $n_{sim} = 500$ draws from the posterior distribution of the parameters are within the 95% Bayesian credible interval across $n_D = 100$ datasets, for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 10$, $T = 4000$, and $c = 8$, by LPs. The Bayesian credible interval is constructed based on the covariance matrix of the approximate posterior for the path estimators discussed in [Muller and Petalas, 2010]. The blue lines are the rejection rates, the red lines are 5%. The x-axis refers to period, and y-axis refers to the coverage rate. And $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.

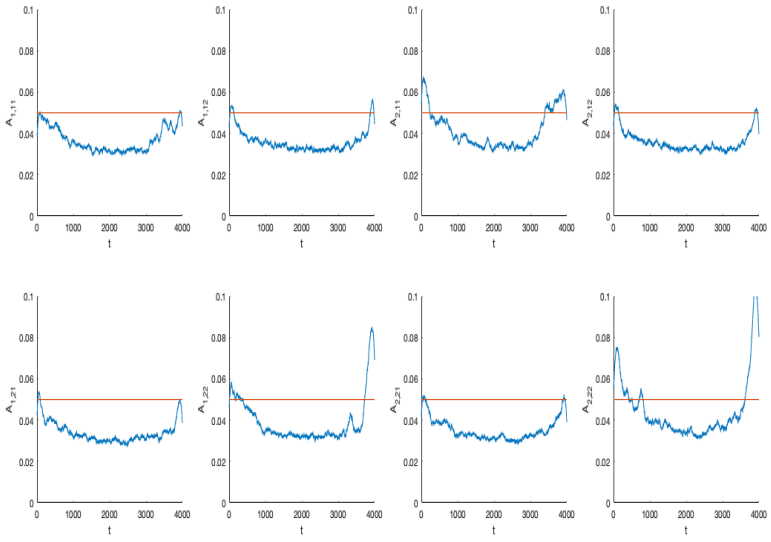
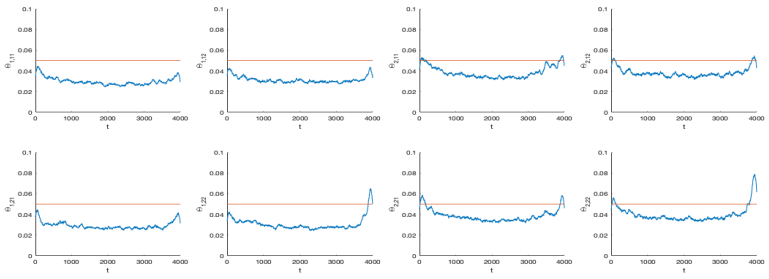
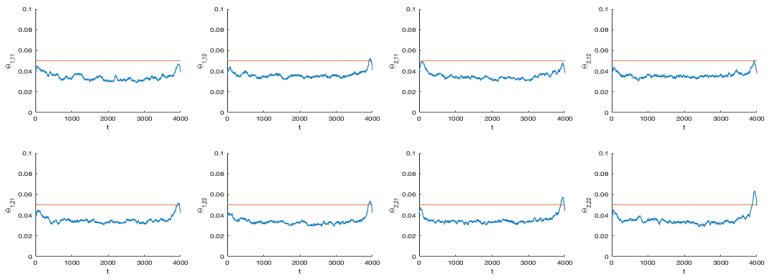


Figure 3.10: Bayesian credible interval check of $\{A_{j,t}\}_{t=1}^{T}$, $c = 12$, VAR

Note: This figure plots the coverage rates of the parameter path estimators for $A_{j,t}^{MP-VAR}, j = 1, 2$ across 5000 replications for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 10$, $T = 4000$, and $c = 12$, under VAR setting. The blue lines are the rejection rates, the red lines are 5%. The x-axis refers to period, and y-axis refers to the parameters. For example, $A_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $A_{j,t}, j = 1, 2$.



(a) $h = 2$



(b) $h = 3$

Figure 3.11: Bayesian credible interval check of $\{\Theta_{j,t}^{(h),MP-LP}\}_{t=1}^{T}$, $c = 12$, LP

Note: This figure plots the coverage rate of the parameter path estimators $\Theta_{j,t}^{(h),MP-LP}, j = 1, 2, h = 2, 3$ across 5000 replications, i.e., the frequency when $n_{sim} = 500$ draws from the posterior distribution of the parameters are within the 95% Bayesian credible interval across $n_D = 100$ datasets, for DGP TVP-VAR(2) with $K = 2$, $p = 2$, $q = 10$, $T = 4000$, and $c = 12$, by LPs. The Bayesian credible interval is constructed based on the covariance matrix of the approximate posterior for the path estimators discussed in [Muller and Petalas, 2010]. The blue lines are the rejection rates, the red lines are 5%. The x-axis refers to period, and y-axis refers to the coverage rate. And $\Theta_{j,rs}, j = 1, 2$ refers to the r^{th} row and s^{th} column element in the parameter matrix $\Theta_{j,t}, j = 1, 2$.

