ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 36

Settore Concorsuale: 13/A5 - ECONOMETRIA

Settore Scientifico Disciplinare: SECS-P/05 - ECONOMETRIA

ESSAYS ON BOOTSTRAP METHODS IN ECONOMETRICS

Presentata da: Edoardo Zanelli

Coordinatore Dottorato

Andrea Mattozzi

Supervisore

Giuseppe Cavaliere

Esame finale anno 2025

Essays on Bootstrap Methods in Econometrics

Edoardo Zanelli
University of Bologna

Acknowledgments

Writing a doctoral thesis is a long and intense process, full of uncertainty, doubts and obstacles that may seem insurmountable for long times. If I apparently managed to be here writing this page is mostly due to the people I had the luck to encounter during my journey, whose help I partly want to acknowledge here.

First of all, I want to express all my gratitude to Giuseppe Cavaliere, for being the best supervisor I could have asked for. Since your first lecture I attended as a Master student, I wanted to be as cool and smart as you are. You guided my path since the beginning, and saw potential in my work even when my knowledge was extremely limited and the mistakes were trivial. But most of all, you never made me feel uncomfortable through my insecurities. I will never forget the time spent together, and will forever be grateful for the unparalleled honor to be calling myself your coauthor.

I want to thank my mentors and coauthors Sílvia Gonçalves, Morten Ørregaard Nielsen, Iliyan Georgiev and Prosper Dovonon for giving me the chance to learn how an econometrician should work from the bests. All of you have gifted me with different though valuable perspectives on how this profession should be done. Most importantly, you are really enjoyable people whom I like spending my time with.

Thanks to the Department of Economics at the University of Bologna for supporting my process as a PhD student, and especially to Michele Costa and Luca Fanelli for always having a word of encouragement in my regards and for making me feel a valuable resource for this prestigious department. Moreover, a special word goes to my colleague, desk-mate, but most of all, my best friend, Claudio Lissona, for the life-lasting conversations and our competition-less bond. I also want to thank the Departments of Economics at Concordia University and McGill University for hosting me during my visiting periods, which were extremely fruitful experiences, both personally and from a professional viewpoint.

I thank my mother and my father for taking all their effort in trying to understand choices that seemed unreasonable for them, for staying close to me at my lowest and for showing their pride in me to the world. Even if I seem to be reluctant to it, this is one thing that has always pushed me on. I thank my brother Filippo for never asking questions, but simply showing me his unconditional love everyday. You are, and will always be, my role model. I thank my family as a whole for following every step. In particular, I want to thank my little brother Tommaso and my grandparents: to those who are still here physically, because every kiss and hug is a gift, to those who are not, because your loving memory will always be part of me. Ciao Giorgio, mi manchi e ti penso ogni giorno.

I want to dedicate this work to a person who has profoundly and irreversibly made my life a better place to be in. Every smile, every song, every trip, every lunch, every poem are just realizations of what has always been inside me and us. Thank you for never discouraging expressing myself, thank you for making everyday a present and thank you for showing interest in the concept of indicator function. All the obstacles I mentioned are irrelevant if you are with me, as your sweet care is the purest form of love I have ever experienced. Simply, thank you for making me, everyday, a better man. This is to Giulia, the love of my life.

Abstract

This thesis proposes novel implementations of the bootstrap in econometrics. Chapter 1 explores bootstrap methods for test statistics showing an asymptotic bias which is difficult, or impossible, to estimate, proposing modifications of standard bootstrap methods delivering valid inference, totally bypassing bias estimation. Chapter 2 develops enhanced inference techniques for nonparametric regression and regression-discontinuity designs, introducing novel bootstrap approaches for debiasing with greater efficiency than the current state-of-the-art. Chapter 3 tackles the problem of invalidity of "standard" bootstrap methods in a predictive regression setup, when the predictability parameter may lie on the boundary of the parameter space, proposing a modified approach restoring bootstrap validity. Chapter 4 investigates the flattening of the Phillips Curve, presenting a time-varying structural estimation framework to disentangle underlying drivers of macroeconomic shifts. Finally, Chapter 5 contributes to robust inference on stochastic time-varying coefficients, proposing new confidence intervals which are robust to "large" – and more efficient – bandwidths. Collectively, these contributions advance theoretical and practical econometric tools for addressing complex real-world economic problems.

General Introduction

The development in analytical and computational methods at the economists' disposal over the last decades have allowed for the revision of old challenges with new tools. Among these, the bootstrap is a powerful method for several purposes related to inference, with bias correction, variance estimation and hypothesis testing being its main – but not exclusive – implementations.

This thesis explores novel and heterogeneous applications of the bootstrap in the field of econometrics and it is divided in 5 chapters, with each chapter targeting a distinct yet interconnected set of challenges, contributing to the advancement of robust and efficient econometric inference.

Chapter 1 is based on the paper *"Bootstrap Inference in the Presence of Bias"*, a joint work with Giuseppe Cavaliere, Sílvia Gonçalves and Morten Ørregaard Nielsen, recently published in the *Journal of the American Statistical Association*, focusing on the application of the bootstrap when test statistics show asymptotic bias. In particular, we focus on situations in which such asymptotic bias is difficult, or impossible, to estimate and "standard" bootstrap methods are invalid. This chapter proposes a solution that leverages the idea of prepivoting – originally proposed by Beran (1987, 1988) to deliver asymptotic refinements over first order asymptotics – to obtain valid (i.e., unbiased) confidence intervals and test statistics without the need to consistently estimate such bias term, thus ensuring asymptotically valid inference.

In Chapter 2, based on the paper *"Improved Inference for Nonparametric Regression and Regression-Discontinuity Designs"*, a joint work with Giuseppe Cavaliere, Sílvia Gonçalves and Morten Ørregaard Nielsen, we give an in-depth analysis on the applicability of the bootstrap and prepivoting to the nonparametric problem of inference on unknown function at a fixed point in their support. While being related to the results in the previous one, this chapter defines novel and insightful conclusions in this setup. Specifically, on the one hand, we show that prepivoting can be applied via two bootstrap algorithms, which we label the local-polynomial (LP) bootstrap and fixed-local (FL) bootstrap; on the other hand, we note that "standard" prepivoting might fail at the boundary of the support of the regressors, and we provide an ad hoc modification which is robust to the entire support (i.e., for boundary and interior points). Moreover, we show that the current state-of-the-art method in this class of problems (i.e., robust bias correction, see Calonico et al., 2014, 2018) is asymptotically equivalent to the proposed FL bootstrap. Finally, we compare the two proposed methods on the grounds of efficiency, showing that the LP bootstrap achieves up to $\sim 20\%$ shorter CIs asymptotically.

Chapter 3, based on the paper *"Parameters on the Boundary in Predictive Regression"*, a joint work with Giuseppe Cavaliere and Iliyan Georgiev, recently accepted for publication on *Econometric Theory*, investigates predictive regressions with parameters on the boundary of the parameter space, a scenario that invalidates standard bootstrap inference. We propose a modifications to standard bootstrap methods by shifting the bootstrap parameter space using data-dependent functions, therefore restoring bootstrap validity. These contributions are particularly relevant for testing hypotheses about predictability in financial markets, where parameter constraints often arise from economic theory.

The empirical relevance of econometric techniques is underscored in Chapter 4, based on the paper *"When did the Phillips Curve Become Flat? A Time-varying Estimate of Structural Parameters"*, written jointly with Claudio Lissona and Antonio Marsi. This chapter examines the flattening of the Phillips Curve by considering a time-varying structural estimation framework that combines instrumental variable methods with nonparametric estimation of impulse response functions. Inference, in this setup, is based on ad hoc bootstrap methods which guarantee robustness to time-variation of the time-varying parameters. By analyzing US data, the chapter identifies a declining structural slope in the Phillips Curve, attributing this to shifts in macroeconomic dynamics rather than increased monetary policy responsiveness. The findings provide nuanced insights into the interplay between inflation and unemployment over time, with significant implications for monetary policy.

Finally, Chapter 5 is a short note about models with time-varying coefficient that evolve stochastically as a random walk process with bounded variation. The chapter underlines the importance of adopting "large" bandwidths (e.g., chosen in a MSE-optimal sense) and shows that, under such choices, standard methods for constructing confidence intervals fail. We propose a alternative CIs that restore validity of inference by considering appropriately higher standard errors. Numerical simulations provide evidence in support of the proposed CIs, as well as the practical relevance of the bootstrap in this setup.

The contributions of this thesis lie at the intersection of theory and practice, addressing longstanding challenges while opening avenues for future research. By integrating novel bootstrap techniques, addressing boundary issues, and developing time-varying structural models, this work provides a robust toolkit for econometricians seeking to analyze complex economic phenomena.

Contents

Appendices

Chapter 1

Bootstrap Inference in the Presence of Bias

(written with Giuseppe Cavaliere, Sílvia Gonçalves and Morten Ørregaard Nielsen)

1.1 Introduction

Suppose that θ is a scalar parameter of interest and let $\hat{\theta}_n$ denote an estimator for which

$$T_n := g(n)(\hat{\theta}_n - \theta) \xrightarrow{d} B + \xi_1, \tag{1.1.1}$$

where $g(n) \to \infty$ is the rate of convergence of $\hat{\theta}_n$, ξ_1 is a continuous random variable centered at zero, and B is an asymptotic bias (our theory in fact allows for a more general formulation of the bias). A typical example is $g(n) = n^{1/2}$ and $\xi_1 \sim N(0, \sigma^2)$. Unless B can be consistently estimated, which is often difficult or impossible, classic (first-order) asymptotic inference on θ based on quantiles of ξ_1 in (1.1.1) is not feasible. Furthermore, the bootstrap, which is well known to deliver asymptotic refinements over first-order asymptotic approximations as well as bias corrections (Hall, 1992; Horowitz, 2001; Cattaneo and Jansson, 2018, 2022; Cattaneo, Jansson, and Ma, 2019), cannot in general be applied to solve the asymptotic bias problem when a consistent estimator of B does not exist. Examples are given below.

Our goal is to justify bootstrap inference based on T_n in the context of asymptotically biased estimators and where a consistent estimator of B does not exist. Consider the bootstrap statistic $T_n^* := g(n)(\hat{\theta}_n^* - \hat{\theta}_n)$, where $\hat{\theta}_n^*$ is a bootstrap version of $\hat{\theta}_n$, such that

$$T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1, \tag{1.1.2}$$

where $\hat{B}_n$ is the implicit bootstrap bias, and '$\xrightarrow{d^*}_p$' denotes weak convergence in probability (defined below). When $\hat{B}_n - B = o_p(1)$, the bootstrap is asymptotically valid in the usual sense that the bootstrap distribution of T_n^* is consistent for the asymptotic distribution of T_n, i.e., $\sup_{x \in \mathbb{R}} |P^*(T_n^* \leq x) - P(T_n \leq x)| = o_p(1)$.

We consider situations where $\hat{B}_n - B$ is not asymptotically negligible so the bootstrap fails to replicate the asymptotic bias. For example, this happens when the asymptotic bias term in the bootstrap world includes a random (additive) component, i.e.

$$\hat{B}_n - B \xrightarrow{d} \xi_2 \text{ (jointly with (1.1.1))}, \tag{1.1.3}$$

where ξ_2 is a random variable centered at zero. In this case, the bootstrap distribution is random in the limit and hence cannot mimic the asymptotic distribution given in (1.1.1). Moreover, the distribution of the bootstrap p-value, $\hat{p}_n := P^*(T_n^* \leq T_n)$, is not asymptotically uniform, and the bootstrap cannot in general deliver hypothesis tests (or confidence intervals) with the desired null rejection probability (or coverage probability).

In this paper, we show that in this non-standard case valid inference can successfully be restored by proper implementation of the bootstrap. This is done by focusing on properties of the bootstrap p-value rather than on the bootstrap as a means of estimating limiting distributions, which is infeasible due to the asymptotic bias. In particular, we show that such implementations lead to bootstrap inferences that are valid in the sense that they provide asymptotically uniformly distributed p-values.

Our inference strategy is based on the fact that, for some bootstrap schemes, the large-sample distribution of the bootstrap p-value, say $H(u)$, $u \in [0,1]$, although not uniform, does not depend on B. That is, we can search for bootstrap algorithms which generate bootstrap p-values that, in large samples, are not affected by unknown bias terms. When this is possible, we can make use of the prepivoting approach of Beran (1987, 1988), which — as we will show in this paper — allows to restore bootstrap validity. Specifically, our proposed modified p-value is defined as

$$\tilde{p}_n := \hat{H}_n(\hat{p}_n),$$

where $\hat{H}_n(u)$ is any consistent estimator of $H(u)$, uniformly over $u \in [0,1]$. The (asymptotic) probability integral transform $\hat{p}_n \mapsto H(\hat{p}_n)$, continuity of $H(u)$, and consistency of $\hat{H}_n(u)$ then guarantee that $\tilde{p}_n$ is asymptotically uniformly distributed. Interestingly, Beran (1987, 1988) proposed this approach to obtain asymptotic refinements for the bootstrap, but did not consider asymptotically biased estimators as we do here.

We propose two approaches to estimating H. First, if $H = H_\gamma$, where γ is a finite-dimensional parameter vector, and a consistent estimator $\hat{\gamma}_n$ of γ is available, then a 'plug-in' approach setting $\hat{H}_n = H_{\hat{\gamma}_n}$ can deliver asymptotically uniform p-values. Second, if estimation of γ is difficult (e.g., when γ does not have a closed form expression), we can use a 'double bootstrap' scheme (Efron, 1983; Hall, 1986), where estimation of H is achieved by resampling from the bootstrap data originated in the first level.

For both methods, we provide general high-level conditions that imply validity of the proposed approach. Our conditions are not specific to a given bootstrap method; rather, they can in principle be applied to any bootstrap scheme satisfying the proposed sufficient conditions for asymptotic validity.

Our approach is related to recent work by Shao and Politis (2013) and Cavaliere and Georgiev (2020). In particular, a common feature is that the distribution function of the bootstrap statistic, conditional on the original data, is random in the limit. Cavaliere and Georgiev (2020) emphasize that randomness of the limiting bootstrap measure does not prevent the bootstrap from delivering an asymptotically uniform p-value (bootstrap 'unconditional' validity), and provide results to assess such asymptotic uniformity. Our context is different, since the presence of an asymptotic bias term renders the distribution of the bootstrap p-value non-uniform, even asymptotically. In this respect, our work is related to Shao and Politis (2013), who show that t-statistics based on subsampling or block bootstrap methods with bandwidth proportional to sample size may deliver non-uniformly distributed p-values that, however, can be estimated.

To illustrate the practical relevance of our results and to show how to implement them in applied problems, we consider three examples involving estimators that feature an asymptotic bias term. In the first two examples (model averaging and ridge regression), B is not consistently estimable due to the presence of local-to-zero parameters and the standard bootstrap fails. In the third example (nonparametric regression), the bootstrap fails because B depends on the second-order derivative of the conditional mean function, whose estimation requires the use of a different (suboptimal) bandwidth. In these examples, ξ_1 is normal, but $g(n)$ and B are example-specific. Two additional examples are presented in the supplement. The fourth is a simple location model without the assumption of finite variance, where ξ_1 is not normal and estimators converge at an unknown rate. The fifth example considers inference for dynamic panel data models, where B is the incidental parameter bias.