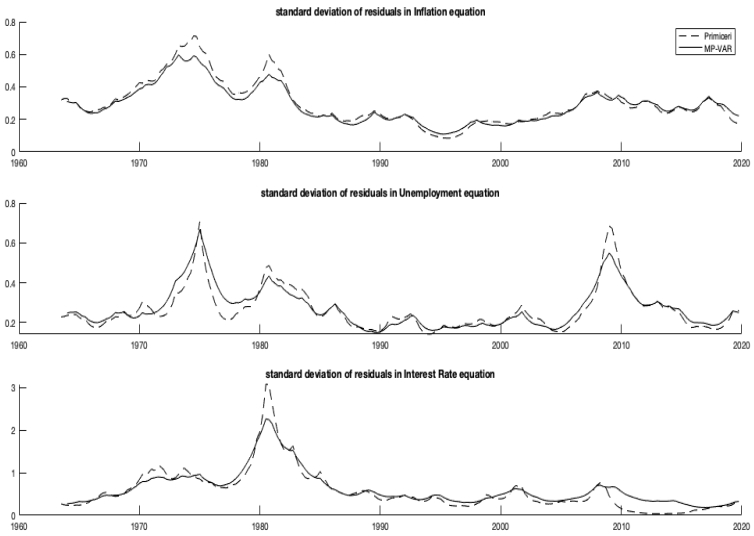


Figure 3.12: Standard deviation of the residuals

Note: This figure plots the standard deviation of (a) residuals of the inflation equation, (b) residuals of the unemployment equation and (c) residuals of the interest rate equation or monetary policy shocks.

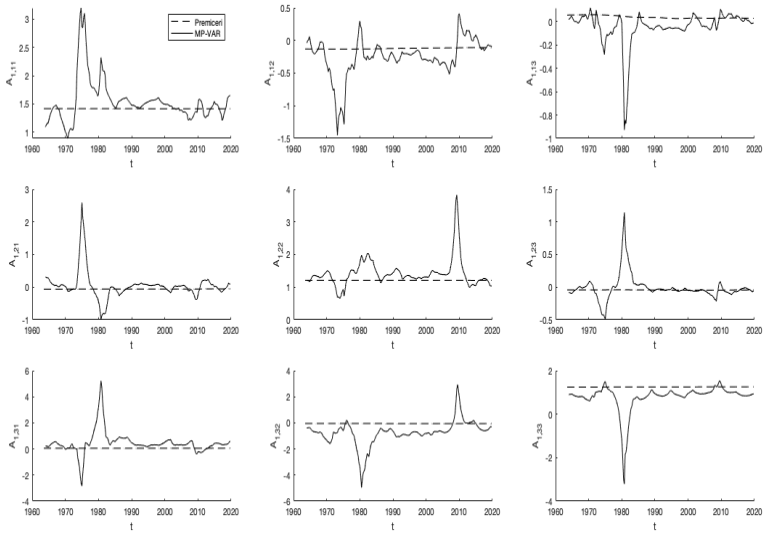


Figure 3.13: Path estimators

Note: This figure plots the path estimators of all the elements in parameter matrix A_1. $A_{1,rs}$ refers to the r^{th} row and s^{th} column element in the parameter matrix A_1.

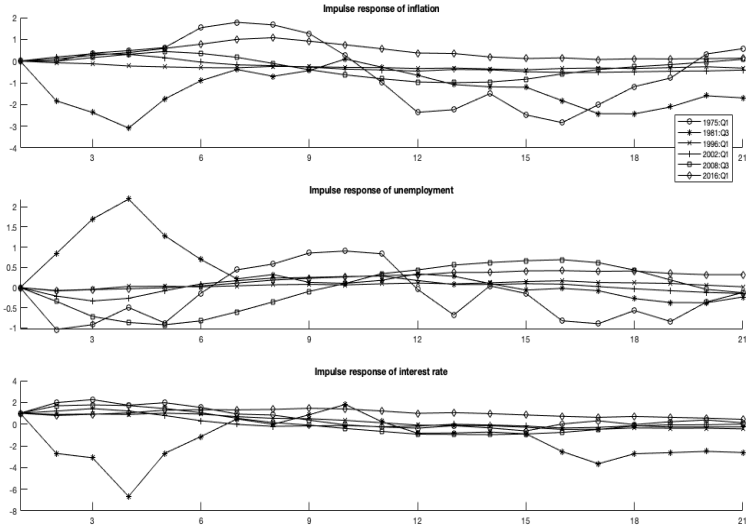


Figure 3.14: Impulse responses at different periods

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections, at various periods: 1975:Q1, 1981:Q3, 1996:Q1, 2002:Q1, 2008:Q3, and 2016:Q1. The x-axis refers to horizons.