The remainder of the paper is organized as follows. In Section 1.2 we introduce our three leading examples. Section 1.3 contains our general results, which we apply to the three examples in Section 1.4. Section 1.5 concludes. The supplemental material contains two appendices. Appendix A.1 specializes the general theory to the case of asymptotically Gaussian statistics, and Appendix A.2 contains details and proofs for the three leading examples, as well as two additional examples.

Notation

Throughout this paper, the notation $\sim$ indicates equality in distribution. For instance, $Z \sim N(0,1)$ means that Z is distributed as a standard normal random variable. We write '$x := y$' and '$y =: x$' to mean that x is defined by y. The standard Gaussian cumulative distribution function (cdf) is denoted by Φ; $U_{[0,1]}$ is the uniform distribution on $[0,1]$, and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. If F is a cdf, F^{-1} denotes the generalized inverse, i.e. the quantile function, $F^{-1}(u) := \inf\{v \in \mathbb{R} : F(v) \geq u\}$, $u \in \mathbb{R}$. Unless specified otherwise, all limits are for $n \to \infty$. For matrices a, b, c with n rows, we let $S_{ab} := a'b/n$

and $S_{ab.c} := S_{ab} - S_{ac}S_{cc}^{-1}S_{cb}$, assuming that S_{cc} has full rank.

For a (single level or first-level) bootstrap sequence, say Y_n^*, we use $Y_n^* \xrightarrow{p^*}_p 0$, or equivalently $Y_n^* \xrightarrow{p^*} 0$, in probability, to mean that, for any $\epsilon > 0$, $P^*(|Y_n^*| > \epsilon) \to_p 0$, where P^* denotes the probability measure conditional on the original data D_n. An equivalent notation is $Y_n^* = o_{p^*}(1)$ (where we omit the qualification "in probability" for brevity). Similarly, for a double (or second-level) bootstrap sequence, say Y_n^{**}, we write $Y_n^{**} = o_{p^{**}}(1)$ to mean that for all $\epsilon > 0$, $P^{**}(|Y_n^{**}| > \epsilon) \xrightarrow{p^*}_p 0$, where P^{**} is the probability measure conditional on the first-level bootstrap data D_n^* and on D_n.

We use $Y_n^* \xrightarrow{d^*}_p \xi$, or equivalently $Y_n^* \xrightarrow{d^*} \xi$, in probability, to mean that, for all continuity points $u \in \mathbb{R}$ of the cdf of ξ, say $G(u) := P(\xi \leq u)$, it holds that $P^*(Y_n^* \leq u) - G(u) \to_p 0$. Similarly, for a double bootstrap sequence Y_n^{**}, we use $Y_n^{**} \xrightarrow{d^{**}}_{p^*} \xi$, in probability, to mean that $P^{**}(Y_n^{**} \leq u) - G(u) \xrightarrow{p^*}_p 0$ for all continuity points u of G.

1.2 Examples

In this section we introduce our three leading examples. Example-specific regularity conditions, formally stated results, and additional definitions are given in Appendix A.2. For each of these examples, we argue that (1.1.1), (1.1.2), and (1.1.3) hold, such that the bootstrap p-values $\hat{p}_n$ are not uniformly distributed rendering standard bootstrap inference invalid. We then return to each example in Section 1.4, where we discuss how to implement our proposed method and prove its validity.

1.2.1 Inference after model averaging

Setup. We consider inference based on a model averaging estimator obtained as a weighted average of least squares estimates (Hansen, 2007). Assume that data are generated according to the linear model

$$y = x\beta + Z\delta + \varepsilon, \tag{1.2.1}$$

where β is the (scalar) parameter of interest and ε is an n-vector of identically and independently distributed random variables with mean zero and variance σ^2 (henceforth i.i.d.$(0, \sigma^2)$), conditional on $W := (x, Z)$.

The researcher fits a set of M models, each of them based on different exclusion restrictions on the q-dimensional vector δ. This setup allows for model averaging both explicitly and implicity. The former follows, e.g., Hansen (2007). The latter includes the common practice of robustness checks in applied research, where the significance of a target coefficient is evaluated through an (often informal) assessment of its significance across a set of regressions based on different sets of controls; see Oster (2019) and the references therein. Specifically, letting R_m denote a $q \times q_m$ selection matrix, the m^{th} model includes x and $Z_m := ZR_m$ as regressors, and the corresponding OLS estimator of β is $\tilde{\beta}_{m,n} = S_{xx.Z_m}^{-1} S_{xy.Z_m}$. Given a set of fixed weights $\omega := (\omega_1, \ldots, \omega_M)'$ such that

$\omega_m \in [0,1]$ and $\sum_{m=1}^M \omega_m = 1$, the model averaging estimator is $\tilde{\beta}_n := \sum_{m=1}^M \omega_m \tilde{\beta}_{m,n}$. Then $T_n := n^{1/2}(\tilde{\beta}_n - \beta)$ satisfies $T_n - B_n \to_d \xi_1 \sim N(0, v^2)$, where $v^2 > 0$ and

$$B_n := Q_n n^{1/2}\delta, \quad Q_n := \sum_{m=1}^M \omega_m S^{-1}_{xx.Z_m} S_{xZ.Z_m}.$$

Thus, the magnitude of the asymptotic bias B_n depends on $n^{1/2}\delta$. If δ is local to zero in the sense that $\delta = cn^{-1/2}$ for some vector $c \in \mathbb{R}^q$ (as in, e.g., Hjort and Claeskens, 2003; Liu, 2015; Hounyo and Lahiri, 2023), then $B_n \to_p B := Qc$ with $Q := \operatorname{plim} Q_n$, so that (1.1.1) is satisfied with nonzero B in general. Because B depends on c, which is not consistently estimable, we cannot obtain valid inference from a Gaussian distribution based on sample analogues of B and v^2.

Fixed regressor bootstrap. We generate the bootstrap sample as $y^* = x\hat{\beta}_n + Z\hat{\delta}_n + \varepsilon^*$, where $\varepsilon^*|D_n \sim N(0, \hat{\sigma}^2_n I_n)$, $(\hat{\beta}_n, \hat{\delta}'_n, \hat{\sigma}^2_n)$ is the OLS estimator from the full model, and $D_n = \{y, W\}$. Similar results can be established for the nonparametric bootstrap where ε^* is resampled from the full model residuals. The bootstrap model averaging estimator is given by $\tilde{\beta}^*_n := \sum_{m=1}^M \omega_m \tilde{\beta}^*_{m,n}$, where $\tilde{\beta}^*_{m,n} := S^{-1}_{xx.Z_m} S_{xy^*.Z_m}$. Letting $T^*_n := n^{1/2}(\tilde{\beta}^*_n - \hat{\beta}_n)$, we can show that (1.1.2) holds with $\hat{B}_n := Q_n n^{1/2}\hat{\delta}_n$ such that, as in (1.1.3),

$$\hat{B}_n - B_n = Q_n n^{1/2}(\hat{\delta}_n - \delta) \xrightarrow{d} \xi_2 \sim N(0, v_{22}), \quad v_{22} > 0,$$

given in particular the asymptotic normality of $n^{1/2}(\hat{\delta}_n - \delta)$. Because the bias term in the bootstrap world is random in the limit, the conditional distribution of T^*_n is also random in the limit, and in particular does not mimic the asymptotic distribution of the original statistic T_n.

Pairs bootstrap. Consider now a pairs (random design) bootstrap sample $\{y^*_t, x^*_t, z^*_t; t = 1, \ldots, n\}$, based on resampling with replacement from the tuples $\{y_t, x_t, z_t; t = 1, \ldots, n\}$. As is standard, it is useful to recall that the bootstrap data have the representation

$$y^* = x^*\hat{\beta}_n + Z^*\hat{\delta}_n + \varepsilon^*,$$

where $\varepsilon^* = (\varepsilon^*_1, \ldots, \varepsilon^*_n)'$ and ε^*_t is an i.i.d. draw from $\hat{\varepsilon}_t = y_t - x_t\hat{\beta}_n - z'_t\hat{\delta}_n$. The pairs bootstrap model averaging estimator is

$$\tilde{\beta}^*_n := \sum_{m=1}^M \omega_m \tilde{\beta}^*_{m,n} \text{ with } \tilde{\beta}^*_{m,n} := S^{-1}_{x^*x^*.Z^*_m} S_{x^*y^*.Z^*_m}$$

and $Z^*_m = Z^* R_m$. The pairs bootstrap statistic is then

$$T^*_n := n^{1/2}(\tilde{\beta}^*_n - \hat{\beta}_n) = B^*_n + n^{1/2} S^{-1}_{x^*x^*} S_{x^*\varepsilon^*},$$

where

$$B^*_n := \sum_{m=1}^M \omega_m S^{-1}_{x^*x^*.Z^*_m} S_{x^*Z^*.Z^*_m} n^{1/2}\hat{\delta}_n.$$

Therefore, and in contrast with the fixed regressor bootstrap (FRB), the term B_n^* is stochastic under the bootstrap probability measure and replaces the bias term $\hat{B}_n$. This difference is not innocuous because it implies that $T_n^* - \hat{B}_n$ no longer replicates the asymptotic distribution of $T_n - B_n$ and (1.1.2) does not hold. However, this does not prevent our method from working, but it will require a different set of conditions which we will give in Section 1.3.5.

1.2.2 Ridge regression

Setup. We consider estimation of a vector of regression parameters through regularization; in particular, by using a ridge estimator. The model is $y_t = \theta' x_t + \varepsilon_t$, $t = 1, \ldots, n$, where x_t is a $p \times 1$ non-stochastic vector and $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$. Interest is on testing $\mathsf{H}_0 : g'\theta = r$, based on ridge estimation of θ. Specifically, the ridge estimator has closed form expression $\tilde{\theta}_n = \tilde{S}_{xx}^{-1} S_{xy}$, where $\tilde{S}_{xx} := S_{xx} + n^{-1} c_n I_p$ and c_n is a tuning parameter that controls the degree of shrinkage towards zero. Clearly, $c_n = 0$ corresponds to the OLS estimator, $\hat{\theta}_n$. We are interested in the case where the regressors have limited explanatory power, i.e., where $\theta = \delta n^{-1/2}$ is local to zero, which can in fact be taken as a motivation for shrinkage towards zero and hence for ridge estimation. To test H_0, we consider the test statistic $T_n = n^{1/2}(g'\tilde{\theta}_n - r)$. If $n^{-1} c_n \to c_0 \geq 0$ (as in, e.g., Fu and Knight, 2000) then, under the null, it holds that $T_n - B_n \to_d \xi_1 \sim N(0, v^2)$, where

$$B_n := -c_n n^{-1/2} g' \tilde{S}_{xx}^{-1} \theta = -c_n n^{-1} g' \tilde{S}_{xx}^{-1} \delta \to B := -c_0 g' \tilde{\Sigma}_{xx}^{-1} \delta$$

with $\tilde{\Sigma}_{xx} := \Sigma_{xx} + c_0 I_p$ and $\Sigma_{xx} := \lim S_{xx}$. Hence, for $c_0 > 0$, $\tilde{\theta}_n$ is asymptotically biased and the bias term cannot be consistently estimated. Consequently, (1.1.1) is satisfied, and inference based on the quantiles of the $N(0, v^2)$ distribution is invalid unless $c_0 = 0$.

Bootstrap. Consider a pairs (random design) bootstrap sample $\{y_t^*, x_t^*; t = 1, \ldots, n\}$ built by i.i.d. resampling from the tuples $\{y_t, x_t; t = 1, \ldots, n\}$. The bootstrap analogue of the ridge estimator is $\tilde{\theta}_n^* := \tilde{S}_{x^*x^*}^{-1} S_{x^*y^*}$, where $\tilde{S}_{x^*x^*} := S_{x^*x^*} + n^{-1} c_n I_p$. The bootstrap statistic is $T_n^* := n^{1/2} g'(\tilde{\theta}_n^* - \hat{\theta}_n)$, which is centered using $\hat{\theta}_n$ to guarantee that ε_t^* and x_t^* are uncorrelated in the bootstrap world. Because we have used a pairs bootstrap, we now have $T_n^* - B_n^* \xrightarrow{d}_{p^*} \xi_1$ for $B_n^* := -c_n n^{-1/2} g' \tilde{S}_{x^*x^*}^{-1} \hat{\theta}_n$. However, $B_n^* - \hat{B}_n = o_{p^*}(1)$ with $\hat{B}_n := -c_n n^{-1/2} g' \tilde{S}_{xx}^{-1} \hat{\theta}_n$, such that $T_n^* - \hat{B}_n$ still satisfies (1.1.2). Then (1.1.3) holds with

$$\hat{B}_n - B_n = -c_n n^{-1} g' \tilde{S}_{xx}^{-1} n^{1/2} (\hat{\theta}_n - \theta) \xrightarrow{d} \xi_2 \sim N(0, v_{22}), \quad v_{22} > 0,$$

so the bootstrap fails to approximate the asymptotic distribution of T_n (see also Chatterjee and Lahiri, 2010, 2011).

1.2.3 Nonparametric regression

Setup. Consider the model

$$y_t = \beta(x_t) + \varepsilon_t, \quad t = 1, \dots, n, \tag{1.2.2}$$

where $\beta(\cdot)$ is a smooth function and $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$. For simplicity, we consider a fixed-design model; i.e., $x_t = t/n$. The goal is inference on $\beta(x)$ for a fixed $x \in (0,1)$. We apply the standard Nadaraya-Watson (fixed-design) estimator $\hat{\beta}_h(x) = (nh)^{-1} \sum_{t=1}^n K((x_t - x)/h)y_t$, where $h = cn^{-1/5}$ for some $c > 0$ is the MSE-optimal bandwidth and K is the kernel function. We do not consider the more general local polynomial regression case, although we conjecture that very similar results will hold. We leave that case for future research. The statistic $T_n = (nh)^{1/2}(\hat{\beta}_h(x) - \beta(x))$ satisfies $T_n - B_n \to_d \xi_1 \sim N(0, v^2)$, where $v^2 := \sigma^2 \int K(u)^2 du > 0$ and

$$B_n := (nh)^{1/2} \left(\frac{1}{nh} \sum_{t=1}^n k_t \beta(x_t) - \beta(x) \right) \tag{1.2.3}$$

with $k_t := K((x_t - x)/h)$. The bias B_n satisfies

$$B_n = (nh)^{1/2}(h^2 \beta''(x)\kappa_2/2 + o(h^2)) \to B := c^{5/2}\beta''(x)\kappa_2/2, \tag{1.2.4}$$

where $\kappa_2 := \int u^2 K(u) du$ and $\beta''(x)$ denotes the second-order derivative of $\beta(x)$. Thus, (1.1.1) is satisfied. Estimating B or B_n is challenging because it involves estimating $\beta''(x)$, and although theoretically valid estimators exist, they perform poorly in finite samples. This issue is pointed out by Calonico, Cattaneo, and Titunik (2014) and Calonico, Cattaneo, and Farrell (2018), who propose more accurate bias correction techniques specifically for regression discontinuity designs and nonparametric curve estimation.

Bootstrap. The (parametric) bootstrap sample is generated as $y_t^* = \hat{\beta}_h(x_t) + \varepsilon_t^*$, $t = 1, \dots, n$, where $\varepsilon_t^* | D_n \sim$ i.i.d.$N(0, \hat{\sigma}_n^2)$ with $D_n = \{y_t, t = 1, \dots, n\}$ and $\hat{\sigma}_n^2$ denotes a consistent estimator of σ^2; e.g. the residual variance. Let $\hat{\beta}_h^*(x) = (nh)^{-1} \sum_{t=1}^n k_t y_t^*$ and $T_n^* = (nh)^{1/2}(\hat{\beta}_h^*(x) - \hat{\beta}_h(x))$. Then (1.1.2) is satisfied with

$$\hat{B}_n := (nh)^{1/2} \left(\frac{1}{nh} \sum_{t=1}^n k_t \hat{\beta}_h(x_t) - \hat{\beta}_h(x) \right).$$

Because $h = cn^{-1/5}$, (1.1.3) holds with

$$\hat{B}_n - B_n = (nh)^{1/2} \left(\frac{1}{nh} \sum_{t=1}^n k_t (\hat{\beta}_h(x_t) - \beta(x_t)) - (\hat{\beta}_h(x) - \beta(x)) \right) \xrightarrow{d} \xi_2 \sim N(0, v_{22}),$$

where $v_{22} > 0$, so the bootstrap is invalid. Two possible solutions to this problem are to generate the bootstrap sample as $y_t^* = \hat{\beta}_g(x_t) + \varepsilon_t^*$, where g is an oversmoothing bandwidth satisfying $ng^5 \to \infty$ (e.g., Härdle and Marron, 1991) or to center the bootstrap statistic at its expected value and add a consistent estimator of B (e.g., Härdle and Bowman, 1988; Eubank and Speckman, 1993). Both approaches require selecting two bandwidths, which is not straightforward. An alternative approach suggested by Hall and Horowitz (2013) focuses on an asymptotic theory-based confidence interval and applies the bootstrap to calibrate its coverage probability. However, this requires an additional averaging step across a grid of x (their step 6) to asymptotically eliminate ξ_2, and it results in an asymptotically conservative interval. Finally, a non-bootstrap-based solution is undersmoothing using a bandwidth h satisfying $nh^5 \to 0$, although of course that is not MSE-optimal and may result in trivial power against certain local alternatives; see Section 1.4.3.

1.3 General results

1.3.1 Framework and invalidity of the standard bootstrap

The general framework is as follows. We have a statistic T_n defined as a general function of a sample D_n, for which we would like to compute a valid bootstrap p-value. Usually T_n is a test statistic or a (possibly normalized) parameter estimator; for example, $T_n = g(n)(\hat{\theta}_n - \theta_0)$. Let D_n^* denote the bootstrap sample, which depends on the original data and on some auxiliary bootstrap variates (which we assume defined jointly with D_n on a possibly extended probability space). Let T_n^* denote the bootstrap version of T_n computed on D_n^*; for example, $T_n^* = g(n)(\hat{\theta}_n^* - \hat{\theta}_n)$. Let $\hat{L}_n(u) := P^*(T_n^* \leq u)$, $u \in \mathbb{R}$, denote its distribution function, conditional on the original data. The bootstrap p-value is defined as

$$\hat{p}_n := P^*(T_n^* \leq T_n) = \hat{L}_n(T_n).$$

First-order asymptotic validity of $\hat{p}_n$ requires that $\hat{p}_n$ converges in distribution to a standard uniform distribution; i.e., that $\hat{p}_n \to_d U_{[0,1]}$. In this section we focus on a class of statistics T_n and T_n^* for which this condition is not necessarily satisfied. The main reason is the presence of an additive 'bias' term B_n that contaminates the distribution of T_n and cannot be replicated by the bootstrap distribution of T_n^*.

$T_n - B_n \to_d \xi_1$, where ξ_1 is centered at zero and the cdf $G_\gamma(u) = P(\xi_1 \leq u)$ is continuous and strictly increasing over its support.

When B_n converges to a nonzero constant B, Assumption 1.3.1 can be written $T_n \to_d B + \xi_1$ as in (1.1.1). If T_n is a normalized version of a (scalar) parameter estimator, i.e., $T_n = g(n)(\hat{\theta}_n - \theta_0)$, then we can think of B as the asymptotic bias of $\hat{\theta}_n$ because ξ_1 is centered at zero. Although we allow for the possibility that B_n does not have a limit (and it may even diverge), we will still refer to B_n as a 'bias term'. More generally, in Assumption 1.3.1 we cover any statistic T_n that is not necessarily Gaussian (even asymptotically) and whose limiting distribution is G_γ only after we subtract the sequence B_n. We index the limiting distribution G_γ by a parameter γ to allow for the possibility that

$T_n - B_n$ is not an asymptotic pivot.

Inference based on the asymptotic distribution of T_n requires estimating B_n and γ. Alternatively, we can use the bootstrap to bypass estimation of B_n and γ and directly compute a bootstrap p-value that relies on T_n^* and T_n alone; that is, we consider $\hat{p}_n := P^*(T_n^* \leq T_n)$. A set of high-level conditions on T_n^* and T_n that allow us to derive the asymptotic properties of this p-value are described next.

For some D_n-measurable random variable $\hat{B}_n$, it holds that: (i) $T_n^* - \hat{B}_n \overset{d^*}{\rightarrow}_p \xi_1$, where ξ_1 is described in Assumption 1.3.1; (ii)

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} \overset{d}{\rightarrow} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

where ξ_2 is centered at zero and $F(u) = P(\xi_1 - \xi_2 \leq u)$ is a continuous cdf.

Assumption 1.3.1(i) states that $T_n^* - \hat{B}_n$ converges in distribution to a random variable ξ_1 having the same distribution function G_γ as $T_n - B_n$.[1] Thus, $\hat{B}_n$ can be thought of as an implicit bootstrap bias that affects the statistic T_n^*, in the same way that B_n affects the original statistic T_n. Assumption 1.3.1(ii) complements Assumption 1.3.1 by requiring the joint convergence of $T_n - B_n$ and $\hat{B}_n - B_n$ towards ξ_1 and ξ_2, respectively; see also (1.1.1)–(1.1.3).

Given Assumption 1.3.1(i), we could use the bootstrap distribution of $T_n^* - \hat{B}_n$ to approximate the distribution of $T_n - B_n$. Since B_n is typically unknown, this result is not very useful for inference unless $\hat{B}_n$ is consistent for B_n. In this case, Assumption 1.3.1 together with Assumption 1.3.1 imply that $\hat{p}_n$ is asymptotically distributed as $U_{[0,1]}$. This follows by noting that if $\hat{B}_n - B_n = o_p(1)$, then $\xi_2 = 0$ a.s., implying that $F_\phi(u) = G_\gamma(u)$. Consequently,

$$\begin{aligned} \hat{p}_n &:= P^*(T_n^* \leq T_n) = P^*(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n) \\ &= G_\gamma(T_n - \hat{B}_n) + o_p(1) \text{ (by Assumption 1.3.1(i))} \\ &\overset{d}{\rightarrow} G_\gamma(\xi_1 - \xi_2) \text{ (by Assumption 1.3.1(ii) and continuity of } G_\gamma) \\ &\sim U_{[0,1]}, \end{aligned}$$

where the last distributional equality holds by $F_\phi = G_\gamma$ and the probability integral transform. However, this result does not hold if $\hat{B}_n - B_n$ does not converge to zero in probability. Specifically, if $\hat{B}_n - B_n \rightarrow_d \xi_2$ (jointly with $T_n - B_n \rightarrow_d \xi_1$), then

$$T_n - \hat{B}_n = (T_n - B_n) - (\hat{B}_n - B_n) \overset{d}{\rightarrow} \xi_1 - \xi_2 \sim F_\phi^{-1}(U_{[0,1]})$$

[1] Note that we write $T_n^* - \hat{B}_n \overset{d^*}{\rightarrow}_p \xi_1$ to mean that $T_n^* - \hat{B}_n$ has (conditionally on D_n) the same asymptotic distribution function as the random variable ξ_1. We could alternatively write that $T_n^* - \hat{B}_n \overset{d^*}{\rightarrow}_p \xi_1^*$ and $T_n - B_n \overset{d}{\rightarrow} \xi_1$ where ξ_1^* and ξ_1 are two independent copies of the same distribution, i.e. $P(\xi_1 \leq u) = P(\xi_1^* \leq u)$. We do not make this distinction because we care only about distributional results, but it should be kept in mind.

under Assumptions 1.3.1 and 1.3.1(ii). When ξ_2 is nondegenerate, $F_\phi \neq G_\gamma$, implying that $\hat{p}_n = G_\gamma(T_n - \hat{B}_n) + o_p(1)$ is not asymptotically distributed as a standard uniform random variable. This result is summarized in the following theorem.

THEOREM 1.3.1 *Suppose Assumptions 1.3.1 and 1.3.1 hold. Then* $\hat{p}_n \to_d G_\gamma(F_\phi^{-1}(U_{[0,1]}))$.

PROOF. First notice that $\hat{p}_n$ and $G_\gamma(T_n - \hat{B}_n)$ have the same asymptotic distribution because Assumption 1.3.1(i) and continuity of G_γ imply that, by Polya's Theorem,

$$|\hat{p}_n - G_\gamma(T_n - \hat{B}_n)| \leq \sup_{u\in\mathbb{R}} |P^*(T_n^* - \hat{B}_n \leq u) - G_\gamma(u)| \xrightarrow{p} 0.$$

Next, by Assumption 1.3.1(ii), $T_n - \hat{B}_n \to_d \xi_1 - \xi_2$, such that

$$G_\gamma(T_n - \hat{B}_n) \xrightarrow{d} G_\gamma(\xi_1 - \xi_2)$$

by continuity of G_γ and the continuous mapping theorem. Since $\xi_1 - \xi_2$ has continuous cdf F_ϕ, it holds that $\xi_1 - \xi_2 \sim F_\phi^{-1}(U_{[0,1]})$, which completes the proof. □

REMARK 1.3.1 *The value of $\hat{B}_n$ in Assumption 1.3.1(i) depends on the chosen bootstrap algorithm. It is possible that $\hat{B}_n \to_p 0$ for some bootstrap algorithms; examples are given in Remark A.2.2 and Appendix A.2.5. If this is the case, then $\xi_2 = -B$ a.s., which implies that*

$$F_\phi(u) := P(\xi_1 - \xi_2 \leq u) = P(\xi_1 \leq u - B) = G_\gamma(u - B),$$

and hence Assumption 1.3.1(ii) is not satisfied. In this case the bootstrap p-value satisfies

$$\hat{p}_n \xrightarrow{d} G_\gamma(G_\gamma^{-1}(U_{[0,1]}) + B).$$

Note that this distribution is uniform only if $B = 0$. Hence, the p-value depends on B, even in the limit.

REMARK 1.3.2 *Under Assumptions 1.3.1 and 1.3.1, standard bootstrap (percentile) confidence sets are also in general invalid. Consider, e.g., the case where $T_n = g(n)(\hat{\theta}_n - \theta_0)$ and T_n^* is its bootstrap analogue with (conditional) distribution function $\hat{L}_n(u)$. A right-sided confidence set for θ_0 at nominal confidence level $1-\alpha \in (0,1)$ can be obtained as (e.g., Horowitz, 2001, p. 3171) $CI_n^{1-\alpha} := [\hat{\theta}_n - g(n)^{-1}\hat{q}_n(1-\alpha), +\infty)$, where $\hat{q}_n(1-\alpha) := \hat{L}_n^{-1}(1-\alpha)$. Then*

$$\begin{aligned} P(\theta_0 \in CI_n^{1-\alpha}) &= P(\hat{\theta}_n - g(n)^{-1}\hat{q}_n(1-\alpha) \leq \theta_0) = P(T_n \leq \hat{q}_n(1-\alpha)) \\ &= P(\hat{L}_n(T_n) \leq 1-\alpha) = P(\hat{p}_n \leq 1-\alpha) \nrightarrow 1-\alpha \end{aligned}$$

because by Theorem 1.3.1 $\hat{p}_n$ is not asymptotically uniformly distributed.

Remark 1.3.3 *It is worth noting that, under Assumptions 1.3.1 and 1.3.1, the bootstrap (conditional) distribution is random in the limit whenever* ξ_2 *is non-degenerate. Specifically, assume for simplicity that* $B_n \to_p B$*. Recall that* $\hat{L}_n(u) := P^*(T_n^* \leq u)$*,* $u \in \mathbb{R}$*, and let* $\hat{G}_{\gamma,n}(u) := P^*(T_n^* - \hat{B}_n \leq u)$*. It then holds that*

$$\hat{L}_n(u) = \hat{G}_{\gamma,n}(u - \hat{B}_n) = G_\gamma(u - B - (\hat{B}_n - B)) + \hat{a}_n(u),$$

where $\hat{a}_n(u) \leq \sup_{u\in\mathbb{R}} |\hat{G}_{\gamma,n}(u) - G_\gamma(u)| = o_p(1)$ *by Assumption 1.3.1(i), continuity of* G_γ*, and Polya's Theorem. Because* $\hat{B}_n - B \to_d \xi_2$*, it follows that when* ξ_2 *is non-degenerate,* $\hat{L}_n(u) \to_w G_\gamma(u - B - \xi_2)$*, where* $\to_w$ *denotes weak convergence of cdf's as (random) elements of a function space (see Cavaliere and Georgiev, 2020). The presence of* ξ_2 *in* $G_\gamma(u - B - \xi_2)$ *makes this a random cdf.*[2] *Therefore, the bootstrap is unable to mimic the asymptotic distribution of* T_n*, which is* $G_\gamma(u - B)$ *by Assumption 1.3.1.*

Next, we describe two possible solutions to the invalidity of the standard bootstrap p-value $\hat{p}_n$. One relies on the prepivoting approach of Beran (1987, 1988); see Section 1.3.2. The basic idea is that we modify $\hat{p}_n$ by applying the mapping $\hat{p}_n \mapsto H(\hat{p}_n)$, where $H(u)$ is the asymptotic cdf of $\hat{p}_n$, which makes the modified p-value $H(\hat{p}_n)$ asymptotically standard uniform. Contrary to Beran (1987, 1988), who proposed prepivoting as a way of providing asymptotic refinements for the bootstrap, here we show how to use prepivoting to solve the invalidity of the standard bootstrap p-value $\hat{p}_n$. This result is new in the bootstrap literature. The second approach relies on computing a standard bootstrap p-value based on the modified statistic given by $T_n - \hat{B}_n$; see Section 1.3.4. Thus, we modify the test statistic rather than modifying the way we compute the bootstrap p-value.

1.3.2 Prepivoting

Theorem 1.3.1 implies that

$$P(\hat{p}_n \leq u) \to P(G_\gamma(F_\phi^{-1}(U_{[0,1]})) \leq u) = P(U_{[0,1]} \leq F_\phi(G_\gamma^{-1}(u))) = F_\phi(G_\gamma^{-1}(u)) =: H_{\phi,\gamma}(u) =: H(u)$$

uniformly over $u \in [0, 1]$ by Polya's Theorem, given the continuity of G_γ and F_ϕ. Although H is not the uniform distribution, unless $G_\gamma = F_\phi$, it is continuous because G_γ is strictly increasing. Thus, the following corollary to Theorem 1.3.1 holds by the probability integral transform.