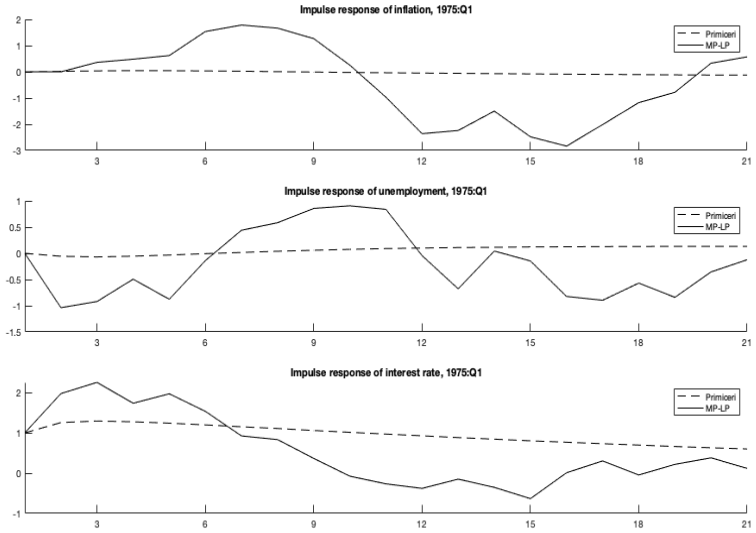


Figure 3.15: Impulse responses, 1975:Q1

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections and time-varying SVAR in [Primiceri, 2005] respectively. The x-axis refers to horizons.

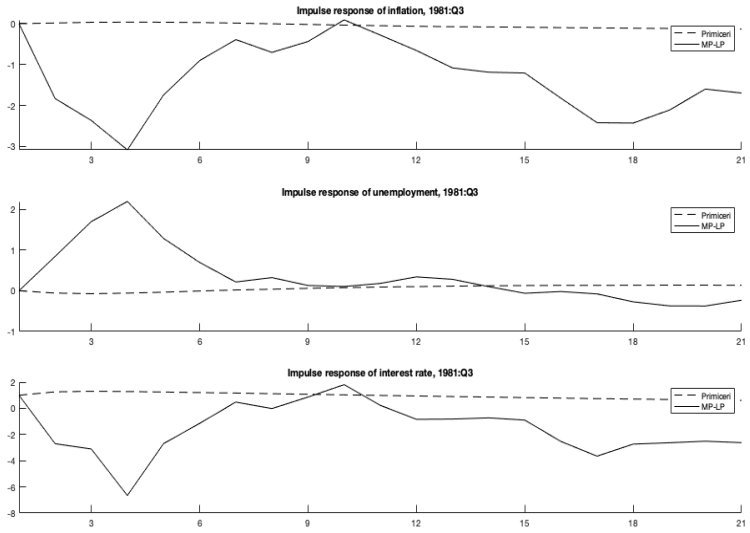


Figure 3.16: Impulse responses, 1981:Q3

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections and time-varying SVAR in [Primiceri, 2005] respectively. The x-axis refers to horizons.

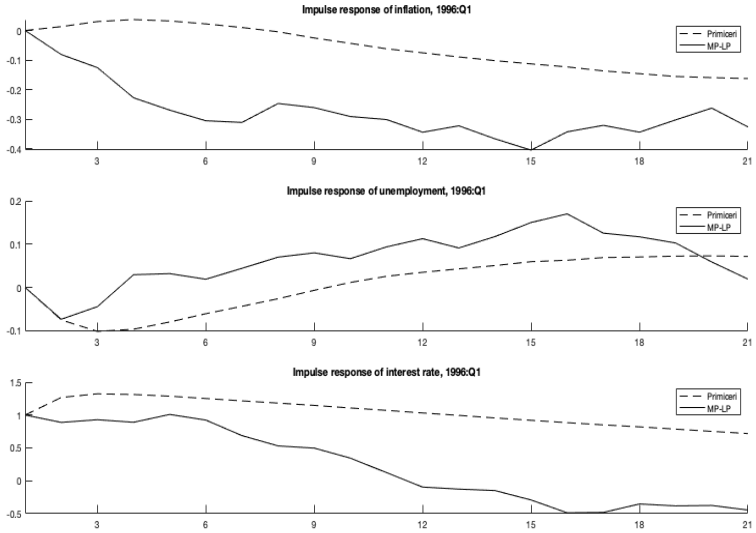


Figure 3.17: Impulse responses, 1996:Q1

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections and time-varying SVAR in [Primiceri, 2005] respectively. The x-axis refers to horizons.