Corollary 1.3.1 *Under the conditions of Theorem 1.3.1,* $H(\hat{p}_n) \to_d U_{[0,1]}$*.*

Therefore, the mapping of $\hat{p}_n$ into $H(\hat{p}_n)$ transforms $\hat{p}_n$ into a new p-value, $H(\hat{p}_n)$, whose asymptotic distribution is the standard uniform distribution on $[0, 1]$. Inference

[2] The same result follows in terms of weak convergence in distribution of $T_n^*|D_n$. Specifically, because $T_n^* = (T_n^* - \hat{B}_n) + (\hat{B}_n - B_n) + B_n$, where $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1^*$ and (jointly) $\hat{B}_n - B_n \xrightarrow{d} \xi_2$ with $\xi_1^* \sim \xi_1$ independent of ξ_2, we have that $T_n^*|D_n \xrightarrow{w} (B + \xi_1^* + \xi_2)|\xi_2$.

based on $H(\hat{p}_n)$ is generally infeasible, because we do not observe $H(u)$. However, if we can replace $H(u)$ with a uniformly consistent estimator $\hat{H}_n(u)$ then this approach will deliver a feasible modified p-value $\tilde{p}_n := \hat{H}_n(\hat{p}_n)$. Since the limit distribution of $\tilde{p}_n$ is the standard uniform distribution, $\tilde{p}_n$ is an asymptotically valid p-value. The mapping of $\hat{p}_n$ into $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$ by the estimated distribution of the former corresponds to what Beran (1987) calls 'prepivoting'. In the following sections, we describe two methods of obtaining a consistent estimator of $H(u)$.

Remark 1.3.4 *The prepivoting approach can also be used to solve the invalidity of confidence sets based on the standard bootstrap; see Remark 1.3.2. In particular, replace the nominal level* $1-\alpha$ *by* $\hat{H}_n^{-1}(1-\alpha)$ *and consider* $\widetilde{CI}_n^{1-\alpha} := [\hat{\theta}_n - g(n)^{-1}\hat{q}_n(\hat{H}_n^{-1}(1-\alpha)), +\infty)$. *Then*

$$P(\theta_0 \in \widetilde{CI}_n^{1-\alpha}) = P(\hat{p}_n \leq \hat{H}_n^{-1}(1-\alpha)) = P(\hat{H}_n(\hat{p}_n) \leq 1-\alpha) \to 1-\alpha,$$

where the last convergence is implied by Corollary 1.3.1 and consistency of $\hat{H}_n$.

Remark 1.3.5 *Corollary 1.3.1 can also be applied to right-tailed or two-tailed tests. The right-tailed p-value, say* $\hat{p}_{n,r} := P^*(T_n^* > T_n) = 1 - \hat{L}_n(T_n) = 1 - \hat{p}_n$, *has cdf* $P(\hat{p}_{n,r} \leq u) = P(\hat{p}_n \geq 1-u) = 1 - P(\hat{p}_n < 1-u) = 1 - H(1-u) + o(1)$ *uniformly in* u. *Note that, because the conditional cdf of* T_n^* *is continuous in the limit, the p-value* $\hat{p}_{n,r}$ *is asymptotically equivalent to* $P^*(T_n^* \geq T_n)$. *Thus, by Corollary 1.3.1, the modified right-tailed p-value,* $\tilde{p}_{n,r} := 1 - \hat{H}_n(\hat{p}_{n,r})$, *satisfies*

$$\tilde{p}_{n,r} = 1 - H(1-\hat{p}_{n,r}) + o_p(1) = 1 - H(\hat{p}_n) + o_p(1) \xrightarrow{d} U_{[0,1]}.$$

Similarly, for two-tailed tests the equal-tailed bootstrap p-value, $\tilde{p}_{n,et} := 2\min\{\tilde{p}_n, \tilde{p}_{n,r}\} = 2\min\{\tilde{p}_n, 1-\tilde{p}_n\}$, *satisfies* $\tilde{p}_{n,et} \to_d U_{[0,1]}$ *by Corollary 1.3.1 and the continuous mapping theorem.*

Plug-in approach

In view of Theorem 1.3.1, a simple approach to estimating $H(u)$ is to use

$$\hat{H}_n(u) = H_{\hat{\phi}_n,\hat{\gamma}_n}(u),$$

where $\hat{\gamma}_n$ and $\hat{\phi}_n$ denote consistent estimators of γ and ϕ, respectively. This leads to a plug-in modified p-value defined as

$$\tilde{p}_n = H_{\hat{\phi}_n,\hat{\gamma}_n}(\hat{p}_n).$$

By consistency of $\hat{\gamma}_n$ and $\hat{\phi}_n$ and under the assumption that $H_{\phi,\gamma}$ is continuous in (ϕ,γ), it follows immediately that

$$\tilde{p}_n = H(\hat{p}_n) + o_p(1) \xrightarrow{d} F_\phi(G_\gamma^{-1}(G_\gamma(F_\phi^{-1}(U_{[0,1]})))) = U_{[0,1]}.$$

This result is summarized next.

Corollary 1.3.2 *Let Assumptions 1.3.1 and 1.3.1 hold, and suppose $H_{\phi,\gamma}(u)$ is continuous in (ϕ,γ) for every u. If $(\hat{\gamma}_n,\hat{\phi}_n)\rightarrow_p(\gamma,\phi)$ then $\tilde{p}_n = H_{\hat{\phi}_n,\hat{\gamma}_n}(\hat{p}_n)\rightarrow_d U_{[0,1]}$.*

The plug-in approach relies on a consistent estimator of the asymptotic distribution H, but does not require estimating the 'bias term' B_n. When estimating γ and ϕ is simple, this approach is attractive since it does not require any double resampling. Examples are given in Section 1.4. However, computation of γ and ϕ is case-specific and may be cumbersome in practice. An automatic approach is to use the bootstrap to estimate $H(u)$, as we describe next.

Double bootstrap

Following Beran (1987, 1988), we can estimate $H(u)$ with the bootstrap. That is, we let

$$\hat{H}_n(u) = P^*(\hat{p}_n^* \leq u),$$

where $\hat{p}_n^*$ is the bootstrap analogue of $\hat{p}_n$. Since $\hat{p}_n$ is itself a bootstrap p-value, computing $\hat{p}_n^*$ requires a double bootstrap. In particular, let D_n^{**} denote a further bootstrap sample of size n based on D_n^* and some additional bootstrap variates (defined jointly with D_n and D_n^* on a possibly extended probability space), and let T_n^{**} denote the bootstrap version of T_n^* computed on D_n^{**}. With this notation, the second-level bootstrap p-value is defined as

$$\hat{p}_n^* := P^{**}(T_n^{**} \leq T_n^*),$$

where P^{**} denotes the bootstrap probability measure conditional on D_n^* and D_n (making $\hat{p}_n^*$ a function of D_n^* and D_n). This leads to a double bootstrap modified p-value, as given by

$$\tilde{p}_n := \hat{H}_n(\hat{p}_n) = P^*(\hat{p}_n^* \leq \hat{p}_n).$$

In order to show that $\tilde{p}_n = \hat{H}_n(\hat{p}_n)\rightarrow_d U_{[0,1]}$, we add the following assumption.

Let ξ_1 and ξ_2 be as defined in Assumptions 1.3.1 and 1.3.1. For some (D_n^*, D_n)-measurable random variable $\hat{B}_n^*$, it holds that: (i) $T_n^{**} - \hat{B}_n^* \overset{d^{**}}{\rightarrow}_{p^*} \xi_1$, in probability, and (ii) $T_n^* - \hat{B}_n^* \overset{d^*}{\rightarrow}_p \xi_1 - \xi_2$.

Assumption 1.3.2 complements Assumptions 1.3.1 and 1.3.1 by imposing high-level conditions on the second-level bootstrap statistics. Specifically, Assumption 1.3.2(i) assumes that T_n^{**} has asymptotic distribution G_γ only after we subtract $\hat{B}_n^*$. This term is the second-level bootstrap analogue of $\hat{B}_n$. It depends only on the first-level bootstrap data D_n^* and is not random under P^{**}. The second part of Assumption 1.3.2 follows from Assumption 1.3.1 in the special case that $\hat{B}_n^* - \hat{B}_n = o_{p^*}(1)$, in probability; i.e., when $\xi_2 = 0$ a.s., implying $F_\phi = G_\gamma$. When $F_\phi \neq G_\gamma$, $\hat{B}_n^*$ is not a consistent estimator of $\hat{B}_n$. However, under Assumption 1.3.2,

$$T_n^* - \hat{B}_n^* = (T_n^* - \hat{B}_n) - (\hat{B}_n^* - \hat{B}_n) \overset{d^*}{\rightarrow}_p \xi_1 - \xi_2 = F_\phi^{-1}(U_{[0,1]})$$

implying that $T_n^* - \hat{B}_n^*$ mimics the distribution of $T_n - \hat{B}_n$. This suffices for proving the asymptotic validity of the double bootstrap modified p-value, $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$, as proved next.

THEOREM 1.3.2 *Under Assumptions 1.3.1, 1.3.1, and 1.3.2, it holds that* $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$.

PROOF. To prove this result, recall that $\hat{H}_n(u) = P^*(\hat{p}_n^* \leq u)$ and $P(\hat{p}_n \leq u) \to H(u) = F_\phi(G_\gamma^{-1}(u))$ uniformly in $u \in \mathbb{R}$, since H is a continuous distribution function by Assumptions 1.3.1 and 1.3.1. We have that

$$\begin{aligned}
\hat{p}_n^* &= P^{**}(T_n^{**} \leq T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \leq T_n^* - \hat{B}_n^*) \\
&= G_\gamma(T_n^* - \hat{B}_n^*) + o_{p^*}(1), \quad \text{by Assumption 1.3.2(i)}, \\
&= G_\gamma(F_\phi^{-1}(U_{[0,1]})) + o_{p^*}(1), \quad \text{by Assumption 1.3.2(ii)},
\end{aligned}$$

where $G_\gamma(F_\phi^{-1}(U_{[0,1]}))$ is a random variable whose distribution function is H. Hence,

$$\sup_{u \in \mathbb{R}} |\hat{H}_n(u) - H(u)| = o_p(1).$$

Since $H(\hat{p}_n) \to_d U_{[0,1]}$, we can conclude that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$. □

Theorem 1.3.2 shows that prepivoting the standard bootstrap p-value $\hat{p}_n$ by applying the mapping $\hat{H}_n$ transforms it into an asymptotically uniformly distributed random variable. This result holds under Assumptions 1.3.1, 1.3.1, and 1.3.2, independently of whether $G_\gamma = F_\phi$ or not. When $G_\gamma = F_\phi$ then $\hat{p}_n \to_d U_{[0,1]}$ (as implied by Theorem 1.3.1). In this case, the prepivoting approach is not necessary to obtain a first-order asymptotically valid test, although it might help further reducing the size distortion of the test. This corresponds to the setting of Beran (1987, 1988), where prepivoting was proposed as a way of reducing the level distortions of confidence intervals. When $G_\gamma \neq F_\phi$ then $\hat{p}_n$ is not asymptotically uniform and a standard bootstrap test based on $\hat{p}_n$ is asymptotically invalid, as shown in Theorem 1.3.1. In this case, prepivoting transforms an asymptotically invalid bootstrap p-value into one that is asymptotically valid. This setting was not considered by Beran (1987, 1988) and is new to our paper.

1.3.3 Power of tests

In this section we explicitly consider a testing situation. Suppose we are interested in testing $\mathsf{H}_0 : \theta = \bar{\theta}$ against $\mathsf{H}_1 : \theta < \bar{\theta}$. Specifically, defining $T_n(\theta) := g(n)(\hat{\theta}_n - \theta)$, we consider the test statistic $T_n(\bar{\theta})$. The corresponding bootstrap p-value is $\hat{p}_n(\bar{\theta})$ with $\hat{p}_n(\theta) := P^*(T_n^* \leq T_n(\theta))$. When the null hypothesis is true, i.e., when $\bar{\theta} = \theta_0$ with θ_0 denoting the true value, we find $T_n(\bar{\theta}) = T_n(\theta_0) = T_n$ and $\hat{p}_n(\bar{\theta}) = \hat{p}_n(\theta_0) = \hat{p}_n$, where T_n and $\hat{p}_n$ are as defined previously. If Assumptions 1.3.1 and 1.3.1 hold under the null, Theorem 1.3.1 and Corollary 1.3.1 imply that tests based on $H(\hat{p}_n(\bar{\theta}))$ have correct asymptotic size, where H continues to denote the asymptotic cdf of $\hat{p}_n$.

To analyze power, we consider $\theta_0 = \bar{\theta} + a_n$ for some deterministic sequence a_n. Then $a_n = 0$ under the null hypothesis, whereas $a_n = a < 0$ corresponds to a fixed alternative and $a_n = a/g(n)$ for $a < 0$ corresponds to a local alternative. Thus, we define $\pi_n := g(n)(\theta_0 - \bar{\theta}) = g(n)a_n$ so that $T_n(\bar{\theta}) = T_n + \pi_n$.

THEOREM 1.3.3 *Suppose Assumptions 1.3.1 and 1.3.1 hold. (i) If $\pi_n \to \pi$ then $H(\hat{p}_n(\bar{\theta})) \to_d F_\phi(F_\phi^{-1}(U_{[0,1]}) + \pi)$. (ii) If $\pi_n \to -\infty$ then $P(H(\hat{p}_n(\bar{\theta})) \leq \alpha) \to 1$ for any nominal level $\alpha > 0$.*

PROOF. As in the proof of Theorem 1.3.1 we have, by Assumption 1.3.1(i),

$$\hat{p}_n(\bar{\theta}) = P^*(T_n^* \leq T_n(\bar{\theta})) = P^*(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n + \pi_n) = G_\gamma(T_n - \hat{B}_n + \pi_n) + o_p(1).$$

If $\pi_n \to \pi$ then $\hat{p}_n(\bar{\theta}) \to_d G_\gamma(F_\phi^{-1}(U_{[0,1]}) + \pi)$ by Assumption 1.3.1(ii), so that

$$H(\hat{p}_n(\bar{\theta})) \xrightarrow{d} H(G_\gamma(F_\phi^{-1}(U_{[0,1]}) + \pi)) = F_\phi(F_\phi^{-1}(U_{[0,1]}) + \pi)$$

by definition of $H(u)$. If $\pi_n \to -\infty$ then $\hat{p}_n(\bar{\theta}) \to_p 0$ because $T_n - \hat{B}_n = O_p(1)$ by Assumption 1.3.1(ii), so that $H(\hat{p}_n(\bar{\theta})) \to_p H(0) = 0$ and $P(H(\hat{p}_n(\bar{\theta})) \leq \alpha) \to 1$ for any $\alpha > 0$. □

It follows from Theorem 1.3.3(ii) that a left-tailed test that rejects for small values of $H(\hat{p}_n(\bar{\theta}))$ is consistent. Furthermore, it follows from Theorem 1.3.3(i) that such a test has non-trivial asymptotic local power against $\pi < 0$. Specifically, the asymptotic local power against π is given by $P(H(\hat{p}_n(\bar{\theta})) \leq \alpha) \to F_\phi(F_\phi^{-1}(\alpha) - \pi)$. Interestingly, this only depends on F_ϕ and not on G_γ. As above, to implement the modified p-value, $H(\hat{p}_n(\bar{\theta}))$, in practice, we would need a (uniformly) consistent estimator of H, i.e., the asymptotic distribution of the bootstrap p-value when the null hypothesis is true. This could be either the plug-in or double bootstrap estimators, as discussed in Sections 1.3.2 and 1.3.2.

Note that Assumption 1.3.1 is still assumed to hold in Theorem 1.3.3. That is, the bootstrap statistic T_n^* is assumed to have the same asymptotic behavior under the null and under the alternative. This is commonly the case when the bootstrap algorithm does not impose the null hypothesis when generating the bootstrap data.

1.3.4 Bootstrap p-value based on $T_n - \hat{B}_n$

The double bootstrap modified p-value $\tilde{p}_n$ depends only on the statistic T_n and their bootstrap analogues T_n^* and T_n^{**}. It does not involve computing explicitly $\hat{B}_n$ or $\hat{B}_n^*$, but in some applications it can be computationally costly as it requires two levels of resampling. As it turns out, $\tilde{p}_n$ is asymptotically equivalent to a single-level bootstrap p-value that is based on bootstrapping the statistic $T_n - \hat{B}_n$, as we show next.

By definition, the double bootstrap modified p-value is given by $\tilde{p}_n := P^*(\hat{p}_n^* \leq \hat{p}_n)$, where

$$\hat{p}_n^* := P^{**}(T_n^{**} \leq T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \leq T_n^* - \hat{B}_n^*) = G_\gamma(T_n^* - \hat{B}_n^*) + o_{p^*}(1),$$

in probability, given Assumption 1.3.2. Similarly, under Assumptions 1.3.1 and 1.3.1,

$$\hat{p}_n := P^*(T_n^* \leq T_n) = P^*(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n) = G_\gamma(T_n - \hat{B}_n) + o_p(1).$$

It follows that

$$\begin{aligned}\tilde{p}_n &:= P^*(\hat{p}_n^* \leq \hat{p}_n) = P^*(G_\gamma(T_n^* - \hat{B}_n^*) \leq G_\gamma(T_n - \hat{B}_n)) + o_p(1) \\ &= P^*(T_n^* - \hat{B}_n^* \leq T_n - \hat{B}_n) + o_p(1)\end{aligned}$$

because G_γ is continuous. We summarize this result in the following corollary.

Corollary 1.3.3 *Under Assumptions 1.3.1, 1.3.1, and 1.3.2,* $\tilde{p}_n = P^*(T_n^* - \hat{B}_n^* \leq T_n - \hat{B}_n) + o_p(1)$.

Theorem 1.3.2 shows that $\tilde{p}_n \to_d U_{[0,1]}$ and hence is asymptotically valid. In view of this, Corollary 1.3.3 shows that removing $\hat{B}_n$ from T_n and computing a bootstrap p-value based on the new statistic, $T_n - \hat{B}_n$, also solves the invalidity problem of the standard bootstrap p-value, $\hat{p}_n = P^*(T_n^* \leq T_n)$. Note that we do not require $\xi_2 = 0$, i.e. $\hat{B}_n - B_n$ and $\hat{B}_n^* - \hat{B}_n$ do not need to converge to zero.

When $\hat{B}_n$ and $\hat{B}_n^*$ are easy to compute, e.g., when they are available analytically as functions of D_n and D_n^*, respectively, Corollary 1.3.3 is useful as it avoids implementing a double bootstrap. When this is not the case, i.e., when deriving $\hat{B}_n$ and $\hat{B}_n^*$ explicitly is cumbersome or impossible, we may be able to estimate $\hat{B}_n$ from the bootstrap and $\hat{B}_n^*$ from a double bootstrap. Corollary 1.3.3 then shows that the double bootstrap modified p-value $\tilde{p}_n$ is a convenient alternative since it depends only on T_n, T_n^*, and T_n^{**}. It is important to note that none of these approaches requires the consistency of $\hat{B}_n$ and $\hat{B}_n^*$.

1.3.5 A more general set of high-level conditions

We conclude this section by providing an alternative set of high-level conditions that cover bootstrap methods for which $T_n^* - \hat{B}_n$ has a different limiting distribution than $T_n - B_n$. This may happen, for example, for the pairs bootstrap; see Section 1.2.1 and Remark 1.3.6.

Assumption 1.3.1 holds with part (i) replaced by (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1$, where ζ_1 is centered at zero and the cdf $J_\gamma(u) = P(\zeta_1 \leq u)$ is continuous and strictly increasing over its support.

Under Assumption 1.3.5, $T_n^* - \hat{B}_n$ does not replicate the distribution of $T_n - B_n$. This is to be understood in the sense that there does not exist a P^*-measurable term $\hat{B}_n$ such that $T_n^* - \hat{B}_n$ has the same asymptotic distribution as $T_n - B_n$.

An important generalization provided by Assumption 1.3.5 compared with Assumption 1.3.1 is to allow for bootstrap methods where the 'centering term', say B_n^*, depends on the bootstrap data. That is, to allow cases where there is a random (with respect to P^*, i.e., depending on the bootstrap data) term B_n^* such that $T_n^* - B_n^* \xrightarrow{d^*}_p \xi_1$ and hence has the same asymptotic distribution as $T_n - B_n$. Clearly, this violates Assumption 1.3.1 unless $B_n^* - \hat{B}_n \xrightarrow{p^*}_p 0$ (as in the ridge regression in Section 1.2.2). However, letting ζ_1 be such that $B_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1 - \xi_1$, then Assumption 1.3.5 covers the former case.

Remark 1.3.6 *A leading example where $T_n^* - B_n^* \xrightarrow{d^*}_p \xi_1$ and hence has the same asymptotic distribution as $T_n - B_n$ is the pairs bootstrap as in Section 1.2.1 for the model averaging example. We study this case in more detail in Section 1.4.1.*

The asymptotic distribution of the bootstrap p-value under Assumption 1.3.5 is given in the following theorem. The proof is identical to that of Theorem 1.3.1, with G_γ replaced by J_γ, and hence omitted.

Theorem 1.3.4 *If Assumptions 1.3.1 and 1.3.5 hold then $\hat{p}_n \rightarrow_d J_\gamma(F_\phi^{-1}(U_{[0,1]}))$.*

Theorem 1.3.4 implies that now $P(\hat{p}_n \le u) \rightarrow P(J_\gamma(F_\phi^{-1}(U_{[0,1]})) \le u) = F_\phi(J_\gamma^{-1}(u)) =: H(u)$. Clearly, a plug-in approach to estimating this $H(u)$ based on G_γ as described in Section 1.3.2 would be invalid because $G_\gamma \neq J_\gamma$ in general. However, it follows straightforwardly by the same arguments as applied in Section 1.3.2 that a plug-in approach based on J_γ will deliver an asymptotically valid plug-in modified p-value.

To implement an asymptotically valid double bootstrap modified p-value we consider the following high-level condition.

Assumption 1.3.2 holds with part (i) replaced by (i) $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} \zeta_1$, in probability, where ζ_1 is defined in Assumption 1.3.5.

Under Assumption 1.3.5, the second-level bootstrap statistic, $T_n^{**} - \hat{B}_n^*$, replicates the distribution of the first-level statistic, $T_n^* - \hat{B}_n$. Thus, the second-level bootstrap p-value is

$$\hat{p}_n^* := P^{**}(T_n^{**} \le T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \le T_n^* - \hat{B}_n^*) = J_\gamma(T_n^* - \hat{B}_n^*) + o_{p^*}(1)$$
$$\xrightarrow{d^*}_p J_\gamma(\xi_1 - \xi_2) = J_\gamma(F_\phi^{-1}(U_{[0,1]}))$$

under Assumption 1.3.5. Hence, the second-level bootstrap p-value has the same asymptotic distribution as the original bootstrap p-value. It follows that the double bootstrap modified p-value, $\tilde{p}_n := \hat{H}_n(\hat{p}_n) = P^*(\hat{p}_n^* \le \hat{p}_n)$, is asymptotically valid, which is stated next. The proof is essentially identical to that of Theorem 1.3.2 and hence omitted.

Theorem 1.3.5 *Under Assumptions 1.3.1, 1.3.5, and 1.3.5, it holds that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \rightarrow_d U_{[0,1]}$.*

Remark 1.3.7 *Consider again the case with a random bootstrap centering term in Remark 1.3.6, where* $B_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1 - \xi_1$ *such that* $T_n^* - B_n^* \xrightarrow{d^*}_p \xi_1$*. Within this setup, we can consider double bootstrap methods such that, for a random (with respect to* P^{**}*) term* B_n^{**} *we have* $T_n^{**} - B_n^{**} \xrightarrow{d^{**}}_{p^*} \xi_1$*, in probability. Thus, the asymptotic distribution of the second-level bootstrap statistic mimics that of the first-level statistic. When* B_n^{**} *and* ζ_1 *are such that* $B_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} \zeta_1 - \xi_1$*, in probability, then Assumption 1.3.5 is satisfied. As in Remark 1.3.6 this setup allows us to cover the pairs bootstrap.*

1.4 Examples continued

In this section we revisit our three leading examples from Section 1.2, where we argued that standard boostrap inference is invalid due to the presence of bias. In this section we show how to apply our general theory in each example. Again, we refer to Appendix A.2 for detailed derivations.

1.4.1 Inference after model averaging

Fixed regressor bootstrap. Extending the arguments in Section 1.2.1, we obtain the following result.

Lemma 1.4.1 *Under regularity conditions stated in Appendix A.2.1, Assumptions 1.3.1 and 1.3.1 are satisfied with* $(\xi_1, \xi_2)' \sim N(0, V)$*, where* $V := (v_{ij}), i, j = 1, 2$*, is positive definite and continuous in* ω*,* σ^2*, and* $\Sigma_{WW} := \text{plim}\, S_{WW}$*.*

By Lemma 1.4.1, the conditions of Theorem 1.3.1 hold with $G_\gamma(u) = \Phi(u/v_{11})$ and $F_\phi(u) = \Phi(u/v_d)$, where $v_d^2 = v_{11} + v_{22} - 2v_{12} > 0$. Then Theorem 1.3.1 implies that the standard bootstrap p-value satisfies $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$ with $m^2 := v_d^2/v^2$. Because ω is known and σ^2, Σ_{WW} are easily estimated, a consistent estimator $\hat{m}_n \to_p m$ is available, and the plug-in approach in Corollary 1.3.2 can be implemented by considering the modified p-value, $\tilde{p}_n = \Phi(\hat{m}_n^{-1}\Phi^{-1}(\hat{p}_n))$. Inspection of the proofs shows that our modified bootstrap approach is asymptotically valid whether δ is fixed or local-to-zero. In the former case, B_n is $O_p(n^{1/2})$ rather than $O_p(1)$, implying that B_n diverges in probability and $\tilde{\beta}_n$ is not even consistent for β. Despite this, the modified bootstrap p-value is asymptotically valid.

Alternatively, we can implement the double bootstrap as in Section 1.3.2. Specifically, let

$$y^{**} = x\hat{\beta}_n^* + Z\hat{\delta}_n^* + \varepsilon^{**},$$

where $\varepsilon^{**}|\{D_n, D_n^*\} \sim N(0, \hat{\sigma}_n^{*2} I_n)$, $(\hat{\beta}_n^*, \hat{\delta}_n^{*\prime}, \hat{\sigma}_n^{*2})$ is the OLS estimator obtained from the full model estimated on the first-level bootstrap data, and $D_n^* = \{y^*, W\}$. The double bootstrap statistic is $T_n^{**} := n^{1/2}(\tilde{\beta}_n^{**} - \hat{\beta}_n^*)$, where $\tilde{\beta}_n^{**} := \sum_{m=1}^M \omega_m \tilde{\beta}_{m,n}^{**}$ with $\tilde{\beta}_{m,n}^{**} := S_{xx.Z_m}^{-1} S_{xy^{**}.Z_m}$ defined as the double bootstrap OLS estimator from the m^{th} model. The double bootstrap modified p-value is then $\tilde{p}_n = P^*(\hat{p}_n^* \le \hat{p}_n)$ with $\hat{p}_n^* = P^{**}(T_n^{**} \le T_n^*)$.

LEMMA 1.4.2 *Under the conditions of Lemma 1.4.1, Assumption 1.3.2 holds with* $\hat{B}_n^* := Q_n n^{1/2}\hat{\delta}_n^*$.

Lemma 1.4.2 shows that Assumption 1.3.2 is verified in this example. The asymptotic validity of the double bootstrap modified p-value now follows from Lemmas 1.4.1 and 1.4.2 and Theorem 1.3.2.

PAIRS BOOTSTRAP. For the pairs bootstrap we verify the high-level conditions in Section 1.3.5. To simplify the discussion we consider the case with scalar z_t in (1.2.1) and where we "average" over only one model ($M = 1$), which is the simplest model in which z_t is omitted from the regression. That is, we estimate β by regression of y on x, i.e., $\tilde{\beta}_n = S_{xx}^{-1}S_{xy}$. In this special case, $T_n - B_n \to_d N(0, v^2)$ with $v^2 = \sigma^2\Sigma_{xx}^{-1}$ and $B_n = S_{xx}^{-1}S_{xz}n^{1/2}\delta$.

LEMMA 1.4.3 *Under regularity conditions stated in Appendix A.2.1, it holds that* $T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v^2 + \kappa^2)$, *where* $\hat{B}_n := S_{xx}^{-1}S_{xz}n^{1/2}\hat{\delta}_n$ *and* $\kappa^2 := d_r(\delta)'\Sigma_r d_r(\delta)$ *with* $d_r(\delta) := \delta(\Sigma_{xx}^{-1}, -\Sigma_{xx}^{-2}\Sigma_{xz})'$.

Notice that, in contrast to the FRB, the asymptotic variance of T_n^* fails to replicate that of T_n because of the term $\kappa^2 > 0$. This implies that the methodology developed in Theorem 1.3.1 and its corollaries no longer applies. Instead we can apply the theory of Section 1.3.5. In particular, Lemma 1.4.3 shows that Assumption 1.3.5(i) holds in this case with $\zeta_1 \sim N(0, v^2 + \kappa^2)$. Lemma 1.4.3 also shows that $\hat{B}_n$ is the same for the pairs bootstrap and the FRB, such that Lemma 1.4.1 shows that Assumptions 1.3.1 and 1.3.1(ii) are verified. This implies that Theorem 1.3.4 holds for this example. Using similar arguments, it can be shown that Assumption 1.3.5 also holds for this example, which would imply that the double bootstrap p-values are asymptotically uniformly distributed.

Under local alternatives of the form $\beta_0 = \bar{\beta} + an^{-1/2}$, where $\bar{\beta}$ is the value under the null (Section 1.3.3), the asymptotic local power function for the modified p-value is given by $\Phi(\Phi^{-1}(\alpha) - a/v_d)$; see Theorem 1.3.3. It is not difficult to verify that this is the same power function as that obtained from a test based directly on $\hat{\beta}_n$ from the full model (1.2.1).