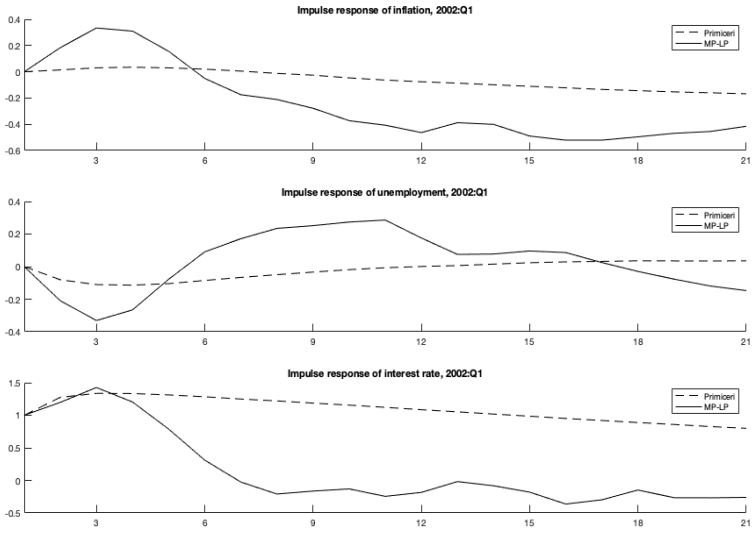


Figure 3.18: Impulse responses, 2002:Q1

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections and time-varying SVAR in [Primiceri, 2005] respectively. The x-axis refers to horizons.

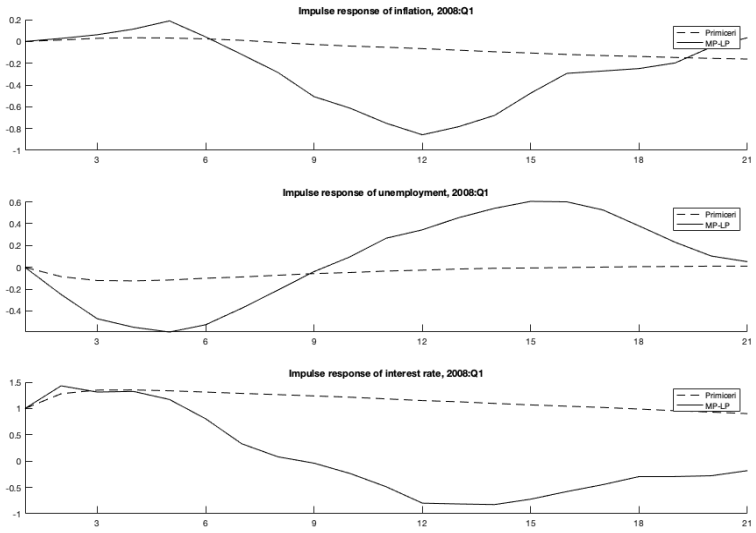


Figure 3.19: Impulse responses, 2008:Q3

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections and time-varying SVAR in [Primiceri, 2005] respectively. The x-axis refers to horizons.

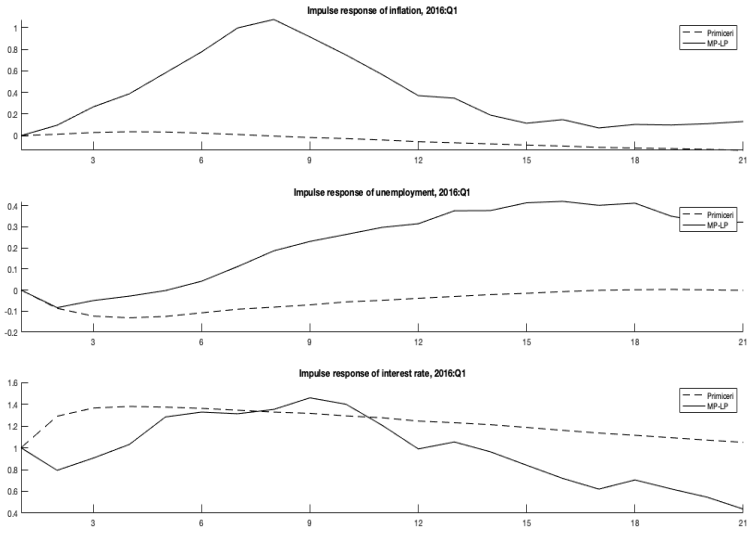


Figure 3.20: Impulse responses, 2016:Q1

Note: This figure plots the impulse responses of each variables to a monatary policy shock, computed based on Muller-Petalas path estimators under local projections and time-varying SVAR in [Primiceri, 2005] respectively. The x-axis refers to horizons.

Bibliography

[Adrian et al., 2019] Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable Growth. *American Economic Review*, 109.

[Ahn and Horenstein, 2013] Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, 80:1203–1227.

[Amisano and Giacomini, 2007] Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, 25.

[Andrews, 1993] Andrews, D. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica: Journal of the Econometric Society*.

[Ang and Bekaert, 2006] Ang, A. and Bekaert, G. (2006). Stock Return Predictability: Is It There? *The Review of Financial Studies*, 20.

[Angrist et al., 2018] Angrist, J., Jorda, O., and Kuersteiner, G. (2018). Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited. *Journal of Business & Economic Statistics*, 36.

[Auerbach and Gorodnichenko, 2012] Auerbach, A. and Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4.

[Azzalini and Capitanio, 2003] Azzalini, A. and Capitanio, A. (2003). Distributions Generated by Perturbation of Symmetry with Emphasis

on a Multivariate Skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65.

[Bai, 2003a] Bai, J. (2003a). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71:135–171.

[Bai, 2003b] Bai, J. (2003b). Testing Parametric Conditional Distributions of Dynamic Models. *Review of Economics and Statistics*, 85.

[Bai and Ng, 2002] Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70:191–221.

[Bai and Ng, 2006] Bai, J. and Ng, S. (2006). Evaluating Latent and Observed Factors in Macroeconomics and Finance. *Journal of Econometrics*, 131:507–537.

[Bai and Perron, 1998] Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66:47–78.

[Baltagi et al., 2017] Baltagi, B. H., Kao, C., and Wang, F. (2017). Identification and Estimation of a Large Factor Model with Structural Instability. *Journal of Econometrics*, 197:87–100.

[Banerjee et al., 2008] Banerjee, A., Marcellino, M., and Masten (2008). Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. In Rapach, D. E. and Wohar, M. E., editors, *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Group Publishing Limited.

[Barigozzi and Trapani, 2017] Barigozzi, M. and Trapani, L. (2017). Sequential testing for structural stability in approximate factor models. *arXiv preprint arXiv:1708.02786.*

[Bates et al., 2013] Bates, B. J., Plagborg-MÃ¸ller, M., Stock, J. H., and Watson, M. W. (2013). Consistent Factor Estimation in Dynamic Factor Models with Structural Instability. *Journal of Econometrics*, 177:289–304.

[Berkowitz, 2001] Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19.

[Bernanke, 2004] Bernanke, B. (2004). The Great Moderation. *Washington, DC.*

[Bernanke and Mihov, 1998] Bernanke, B. and Mihov, I. (1998). Measuring Monetary Policy. *The Quarterly Journal of Economics*, 113.

[Bernanke et al., 2005] Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the Effects of Monetary Policy: a Factor-augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, 120:387–422.

[Blanchard and Simon, 2001] Blanchard, O. and Simon, J. (2001). The Long and Large Decline in us Output Volatility. *Brookings papers on economic activity*, 2001.

[Boivin, 2005] Boivin, J. (2005). Has us Monetary Policy Changed? Evidence from Drifting Coefficients and Real-time Data. *National Bureau of Economic Research. No. w11314*.

[Boivin and Giannoni, 2006] Boivin, J. and Giannoni, M. (2006). Has Monetary Policy Become More Effective? *The Review of Economics and Statistics*, 88.

[Boivin et al., 2009] Boivin, J., Giannoni, M., and Mihov, I. (2009). Sticky Prices and Monetary Policy: Evidence from Disaggregated us Data. *American Economic Review*, 99:350–84.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31.

[Brave and Butters, 2012] Brave, S. and Butters, R. (2012). Diagnosing the Financial System: Financial Conditions and Financial Stress. *BiblioGov.*

[Breitung and Eickmeier, 2011] Breitung, J. and Eickmeier, S. (2011). Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics*, 163:71–84.

[Canova, 1993] Canova, F. (1993). Modelling and Forecasting Exchange Rates with a Bayesian Time-varying Coefficient Model. *Journal of Economic Dynamics and Control*, 17.

[Carter and Kohn, 1994] Carter, C. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81.

[Chen, 2015] Chen, L. (2015). Estimating the Common Break Date in Large Factor Models. *Economics Letters*, 131:70–74.

[Chen et al., 2014] Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting Big Structural Breaks in Large Factor Models. *Journal of Econometrics*, 180:30–48.