1.4.2 Ridge regression

To complete the example in Section 1.2.2, we can proceed as in the previous example.

LEMMA 1.4.4 *Under the null hypotheses and the regularity conditions stated in Appendix A.2.2, Assumptions 1.3.1 and 1.3.1 are satisfied with* $(\xi_1, \xi_2)' \sim N(0, V)$, *where* $V := (v_{ij}), i, j = 1, 2$, *is positive definite and continuous in* c_0, σ^2, *and* Σ_{xx}.

As in Section 1.4.1, Lemma 1.4.4 and Theorem 1.3.1 imply that the standard bootstrap p-value satisfies $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$, where we now have $m^2 = (g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g)^{-1}g'\Sigma_{xx}^{-1}g$. Note that this result holds irrespectively of θ being fixed or local to zero. Thus, the bootstrap is invalid unless $c_0 = 0$ which implies $m = 1$. For the plug-in method, a simple consistent estimator of m is given by $\hat{m}_n^2 := (g'\tilde{S}_{xx}^{-1}S_{xx}\tilde{S}_{xx}^{-1}g)^{-1}g'S_{xx}^{-1}g$, and inference based on the plug-in modified p-value $\tilde{p}_n = \Phi(\hat{m}_n^{-1}\Phi^{-1}(\hat{p}_n))$ is then asymptotically valid by Corollary 1.3.2.

To implement the double bootstrap method, we can draw the double bootstrap sample $\{y_t^{**}, x_t^{**}; t = 1, \ldots, n\}$ as i.i.d. from $\{y_t^*, x_t^*; t = 1, \ldots, n\}$. Accordingly, the second-level bootstrap ridge estimator is $\tilde{\theta}_n^{**} := \tilde{S}_{x^{**}x^{**}}^{-1}S_{x^{**}y^{**}}$ with associated test statistic $T_n^{**} := n^{1/2}g'(\tilde{\theta}_n^{**} - \hat{\theta}_n^*)$, which is centered at the first-level bootstrap OLS estimator, $\hat{\theta}_n^*$. It is straightforward to show that, without additional conditions, Assumption 1.3.2 holds.

LEMMA 1.4.5 *Under the conditions of Lemma 1.4.4, Assumption 1.3.2 holds with* $\hat{B}_n^* := -c_n n^{-1/2}g'\tilde{S}_{x^*x^*}^{-1}\hat{\theta}_n^*$.

Validity of the double bootstrap modified p-value $\tilde{p}_n = P^*(\hat{p}_n^* \le \hat{p}_n)$ now follows by application of Theorem 1.3.2.

1.4.3 NONPARAMETRIC REGRESSION

Again, we complete the example in Section 1.2.3 by proceeding as in the previous examples.

LEMMA 1.4.6 *Under regularity conditions stated in Appendix A.2.3, Assumptions 1.3.1 and 1.3.1 are satisfied with* $(\xi_1, \xi_2)' \sim N(0, V)$*, where* $V := (v_{ij}), i, j = 1, 2,$ *is positive definite and continuous in* σ^2 *and the kernel function.*

As before, Lemma 1.4.6 and Theorem 1.3.1 imply that the standard bootstrap p-value satisfies $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$, where we now have $m^2 := 4 + (\int K^2(u)du)^{-1}(\int(\int K(s-u)K(s)ds)^2du - 4\int K(u)\int K(u-s)K(s)dsdu)$. Thus, in this example, m need not be estimated because it is observed once K is chosen. Therefore, valid inference is feasible with the modified p-value $\tilde{p}_n = H(\hat{p}_n) = \Phi(m^{-1}\Phi^{-1}(\hat{p}_n))$; see Corollary 1.3.1.

We can also apply a double bootstrap modification. Let $y_t^{**} = \hat{\beta}_h^*(x_t) + \varepsilon_t^{**}$, $t = 1, \ldots, n$, where $\varepsilon_t^{**}|\{D_n, D_n^*\} \sim$ i.i.d.$N(0, \hat{\sigma}_n^{*2})$ with $D_n^* := \{y_t^*, t = 1, \ldots, n\}$ and $\hat{\sigma}_n^{*2}$ denoting the residual variance from the first-level bootstrap data. The double bootstrap analogue of T_n is $T_n^{**} := (nh)^{1/2}(\hat{\beta}_h^{**}(x) - \hat{\beta}_h^*(x))$, where $\hat{\beta}_h^{**}(x) := (nh)^{-1}\sum_{t=1}^n k_t y_t^{**}$. This can be decomposed as $T_n^{**} = \xi_{1,n}^{**} + \hat{B}_n^*$, where $\hat{B}_n^* := (nh)^{1/2}((nh)^{-1}\sum_{t=1}^n k_t\hat{\beta}_h^*(x_t) - \hat{\beta}_h^*(x))$. Unfortunately, although $\xi_{1,n}^{**}$ satisfies Assumption 1.3.2(i), $\hat{B}_n^*$ does not satisfy Assumption 1.3.2(ii). The reason is that $\hat{B}_n^* - \hat{B}_n = \xi_{2,n}^* + \hat{B}_{2,n} - \hat{B}_n$, where $\xi_{2,n}^*$ satisfies Assumption 1.3.2(ii), but $\hat{B}_{2,n} := (nh)^{-1}\sum_{t=1}^n k_t\hat{B}_n(x_t)$ is a smoothed version of $\hat{B}_n$ (evaluated at x_t) and although $\hat{B}_{2,n} - \hat{B}_n$ is mean zero it is not $o_p(1)$. However, $\hat{B}_{2,n} - \hat{B}_n$ is observed, so this is easily corrected by defining $\bar{T}_n^{**} := T_n^{**} - (\hat{B}_{2,n} - \hat{B}_n)$. Then we have the following result.

LEMMA 1.4.7 *Under the conditions of Lemma 1.4.6, Assumption 1.3.2 holds with T_n^{**} and $\hat{B}_n^*$ replaced by $\bar{T}_n^{**}$ and $\bar{B}_n^* := \hat{B}_n^* - (\hat{B}_{2,n} - \hat{B}_n)$, respectively.*

The validity of the double bootstrap modified p-value $\tilde{p}_n = P^*(\hat{p}_n^* \leq \hat{p}_n)$, where $\hat{p}_n^* := P^{**}(\bar{T}_n^{**} \leq T_n^*)$, follows from Lemma 1.4.7 and Theorem 1.3.2. This in turn implies that confidence intervals based on the double bootstrap are asymptotically valid; see also Remark 1.3.4. We note that Hall and Horowitz (2013) also proposed, without theory, a version of their calibration method based on the double bootstrap. Our double bootstrap-based method for confidence intervals corresponds to their steps 1–5, and where we need a correction they have instead a step 6 in which they average over a grid of x.

Finally, under local alternatives of the form $\beta_0(x) = \bar{\beta} + an^{-2/5}$, where $\bar{\beta}$ is the value under the null (Section 1.3.3), the asymptotic local power function for the modified p-value is given by $\Phi(\Phi^{-1}(\alpha) - a/v_d)$; see Theorem 1.3.3. Alternatively, we could consider a "bias-free" test based on undersmoothing; that is using a bandwidth h satisfying $nh^5 \to 0$ such that $B_n \to 0$ and inference can be based on quantiles of $\xi_1 \sim N(0, v_{11}^2)$. In contrast to our procedure, however, such a test has only trivial power against $\bar{\beta} + an^{-2/5}$ because $(nh)^{1/2}an^{-2/5} \to 0$.

1.5 Concluding remarks

In this paper, we have shown that in statistical problems involving bias terms that cannot be estimated, the bootstrap can be modified to provide asymptotically valid inference. Intuitively, the main idea is the following: in some important cases, the bootstrap can be used to 'debias' a statistic whose bias is non-negligible, but when doing so additional 'noise' is injected. This additional noise does not vanish because the bias cannot be consistently estimated, but it can be handled either by a 'plug-in' method or by an additional (i.e., double) bootstrap layer. Specifically, our solution is simple and involves (i) focusing on the bootstrap p-value; (ii) estimating its asymptotic distribution; (iii) mapping the original (invalid) p-value into a new (valid) p-value using the prepivoting approach. These steps are easy to implement in practice and we provide sufficient conditions for asymptotic validity of the associated tests and confidence intervals.

Our results can be generalized in several directions. For instance, there is a growing literature where inference on a parameter of interest is combined with some auxiliary information in the form of a bound on the bias of the estimator in question. These bounds appear, e.g., in Oster (2019) and Li and Müller (2021). It is of interest to investigate how our analysis can be extended in order to incorporate such bounds. Other possible extensions include non-ergodic problems, large-dimensional models, and multivariate estimators or statistics. All these extensions are left for future research.

Chapter 2

Improved Inference for Nonparametric Regression and Regression-Discontinuity Designs

(written with Giuseppe Cavaliere, Sílvia Gonçalves and Morten Ørregaard Nielsen)

2.1 Introduction

Nonparametric regression for the analysis of (possibly) non-linear economic data have a long tradition. This class of models has the appealing property of relaxing the assumption of linearity of the conditional expectation function of the dependent variable without the need of imposing any parametric structure on its functional form. One of the most important applications of nonparametric regression is the Regression-Discontinuity Design (RDD); a popular tool for the analysis of quasi-experimental phenomena. On the one hand, RDDs have proven to be a reliable method for applied researchers (see, e.g., Black, 1990; Angrist and Lavy, 1999; and Chay et al., 2005), but on the other hand, methodological challenges have raised the attention of theoretical research (see, e.g., Hahn et al., 2001; Imbens and Kalyanaraman, 2012; Calonico et al., 2014; and Imbens and Lemieux, 2008, for a detailed review).

One of the main methodological issues in estimating (possibly) non-linear conditional expectations is that the popular choice of the local polynomial estimator – despite being consistent – is asymptotically biased when implemented with a mean-squared-error-minimizing bandwidth, and this poses a crucial challenge for inference. One possibility to deal with such asymptotic bias is the use of undersmoothing bandwidths, for which the bias term is asymptotically negligible. However, this implies inefficiency of the local polynomial estimator and is in contrast with most bandwidth selectors, which typically tend to pick "large" bandwidths; see Calonico et al. (2014) for a detailed discussion. Another way to deal with the asymptotic bias is direct bias estimation, which generally involves local polynomial estimation of higher-order derivatives of the conditional expectation function. Direct bias estimation has generally proven to be outperformed by undersmoothing techniques when constructing confidence bands for kernel-based estimators (see Hall, 1992, 1993). However, recent important contributions by Calonico et al. (2014, 2018) has proven that proper studentizations of the test statistic to appropriately account for the variability of the bias estimator drastically improves the performance of "direct" bias correction.

The bootstrap is generally considered a useful tool for bias correction. However, invalidity of "standard" bootstrap methods for the estimation of smooth regression curves is a well-known issue when dealing with kernel-based estimators. Such invalidity is due to the fact that the bootstrap test statistic, say T_n^*, is not able to mimic the asymptotic bias of the local polynomial estimator when a "large" bandwidth is considered, resulting in an asymptotic distribution which is random in the limit. Other than undersmoothing, which makes the asymptotic bias negligible both for the asymptotic and the bootstrap statistic, the literature on nonparametric regression has explored various possibilities to remove randomness in the limit distribution of T_n^* in order to restore "standard" bootstrap validity. Härdle and Bowman (1988) show the validity of bootstrap confidence bands based on a version of T_n^* which is centered at a consistent estimator of the asymptotic bias B. Härdle and Marron (1991) propose a fixed-regressor bootstrap in which the conditional expectation of the bootstrap dependent variable is an oversmoothed version of the local polynomial estimator, guaranteeing consistency of the bootstrap bias to B and standard bootstrap validity. However, both these approaches require calibration of two different bandwidths and suffer from undercoverage in finite samples. An approach more related to ours is that considered by Hall and Horowitz (2013), which focuses on an asymptotic theory-based confidence interval and applies the bootstrap to correct its coverage probability. However, their approach is asymptotically conservative and only over a subset of the support of x that does not include boundary points.

We propose a novel bootstrap-based approach to obtain asymptotically valid (unbiased) inference in nonparametric regression and RDD, which does not involve neither undersmoothing nor direct bias estimation. Our method is based on the concept of prepivoting, originally proposed by Beran (1987, 1988) to deliver asymptotic refinements and recently considered by Cavaliere et al. (2024) in the context of asymptotically biased estimators. The idea of prepivoting is the following. Even if the bias of the bootstrap test statistic does not converge in probability to the asymptotic bias, thus implying invalidity of the bootstrap using "standard" arguments, the distribution of the bootstrap p-value often does not depend on the original bias, but only on some nuisance parameters for which consistent estimation is possible. In such cases, the distribution of the bootstrap p-value is not uniform, not even for large samples (thus motivating the need for "non-standard" bootstrap algorithms), but its cdf can be uniformly estimated; see Cavaliere et al. (2024). We show that valid two-sided confidence intervals (CIs) can therefore be obtained by replacing the nominal levels $1-\alpha/2$ and $\alpha/2$, $\alpha \in (0,1)$, in "standard" bootstrap CIs, with the inverse of such a uniformly consistent estimator evaluated at $1-\alpha/2$ and $\alpha/2$. Specifically, we present two bootstrap algorithms, which we label the local polynomial (LP) and fixed-local (FL) bootstraps, that both deliver CIs with asymptotically correct coverage through prepivoting. Of course, even though we restrict our attention to these two algorithms, the application of prepivoting is not exclusive to them.

In the context of estimation of the conditional expectation at a fixed point x of the dependent variable in a bivariate, cross-sectional dataset, the LP bootstrap is based on the commonly considered fixed-regressor wild bootstrap algorithm in which the conditional expectation (conditionally on the original data) of the bootstrap dependent variable is a different local polynomial estimator at each observation point. This or analogous bootstrap algorithms are widely considered in the statistics literature; see, for instance, in Härdle and Bowman (1988), Härdle and Marron (1991) and Hall and Horowitz (2013). We show that standard bootstrap validity does not hold in this setup when a "large" bandwidth is selected and propose prepivoting as a possible solution. Interestingly, we show that "standard" prepivoting (i.e., as presented in Cavaliere et al., 2024) is not sufficient to delivier valid confidence intervals when x is a boundary point. Indeed, we prove that in such cases the large-sample distribution of the bootstrap p-value still depends on the asymptotic bias. Therefore, we propose a "modified" prepivoting approach based on a simple modification of the bootstrap test statistic involving known sample quantities. This modification ensures that the large-sample distribution of the bootstrap p-value is not a function of the asymptotic bias. Crucially, the "modified" prepivoting approach is valid both for interior and boundary points.

The overall idea of the FL bootstrap is similar, but the bootstrap conditional expectation function is based on a single Taylor series approximation of the original conditional expectation at x, where the coefficients of the Taylor series are estimated via local polynomial estimation. If this local polynomial order is larger than that considered to derive the original test statistic, we show that the bootstrap bias is not consistent for the original bias, but it does allow the application of "standard" prepivoting without the need for any modification, both for interior and boundary points.