[Cheng and Schorfheide, 2016] Cheng, X., L. Z. and Schorfheide, F. (2016). Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities. *The Review of Economic Studies*, 83:1511–1543.

[Clark, 2006] Clark, T. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics*, 21:563–587.

[Clark and McCracken, 2001] Clark, T. and McCracken, M. (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics*, 105.

[Clements and Hendry, 1998] Clements, M. and Hendry, D. (1998). In *Forecasting Economic Time Series.* Cambridge University Press.

[Cogley and Sargent, 2005] Cogley, T. and Sargent, T. (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post wwii us. *Review of Economic dynamics*, 8.

[Connor and Korajczyk, 1986] Connor, G. and Korajczyk, R. (1986). Performance Measurement with the Arbitrage Pricing Theory: a New Framework for Analysis. *Journal of Financial Economics*, 15:373–394.

[Connor and Korajczyk, 1988] Connor, G. and Korajczyk, R. (1988). Risk and Return in an Equilibrium APT: Application of a New Test Methodology. *Journal of Financial Economics*, 21:255–289.

[Connor and Korajczyk, 1993] Connor, G. and Korajczyk, R. (1993). A Test for the Number of Factors in an Approximate Factor Model. *Journal of Finance*, 48:1263–1291.

[Corradi and Swanson, 2005] Corradi, V. and Swanson, N. (2005). A Test for Comparing Multiple Misspecified Conditional Interval Models. *Econometric Theory*, 21.

[Corradi and Swanson, 2006] Corradi, V. and Swanson, N. (2006). Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification. *Journal of Econometrics*, 133.

[Corradi and Swanson, 2014] Corradi, V. and Swanson, N. (2014). Testing for Structural Stability of Factor Augmented Forecasting Models. *Journal of Econometrics*, 182:100–118.

[Debortoli et al., 2018] Debortoli, D., GalÃ, J., and Gambetti, L. (2018). On the empirical (ir) relevance of the zero lower bound constraint.

[Diebold et al., 1997] Diebold, F., Gunther, T., and Tay, A. (1997). Evaluating Density Forecasts.

[Diebold et al., 1998] Diebold, F., Gunther, T., and Tay, A. (1998). Evaluating Density Forecasts with Applications to Finance and Management. *International Economic Review*, 39:863–883.

[Diebold and Lopez, 1996] Diebold, F. and Lopez, J. (1996). Forecast Evaluation and Combination. In *Handbook of Statistics*, volume 14, pages 241–268.

[Elliott and Muller, 2006] Elliott, G. and Muller, U. (2006). Efficient Tests for General Persistent Time Variation in Regression Coefficients. *The Review of Economic Studies*, 73.

[Engle, 1982] Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*.

[Fernandez-Villaverde and Rubio-Ramirez, 2004] Fernandez-Villaverde, J. and Rubio-Ramirez, J. (2004). Comparing dynamic equilibrium models to data: a bayesian approach. *Journal of Econometrics*, 123.

[Forni and Reichlin, 2000] Forni, M., H. M. L. M. and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *Review of Economics and Statistics*, 82:540–554.

[Forni and Reichlin, 2004] Forni, M., H. M. L. M. and Reichlin, L. (2004). The Generalized Dynamic Factor Model Consistency and Rates. *Journal of Econometrics*, 119:231–255.

[Forni and Zaffaroni, 2015] Forni, M., H. M. L. M. and Zaffaroni, P. (2015). Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations. *Journal of Econometrics*, 185:359–371.

[Forni and Zaffaroni, 2017] Forni, M., H. M. L. M. and Zaffaroni, P. (2017). Dynamic Factor Models with Infinite-Dimensional Factor Spaces: Asymptotic Analysis. *Journal of Econometrics*, 199:74–92.

[Forni and Lippi, 2001] Forni, M. and Lippi, M. (2001). The Generalized Dynamic Factor Model: Representation Theory. *Econometric Theory*, 17:1113–1141.

[Giacomini and Rossi, 2009a] Giacomini, R. and Rossi, B. (2009a). Detecting and Predicting Forecast Breakdowns. *The Review of Economic Studies*, 76.

[Giacomini and Rossi, 2009b] Giacomini, R. and Rossi, B. (2009b). How Stable is the Forecasting Performance of the Yield Curve for Output Growth? *Oxford Bulletin of Economics and Statistics*, 68:783–795.

[Giacomini and Rossi, 2010] Giacomini, R. and Rossi, B. (2010). Forecast Comparisons in Unstable Environments. *Journal of Applied Econometrics*, 25.

[Giannone et al., 2008] Giannone, D., Lenza, M., and Reichlin, L. (2008). Explaining the Great Moderation: It Is Not the Shocks. *Journal of the European Economic Association*, 6.