Our contribution to the literature is threefold. First, we show that bootstrap validity can be restored in the context of local polynomial estimation of regression curves without the need of undersmoothing or direct bias correction, via the use of prepivoting for the LP and FL bootstrap algorithms. Second, we compare the efficiency properties of the two bootstrap methods. Finally, we show that the FL bootstrap-based prepivoted CIs are asymptotically equivalent to those obtained via robust bias correction (RBC), the leading approach in the literature proposed by Calonico et al. (2014, 2018). By combining the second and the third contribution, we show that the LP bootstrap-based prepivoted CIs are asymptotically more efficient than those obtained through RBC and are about 20%

shorter.

The remainder of this paper is organized as follows. In Section 2, we describe the idea of prepivoting in nonparametric regression. In Section 3, we first present the estimators and review the asymptotic theory. We then present the LP and FL bootstrap algorithms and formalize the validity of their prepivoted CIs. We conclude the section by comparing their efficiency and relating them with the RBC approach. In Section 4 we show the applicability of our method to (sharp) RDD. Finally, in Section 5 we assess the performance of our methods in finite samples via the results of Monte Carlo simulations, and Section 6 concludes. All technical derivations are included in the Appendix.

Notation

Throughout this chapter, the notation $\sim$ indicates equality in distribution. For instance, $Z \sim N(0,1)$ means that Z is distributed as a standard normal random variable. We write '$x := y$' and '$y =: x$' to mean that x is defined by y. The standard Gaussian cumulative distribution function (cdf) is denoted by Φ; $U_{[0,1]}$ is the uniform distribution on $[0,1]$, and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. If F is a cdf, F^{-1} denotes the right-continuous generalized inverse, i.e., $F^{-1}(u) := \sup\{v \in \mathbb{R} : F(v) \leq u\}$, $u \in \mathbb{R}$. Unless specified otherwise, all limits are for $n \to \infty$. To define a matrix A we write $A := (a_{ij})$ meaning that a_{ij} is the (i,j)-th element of A, and if A is a variance matrix we use the convention that $a_{ii} = a_i^2$. If f_0 and f_1 are a left-continuous and a right-continuous function, respectively, we write $f_0(0-)$ for $\lim_{x\uparrow 0} f_0(x)$ and $f_1(0+)$ for $\lim_{x\downarrow 0} f_1(x)$.

For a bootstrap sequence, say Y_n^*, we use $Y_n^* \xrightarrow{p^*}_p 0$, or equivalently $Y_n^* \xrightarrow{p^*} 0$, in probability, to mean that, for any $\epsilon > 0$, $P^*(|Y_n^*| > \epsilon) \to_p 0$, where P^* denotes the probability measure conditional on the original data D_n. An equivalent notation is $Y_n^* = o_{p^*}(1)$ (where we omit the qualification "in probability" for brevity). Similarly, we use $Y_n^* \xrightarrow{d^*}_p \xi$, or equivalently $Y_n^* \xrightarrow{d^*} \xi$, in probability, to mean that, for all continuity points $u \in \mathbb{R}$ of the cdf of ξ, say $G(u) := P(\xi \leq u)$, it holds that $P^*(Y_n^* \leq u) - G(u) \to_p 0$.

2.2 Prepivoting in Nonparametric Regression

We consider the problem of inference on an unknown smooth function g at a fixed point x. In a standard nonparametric regression, $g(x)$ is defined as the conditional expectation $\mathbb{E}[y_i|x_i = x]$ for an observed bivariate random sample $D_n := \{(y_i, x_i) : i = 1, ..., n\}$. Suppose a consistent estimator $\hat{g}_n(x) = \hat{g}_n(x; h, D_n, K)$ – indexed by a bandwidth $h = h(n) > 0$ and a kernel function K – of $g(x)$ exists, a popular choice for $\hat{g}_n(x)$ being a local approximation of $g(x)$ to a polynomial of order p. Inference based on $\hat{g}_n(x)$ is typically challenging due to the presence of an asymptotic bias. For instance, letting $T_n := \sqrt{nh}(\hat{g}_n(x) - g(x))$, the standard confidence interval

$$CI_{us} := \left[\hat{g}_n(x) - (nh)^{-1/2} v_{1n} \Phi^{-1}(1 - \alpha/2), \hat{g}_n(x) - (nh)^{-1/2} v_{1n} \Phi^{-1}(\alpha/2)\right] \quad (2.2.1)$$

is such that $\mathbb{P}\left(g(x) \in CI_{us}\right) \to 1-\alpha$, for some $\alpha \in (0,1)$, if and only if the condition

$$v_{1n}^{-1} T_n \xrightarrow{d} N(0,1) \tag{2.2.2}$$

holds. However, it is typically the case that (2.2.2) is only satisfied under "undersmoothing" choices of the sequence of bandwidths h – which the label "us" in (2.2.1) refers to. Unfortunately, most bandwidth selectors tend to opt for choices of h which are larger than the undersmoothing bandwidths (see Calonico et al., 2014, for a detailed discussion on the issue), leading to

$$v_{1n}^{-1} T_n \xrightarrow{d} N(v_1^{-1}B, 1) \tag{2.2.3}$$

where v_1 is such that $v_{1n} = v_1 + o_p(1)$ and $B = B(x, g^{(p+1)}(x), K)$ is an asymptotic bias with $g^{(p+1)}$ denoting the $(p+1)$-th order derivative of g.

We propose valid confidence intervals based on the bootstrap. Bootstrap inference in the context of nonparametric regression is challenging, as the bias of the bootstrap estimator is typically not able to mimic the behavior of the "true" bias B, not even asymptotically; see, e.g., Härdle and Marron (1991). To see why, let $D_n^* := \{(y_i^*, x_i^*) : i = 1, ..., n\}$ be a bootstrap sample and $\hat{g}_n^*(x) = \hat{g}_n(x; h, D_n^*, K)$ be the associated bootstrap estimator. A natural candidate for a bootstrap confidence interval would then be:

$$CI_{b,us} := \left[\hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}(1-\alpha/2), \hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}(\alpha/2)\right] \tag{2.2.4}$$

where $\hat{L}_n(u) := \mathbb{P}^*\left(T_n^* \leq u\right)$ with $T_n^* := \sqrt{nh}\left(\hat{g}_n^*(x) - \hat{g}_n(x)\right)$. Similarly to CI_{us}, also $CI_{b,us}$ delivers asymptotically correct coverage when h converges to zero sufficiently fast, ensuring that

$$v_{1n}^{-1} T_n^* \xrightarrow{d^*}_p N(0,1) \tag{2.2.5}$$

so that the bootstrap is said to be valid through standard arguments. On the contrary, when a "large" bandwidth is selected, letting $\xi_{1n}^* = \sqrt{nh}\left(\hat{g}_n^*(x) - \mathbb{E}^*\left[\hat{g}_n^*(x)\right]\right)$ and $\hat{B}_n := \sqrt{nh}(\mathbb{E}^*[\hat{g}_n^*(x)] - \hat{g}_n(x))$, we have that $v_{1n}^{-1}\xi_{1n}^* \xrightarrow{d^*}_p N(0,1)$ but $\hat{B}_n \neq B + o_p(1)$. As shown in Section 3, $\hat{B}_n$ actually converges in distribution to a Gaussian random variable with variance $v_2^2 > 0$ and might not even be centered at B. Therefore, the distribution of T_n^* is random in the limit and the bootstrap cannot be justified through standard arguments, see Cavaliere and Georgiev (2020).

We here show that the bootstrap can be used to deliver asymptotically valid confidence intervals even when a "large" bandwidth is selected, without an explicit bias correction. Our approach is based on Beran's (1987, 1988) prepivoting idea, recently discussed in Cavaliere et al. (2024). We show conditions under which a simple change of the significance levels in (2.2.4) is sufficient to deliver confidence intervals with asymptotically correct coverage. Specifically we propose the prepivoted confidence intervals:

$$\widetilde{CI} := \left[\hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\hat{H}_n^{-1}(1-\alpha/2)\right), \hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\hat{H}_n^{-1}(\alpha/2)\right)\right] \tag{2.2.6}$$

such that the values $1-\alpha/2$ and $\alpha/2$ in (2.2.4) are replaced by $\hat{H}_n^{-1}(1-\alpha/2)$ and $\hat{H}_n^{-1}(\alpha/2)$, respectively, where $\hat{H}_n(u)$ is a uniformly consistent estimator of $H(u)$, i.e. the large-sample distribution function of the bootstrap p-value $\hat{p}_n$, where $\hat{p}_n := \mathbb{P}^*(T_n^* \leq T_n)$.

The intuition is the following. Even if the distributions of T_n and T_n^* depend on the value of the unknown bias term B, we find conditions under which H does not. Therefore, even if H is not uniform (condition which holds if the bootstrap is valid using "standard" arguments), it only depends on nuisance parameters which are relatively easy to estimate, with their estimation not requiring the calibration of additional tuning tools. Therefore,

$$\begin{aligned}\mathbb{P}\left(g(x) \in \widetilde{CI}\right) &= \mathbb{P}\left(\hat{L}_n^{-1}\left(\hat{H}_n^{-1}(\alpha/2)\right) \leq T_n \leq \hat{L}_n^{-1}\left(\hat{H}_n^{-1}(1-\alpha/2)\right)\right)\\ &= \mathbb{P}\left(\hat{H}_n^{-1}(\alpha/2) \leq \hat{p}_n \leq \hat{H}_n^{-1}(1-\alpha/2)\right)\\ &= \mathbb{P}\left(\alpha/2 \leq \hat{H}_n(\hat{p}_n) \leq 1-\alpha/2\right) \to 1-\alpha\end{aligned}$$

where the convergence is given by the fact that uniform consistency of $\bar{H}(u)$ to $H(u)$ implies $\hat{H}_n(\hat{p}_n) \xrightarrow{d} U_{[0,1]}$; see Cavaliere et al. (2024).

In the setup of nonparametric curve estimation, we find that a crucial condition for H not to depend on B is that the large sample distribution of $\hat{B}_n$ is centered at B. We find that for some bootstrap DGPs and test statistics, this is not always the case and show proper modifications of $\hat{L}_n$ which allow such condition to be satisfied. In this regards, notice that prepivoting does not restrict to a single specifications of D_n^* and $\hat{L}_n$. In Section 3 we implement prepivoting through two different procedures, namely the LP and FL bootstraps, which indeed imply different specifications of D_n^* and $\hat{L}_n$, and the applicability of prepivoting to alternative bootstrap procedures is left for future research.

2.3 Main Results

In this section we show the main results of this paper. Specifically, in Section 3.1 we introduce the considered DGP, the main assumptions and the estimator. In Section 3.2 and 3.3 we implement our prepivoted confidence intervals via two different bootstrap methodology, the LP and FL bootstrap, respectively. In Section 3.4 we analyze the efficiency properties of the prepivoted confidence intervals.

2.3.1 Review of asymptotic theory

Let $D_n := \{(y_i, x_i) : i = 1, ..., n\}$ be a random sample from the model

$$y_i = g(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $E(\varepsilon_i|x_i) = 0$, $\mathbb{V}(\varepsilon_i|x_i) =: \sigma^2(x_i)$, x_i is a random variable with bounded support $\mathbb{S}_x := [a,b], (a,b) \in \mathbb{R}^2$, and pdf $f(x)$ such that $f : \mathbb{S}_x \to (0,+\infty)$, while g is a smooth function such that $g : \mathbb{S}_x \to \mathbb{R}$. We consider local polynomial estimation of g at a fixed point x. For the seek of simplicity of exposition, we here make the normalization $[a,b] = [0,1]$ and restrict to the most popular case of the local linear estimator, given by

$$\hat{g}_n(x) = e_1' (Z_x' W_x Z_x)^{-1} Z_x' W_x y,$$

where $e_1' := (1,0)$, $y := (y_1, ..., y_n)'$, $Z_{1x} := (Z_{1x1}, ..., Z_{1xn})'$, $Z_{1xi} := (1, (x_i - x)/h))'$ and $W(x) := \text{diag}(h^{-1}K((x_1 - x)/h), ..., h^{-1}K((x_n - x)/h))$. Letting

$$w_i(x) := e_1' \left(\frac{Z_{1x}' W_x Z_{1x}}{n} \right)^{-1} Z_{1xi} K\left(\frac{x_i - x}{h} \right), \tag{2.3.1}$$

we can rewrite $\hat{g}_n(x)$ as

$$\hat{g}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} w_i(x) y_i,$$

We now focus on the asymptotic behavior of $\hat{g}_n(x)$ when properly centered and scaled. We provide results for both interior points and points on the boundary of the support of x. Although these results are well-known in the literature, they are useful for deriving our bootstrap results and hence we summarize them here.

We make the following assumptions.

Assumption 2.3.1 *(i) (y_i, x_i) are i.i.d. such that $E(\varepsilon_i^4|x_i = x) < +\infty$; (ii) $g : \mathbb{S}_x \to \mathbb{R}$ is three times continuously differentiable, and (iii) $\sigma^2(x) := V(y_i|x_i = x)$ is continuous and bounded away from zero.*

Assumption 2.3.2 *The function $K : \mathbb{R} \to [0,+\infty)$ is a symmetric, continuous and bounded function on $(-1,1)$ which equals zero outside the interval $[-1,1]$. In addition, we assume that K is a second-order kernel function such that $\int_{-1}^{1} K(u)\, du = 1$.*

Assumption 2.3.3 *The bandwidth $h = h(n)$ is such that $h \to 0$ as $n \to \infty$ and $nh^5 \to \kappa$ for some $\kappa \in [0,+\infty)$.*

Let $T_n := \sqrt{nh}(\hat{g}_n(x) - g(x))$, note that we can decompose T_n into a "bias" and a "variance" component,

$$T_n = B_n + \xi_{1n},$$

where

$$B_n = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x)[g(x_i) - g(x)] \qquad \text{and} \qquad \xi_{1n} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) \varepsilon_i.$$

The variance component ξ_{1n} drives the asymptotic Gaussianity of T_n, whereas B_n is a bias term that shifts this asymptotic distribution away from zero. Let $\mathcal{X}_n := (x_1, ..., x_n)'$ and $v_{1n}^2 := \mathbb{V}(\xi_{1n}|\mathcal{X}_n)$, then the following proposition holds.

Proposition 2.3.1 *Let Assumptions 1-3 hold, then:*

$$v_{1n}^{-1}\xi_{1n} \xrightarrow{d} N(0,1).$$

2.3.2 LP bootstrap

Consider a fixed-regressor wild bootstrap DGP of the form:

$$y_i^* = \hat{g}_n(x_i) + \varepsilon_i^* \tag{2.3.2}$$

where $\varepsilon_i^* := \hat{\varepsilon}_i e_i^*$, such that $\hat{\varepsilon}_i$ are the leave-one-out residuals $\hat{\varepsilon}_i := y_i - \hat{g}_{n,-i}(x_i)$, and e_i^* is a iid random variable, conditionally on the original data, satisfying $\mathbb{E}^*[e_i^*] = 0$, $\mathbb{E}^*[e_i^{*2}] = 1$ and $\mathbb{E}^*[e_i^{*4}] < \infty$. As for the asymptotic test statistic, we focus to the case in which the bootstrap conditional expectation function is based on a local linear estimator. Note that fixed-regressor bootstrap DGPs of this or similar forms have been widely adapted to the problem of bootstrapping a kernel-based estimator in nonparametric regression; see, e.g., Härdle and Marron (1988), Härdle and Bowman (1991) and Hall and Horowitz (2013).