[Gneiting and Raftery, 2007] Gneiting, T. and Raftery, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102.

[Good, 1952] Good, I. (1952). Rational Decisions. *Journal of the Royal Statistical Society*.

[Hamilton, 1994] Hamilton, J. (1994). In *Time Series Analysis*, volume 2, pages 690–696. New Jersey: Princeton.

[Han and Inoue, 2015] Han, X. and Inoue, A. (2015). Tests for Parameter Instability in Dynamic Factor Models. *Econometric Theory*, 31:1117–1152.

[Hansen, 2000] Hansen, B. (2000). Testing for Structural Change in Conditional Models. *Journal of Econometrics*, 97.

[Hong and Li, 2004] Hong, Y. and Li, H. (2004). Nonparametric Specification Testing for Continuous-time Models with Applications to Term Structure of Interest Rates. *The Review of Financial Studies*, 18.

[Hong et al., 2007] Hong, Y., Li, H., and Zhao, F. (2007). Can the Random Walk Model be Beaten in Out-of-sample Density Forecasts? Evidence from Intraday Foreign Exchange Rates. *Journal of Econometrics*, 141.

[Inoue, 2001] Inoue, A. (2001). Testing for Distributional Change in Time Series. *Econometric Theory*, 17.

[Inoue and Rossi, 2012] Inoue, A. and Rossi, B. (2012). Out-of-sample Forecast Tests Robust to the Choice of Window Size. *Journal of Business & Economic Statistics*, 30.

[Jorda, 2005] Jorda, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95.

[Kim and Nelson, 1999] Kim, C. and Nelson, C. (1999). Has the us Economy Become More Stable? A Bayesian Approach Based on a Markov-switching Model of the Business Cycle. *Review of Economics and Statistics*, 81.

[Knuppel, 2015] Knuppel, M. (2015). Evaluating the Calibration of Multi-step-ahead Density Forecasts Using Raw Moments. *Journal of Business & Economic Statistics*, 33.

[Koenker and Bassett, 1978] Koenker, R. and Bassett, J. G. (1978). Regression Quantiles. *Econometrica: journal of the Econometric Society*.

[Koop and Korobilis, 2010] Koop, G. and Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. In *Foundations and TrendsÂ® in Econometrics*, volume 3.

[Koop et al., 1996] Koop, G., Pesaran, M., and Potter, S. (1996). Impulse Response Analysis in Nonlinear Multivariate Models. *Journal of econometrics*, 74.

[Korobilis, 2013] Korobilis, D. (2013). Assessing the Transmission of Monetary Policy Using Time-varying Parameter Dynamic Factor Models. *Oxford Bulletin of Economics and Statistics*, 75:157–179.

[Li and Racine, 2007] Li, Q. and Racine, J. (2007). In *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

[Lopez, 2001] Lopez, J. (2001). Evaluating the Predictive Accuracy of Volatility Models. *Journal of Forecasting*, 20.

[Ma and Su, 2016] Ma, S. and Su, L. (2016). Estimation of Large Dimensional Factor Models with an Unknown Number of Breaks. *Working Paper, University of California, Riverside and Singapore Management University*.

[Massacci, 2017] Massacci, D. (2017). Least Squares Estimation of Large Dimensional Threshold Factor Models. *Journal of Econometrics*, 197:101–129.

[McConnell and Perez-Quiros, 2000] McConnell, M. and Perez-Quiros, G. (2000). Output Fluctuations in the United States: What Has Changed Since the Early 1980's?. *American Economic Review*, 90.

[McCracken, 2000] McCracken, M. (2000). Robust Out-of-sample Inference. *Journal of Econometrics*, 99.

[Meese and Rogoff, 1983] Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14.

[Miranda-Agrippino and Ricco, 2018] Miranda-Agrippino, S. and Ricco, G. (2018). The Transmission of Monetary Policy Shocks.

[Muller and Petalas, 2010] Muller, U. and Petalas, P. (2010). Efficient Estimation of the Parameter Path in Unstable Time Series Models. *The Review of Economic Studies*, 77.

[Newey and West, 1987] Newey, W. and West, K. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55:703–708.

[Onatski, 2010] Onatski, A. (2010). Determining the Number of Factors from Empirical Distribution of Eigenvalues. *Review of Economics and Statistics*, 92:1004–1016.

[Primiceri, 2005] Primiceri, G. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72.

[Ramey, 2016] Ramey, V. (2016). Macroeconomic Shocks and Their Propagation. In *Handbook of Macroeconomics*, volume 2, pages 71–162.

[Ramey and Zubairy, 2018] Ramey, V. and Zubairy, S. (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from us Historical Data. *Journal of Political Economy*, 126.

[Reis and Watson, 2010] Reis, R. and Watson, M. W. (2010). Relative Goods' Prices, Pure Inflation, and the Phillips Correlation. *American Economic Journal: Macroeconomics*, 2:128–57.

[Rossi, 2013] Rossi, B. (2013). Advances in Forecasting under Instability. In *Handbook of Economic Forecasting*, volume 2, pages 1203–1324.

[Rossi and Sekhposyan, 2011] Rossi, B. and Sekhposyan, T. (2011). Understanding Models' Forecasting Performance. *Journal of Econometrics*, 164.

[Rossi and Sekhposyan, 2013] Rossi, B. and Sekhposyan, T. (2013). Conditional Predictive Density Evaluation in the Presence of Instabilities. *Journal of Econometrics*, 177.

[Rossi and Sekhposyan, 2017] Rossi, B. and Sekhposyan, T. (2017). Alternative Tests for Correct Specification of Conditional Predictive Densities.

[Sarno and Valente, 2009] Sarno, L. and Valente, G. (2009). Exchange Rates and Fundamentals: Footloose or Evolving Relationship? *Journal of the European Economic Association*, 7.

[Sims, 1980] Sims, C. (1980). Macroeconomics and Reality. *Econometrica*.

[Sims and Zha, 1998] Sims, C. and Zha, T. (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*.

[Stock and Watson, 1996] Stock, J. H. and Watson, M. W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics*, 14:11–30.

[Stock and Watson, 2001] Stock, J. H. and Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15:101–115.

[Stock and Watson, 2002a] Stock, J. H. and Watson, M. W. (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97:1167–1179.

[Stock and Watson, 2002b] Stock, J. H. and Watson, M. W. (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20:147–162.

[Stock and Watson, 2003] Stock, J. H. and Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, 41.

[Stock and Watson, 2009] Stock, J. H. and Watson, M. W. (2009). Forecasting in Dynamic Factor Models Subject to Structural Instability. In Castle, J. and Shephard, N., editors, *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*. Oxford University Press, Oxford.

[Stock and Watson, 2016] Stock, J. H. and Watson, M. W. (2016). Core Inflation and Trend Inflation. *Review of Economics and Statistics*, 98:770–784.

[Stock and Watson, 2018] Stock, J. H. and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal*, 128.

[Su and Wang, 2017] Su, L. and Wang, X. (2017). On Time-Varying Factor Models: Estimation and Testing. *Journal of Econometrics*, 198:84–101.

[Swanson, 1998] Swanson, N. (1998). Money and Output Viewed Through a Rolling Window. *Journal of Monetary Economics*, 41.

[Swanson and White, 1997] Swanson, N. and White, H. (1997). A Model Selection Approach to Real-time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, 79.

[Tay and Wallis, 2000] Tay, A. and Wallis, K. (2000). Density Forecasting: a Survey. *Journal of forecasting*, 19.

[Taylor, 1986] Taylor, S. (1986). Modelling Financial Time Series. *Wiley, Chichester*.

[Uhlig, 1997] Uhlig, H. (1997). Bayesian Vector Autoregressions with Stochastic Volatility. *Econometrica: Journal of the Econometric Society*.

[Vuong, 1989] Vuong, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. *Econometrica: Journal of the Econometric Society*.

[Wallis, 1977] Wallis, K. (1977). Multiple Time Series Analysis and the Final Form of Econometric Models. *Econometrica: Journal of the Econometric Society*.

[West, 1996] West, K. (1996). Asymptotic Inference about Predictive Ability. *Econometrica: Journal of the Econometric Society*.

[West and McCracken, 1998] West, K. and McCracken, M. (1998). Regression-based Tests of Predictive Ability. *International Economic Review*, 39:817–840.

[White, 1982] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica: Journal of the Econometric Society*.

[White, 2001] White, H. (2001). In *Asymptotic Theory for Econometricians*. Academic press.

[White and Domowitz, 1984] White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica: Journal of the Econometric Society*.

[Winkler, 1967] Winkler, R. (1967). The Quantification of Judgment: Some Methodological Suggestions. *Journal of the American Statistical Association*, 62.

[Wooldridge and White, 1988] Wooldridge, J. and White, H. (1988). Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes. *Econometric Theory*, 4.

[Yamamoto and Tanaka, 2015] Yamamoto, Y. and Tanaka, S. (2015). Testing for Factor Loading Structural Change under Common Breaks. *Journal of Econometrics*, 189:187–206.

[Zellner and Palm, 1974] Zellner, A. and Palm, F. (1974). Time Series Analysis and Simultaneous Equation Econometric Models. *Journal of Econometrics*, 2.