The local linear bootstrap estimator is then:

$$\hat{g}_n^*(x) := e_1'(Z_{1x}'W_xZ_{1x})^{-1}(Z_{1x}'W_xy^*)$$

where $y^* := (y_1^*, ..., y_n^*)'$; moreover, we let $T_n^* := \sqrt{nh}\,(\hat{g}_n^*(x) - \hat{g}_n(x))$. It is well known that standard bootstrap validity does not generally apply to this setup as T_n^* does not mimic the asymptotic bias of T_n, making the confidence intervals (2.2.4) invalid unless $\kappa = 0$. Indeed, by letting $\hat{g}_n := (\hat{g}_n(x_1), ..., \hat{g}_n(x_n))'$, we have that the bootstrap bias

$$\hat{B}_n := \sqrt{nh}\,(\mathbb{E}^*\left[\hat{g}_n^*(x)\right] - \hat{g}_n(x)) = \sqrt{nh}\left[e_1'(Z_x'W_xZ_x)^{-1}(Z_x'W_x\hat{g}_n) - \hat{g}_n(x)\right]$$

is such that $T_n^* - \hat{B}_n$ is asymptotically Gaussian and centered at zero, with asymptotic variance equal to the asymptotic variance of ξ_{1n}, but $\hat{B}_n - B_n \neq o_p(1)$. We formalize the first result in the following proposition.

Proposition 2.3.2 *Let Assumptions 1-3 hold, then,*

$$v_{1,n}^{-1}\xi_{1n}^* := v_{1,n}^{-1}(T_n^* - \hat{B}_n) \xrightarrow{d^*}_p N(0,1).$$

In order to analyze the asymptotic behavior of $\hat{B}_n - B_n$, we note that also $\hat{B}_n$ can be split into a "bias" and "variance" component. Specifically, we write:

$$\begin{aligned}\hat{B}_n &= \frac{1}{\sqrt{nh}}\sum_{i=1}^n w_i(x)\left(\frac{1}{nh}\sum_{j=1}^n w_j(x_i)g(x_j) - \frac{1}{nh}\sum_{i=1}^n w_i(x)g(x_i)\right)\\ &\quad + \frac{1}{\sqrt{nh}}\sum_{i=1}^n w_i(x)\left(\frac{1}{nh}\sum_{j=1}^n w_j(x_i)\varepsilon_j - \frac{1}{nh}\sum_{i=1}^n w_i(x)\varepsilon_i\right)\\ &=: B_{2n} + \xi_{2n}\end{aligned}$$

B_{2n} is a stochastic term driving the expectation of $\hat{B}_n$, whereas ξ_{2n} is an asymptotically Gaussian random variable centered at zero. Intuitively, if B_{2n} converged in probability to $B := \text{plim}_{n\to\infty} B_n$ and ξ_{2n} was asymptotically negligible, then standard bootstrap validity would apply and the confidence intervals in (2.2.4) would deliver asymptotically correct coverage. In this specific setup, we note that both such conditions can be violated, justifying the need for alternative implementations of the bootstrap. We start by considering the behavior of ξ_{2n}. Let $\xi_n := (\xi_{1n}, \xi_{2n})'$ and $V_{LP,n} := \mathbb{V}(\xi_n|\mathcal{X}_n) = (v_{ij,n})$, then the following proposition holds.

Proposition 2.3.3 *Let Assumptions 1-3 hold, then: (i)*

$$V_{LP,n}^{-1/2}\xi_n \xrightarrow{d} N(0, I_2);$$

(ii) moreover, if x is an interior point,

$$V_{LP,n} \xrightarrow{p} V_{LP};$$

whereas if x is a boundary point,

$$V_{LP,n} \xrightarrow{p} \ddot{V}_{LP}$$

where $V_{LP} := (v_{ij,LP})$ and $\ddot{V}_{LP} := (\ddot{v}_{ij,LP})$, with $v_{2,LP}, \ddot{v}_{2,LP} > 0$, are defined in Appendix B.

Remark 2.3.1 *Proposition 2.3.3 shows a joint convergence in distribution argument for ξ_n, making a distinction between interior and boundary points (we here focus on left-boundary points for simplicity of exposition, though the analysis for right-boundary points is analogous). In Proposition 2.3.3, as well as in the results below, we refer to boundary points as left-boundary points, i.e. with $x = 0$, for simplicity of exposition, though the conclusions are equivalent for the case $x = 1$.*

Note that, even if the limit of $V_{LP,n}$ changes depending on the location of x, there exist estimators such that they adaptively converge in probability to V_{LP} when x is an interior point and to $\ddot{V}_{LP}$ when x is a boundary point. Such estimators are typically based on a feasible version of $V_{LP,n}$, which replaces the unknown quantity $\sigma^2(x)$ by some functions of the estimated residuals; see, e.g., Calonico et al. (2018) and Bartalotti (2019). For instance, let $\hat{V}_{LP,n} := (\hat{v}_{ij,LP,n})$, where

$$\hat{V}_{LP,n} := \frac{1}{nh}\sum_{i=1}^{n}\begin{pmatrix} w_i(x) & w_i(x)\tilde{w}_i(x) \\ w_i(x)\tilde{w}_i(x) & \tilde{w}_i^2(x)\end{pmatrix}\hat{\varepsilon}_i^2, \tag{2.3.3}$$

and $\tilde{w}(x_i) := (nh)^{-1}\sum_{j=1}^{n}(w_j(x)w_i(x_j) - w_i(x))$. Then, $\hat{V}_{LP,n} - V_{LP,n} = o_p(1)$.

We now consider B_{2n} and show that it may not converge in probability to the same limit as B_n. To motivate this statement, note that a standard result in nonparametric regression states that

$$B_n = B_{AT,n} + o_p(1), \qquad B_{AT,n} := \sqrt{nh^5}\frac{g''(x)}{2}C_n \tag{2.3.4}$$

where $C_n = C_n(x) := (nh)^{-1}\sum_{i=1}^n w_i(x)((x_i - x)/h)^2$. A similar expansion can be made for B_{2n}, for which we note that:

$$B_{2n} = B_{LP,n} + o_p(1), \qquad B_{LP,n} := \sqrt{nh^5}\frac{g''(x)}{2}C_{2n} \tag{2.3.5}$$

where $C_{2n} = C_{2n}(x) := (nh)^{-1}\sum_{i=1}^n w_i(x)C_n(x_i)$. Hence, the limit of $B_{2n} - B_n$ is driven by the limit of $C_{2n} - C_n$. Crucially, we find that such limit depends on the distance of x to the boundaries of $\mathbb{S}_x$, as formalized by the following proposition.

PROPOSITION 2.3.4 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$C_{2n} - C_n = o_p(1) \quad \Rightarrow \quad B_{2n} - B_n = o_p(1)$$

(ii) if x is a boundary point,

$$C_{2n} - C_n \neq o_p(1) \quad \Rightarrow \quad B_{2n} - B_n = A + o_p(1)$$

where $A := \sqrt{\kappa}\frac{g''(0+)}{2}(C_2 - C)$ such that $C_2 := plim_{n\to\infty}C_{2n}$ and $C := plim_{n\to\infty}C_n$ are defined in Appendix B.

Table 2.1: Limits of C_n, C_{2n} and C_n/C_{2n} for different choices of K.

	Interior			Boundary		
K	C	C_2	C/C_2	C	C_2	C/C_2
Triangular	0.1667	0.1667	1.0000	-0.1000	-0.0710	1.4082
Uniform	0.3333	0.3333	1.0000	-0.1667	-0.1389	1.2000
Epanechnikov	0.2000	0.2000	1.0000	-0.1158	-0.0853	1.3571
Biweight	0.1429	0.1429	1.0000	-0.0886	-0.0624	1.4211
Triweight	0.1111	0.1111	1.0000	-0.0718	-0.0493	1.4551

Propositions 2.3.3 and 2.3.4 show the two limiting sources of the invalidity of the "standard" confidence intervals $CI_{b,us}$. Standard invalidity of $CI_{b,us}$ can be view through the lenses of the distribution of the bootstrap p-value $\hat{p}_n := \mathbb{P}(T_n^* \leq T_n)$. Indeed, to allow $CI_{b,us}$ to be valid, the bootstrap p-value should be uniformly distributed:

$$\mathbb{P}(g(x) \in CI_{b,us}) = \mathbb{P}(\alpha/2 \leq \hat{p}_n \leq 1 - \alpha/2) \to 1 - \alpha, \quad \forall \alpha \in (0,1) \quad \Leftrightarrow \quad \hat{p}_n \xrightarrow{d} U_{[0,1]}$$

However, this is not true due to the results in Propositions 2.3.3 and 2.3.4. The limit distribution of $\hat{p}_n$ is derived as follows.

Proposition 2.3.5 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$\hat{p}_n \xrightarrow{d} \Phi\left(m_{LP}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{2.3.6}$$

where $m_{LP} = \sqrt{v_{1,LP}^2 + v_{2,LP}^2 - 2v_{12,LP}}/v_{1,LP}$; and (ii) if x is a boundary point,

$$\hat{p}_n \xrightarrow{d} \Phi\left(\ddot{a}_{LP} + \ddot{m}_{LP}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{2.3.7}$$

where $\ddot{a}_{LP} = A/\ddot{v}_{1,LP}$ and $\ddot{m}_{LP} := \ddot{v}_{d,LP}/\ddot{v}_{1,LP} := \sqrt{\ddot{v}_{1,LP}^2 + \ddot{v}_{2,LP}^2 - 2\ddot{v}_{12,LP}}/\ddot{v}_{1,LP}$.

Proposition 2.3.5 shows that the bootstrap p-value would be uniformly distributed - both for interior and boundary points - if and only if: (1) $m_{LP} = \ddot{m}_{LP} = 1$; and (2) $\ddot{a}_{LP} = 0$. We can see from Proposition 2.3.3 that (1) is violated because $v_{2,LP}, \ddot{v}_{2,LP} > 0$, i.e., the bootstrap bias does not have a probability limit; moreover, (2) is violated because the convolution term C_{2n} entering the definition of $B_{LP,n}$ implies that $\ddot{a}_{LP} \neq 0$.

We here propose prepivoting as a way to restore bootstrap validity. Specifically, we show that our prepivoted confidence intervals (2.2.6) are able to provide asymptotically correct coverage without the need to directly estimate B and despite the invalidity sources arising from Proposition 2.3.3 and 2.3.4. Additionally, the procedure does not require additional tuning parameters. As depicted in Section 2, our approach is based on the inversion of a uniformly consistent estimator the cdf of $\hat{p}_n$. We see that "standard" prepivoting – i.e., as considered in Cavaliere et al. (2024) – can restore validity of the bootstrap when (1) is not satisfied, but is not sufficient if invalidity arises from the violation of condition (2). Therefore, it can only be applied for interior points in the sense of Remark 2.3.1. However, as we will show below, a "modified" prepivoting approach can be applied to restore validity without ex-ante knowledge about the location of x relatively to the boundaries of its support.

We first consider the case in which x is an interior point. Proposition 2.3.5 implies that

$$\mathbb{P}\left(\hat{p}_n \leq u\right) \to \mathbb{P}\left(\Phi\left(m_{LP}\Phi^{-1}\left(U_{[0,1]}\right)\right) \leq u\right) = \mathbb{P}\left(U_{[0,1]} \leq \Phi\left(m_{LP}^{-1}\Phi^{-1}\left(u\right)\right)\right) = \Phi\left(m_{LP}^{-1}\Phi^{-1}\left(u\right)\right) =: H(u)$$

Therefore, even if the distribution of $\hat{p}_n$ is not uniform because $m_{LP} \neq 1$, uniformity can be retrieved by applying its cdf transform, i.e.:

$$H(\hat{p}_n) \xrightarrow{d} U_{[0,1]}$$

As depicted in Proposition 2.3.3, H does not depend on the value of B, but only on nuisance parameter for which consistent estimation is possible and does not involve calibration of additional tuning parameters; see (2.3.3). Hence, letting $\hat{m}_n := (\hat{v}_{1,LP,n}^2 + \hat{v}_{2,LP,n}^2 - 2\hat{v}_{12,LP,n})^{1/2}/\hat{v}_{1,LP,n}$, a uniformly consistent estimator of H is

$$\hat{H}_{LP,n}(u) := \Phi\left(\hat{m}_{LP,n}^{-1}\Phi^{-1}\left(u\right)\right) \tag{2.3.8}$$

Valid confidence intervals can thus be based on $\hat{H}_n$, as stated in the following theorem.

Theorem 2.3.1 *Let Assumptions 1-3 hold and x be an interior point, then*

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LP}\right) \to 1-\alpha, \quad \alpha \in (0,1)$$

where $\widetilde{CI}_{LP}$ is the prepivoted confidence interval in (2.2.6) with $\hat{H}_n = \hat{H}_{LP,n}$ and $\hat{L}_n$ the probability distribution (conditional on the data) of the LP bootstrap statistic T_n^*.

We now move to the case in which x is a boundary point. In this scenario, "standard" prepivoting is not able to restore bootstrap validity as it cannot correct the source of invalidity arising from the presence of $\ddot{a}_{LP}$. In the following, we show how a "modified" prepivoting approach, based on a simple modification of T_n^*, is able to provide asymptotically correct confidence intervals. Crucially, the resulting confidence intervals are valid both for interior and boundary points without ex-ante knowledge about the relative distance of x to the boundaries of its support.

To see how, we note that

$$\ddot{a}_{LP} := \frac{A}{\ddot{v}_{1,LP}} = \frac{\text{plim}_{n\to\infty}(B_{LP,n} - B_{AT,n})}{\ddot{v}_{1,LP}} = \frac{\sqrt{\kappa} g''(0+)(C_2 - C)}{2\ddot{v}_{1,LP}}$$

Clearly, the fact that $\ddot{a}_{LP}$ depends on g'' implies that also the cdf of $\hat{p}_n$ will depend on g'', thus preventing "standard" prepivoting to avoid direct estimation of B to obtain asymptotically valid confidence intervals. However, we note that, $\forall x \in \mathbb{S}_x$:

$$\frac{B_{AT,n}}{B_{LP,n}} = \frac{C_n}{C_{2n}} =: Q_n \tag{2.3.9}$$

where $Q_n = Q_n(x)$ is an observed quantity only depending on the observed K, h and $\mathcal{X}_n$. Moreover,

$$Q_n = Q + o_p(1) \tag{2.3.10}$$

where $Q = C/C_2 = 1$ if x is an interior point $Q \neq 1$ if x is a boundary point. Since Q_n is observed, we can think of a modified bootstrap statistic being $T_{mod,n}^* := Q_n T_n^*$. Clearly, the decomposition of the bootstrap test statistic between a "bias" and a "variance" component can also be applied to such modified bootstrap statistic, so that:

$$T_{mod,n}^* = \hat{B}_{mod,n} + \xi_{1,mod,n}$$

where $\hat{B}_{mod,n} := Q_n \hat{B}_n$ and $\xi_{1,mod,n} = Q_n \xi_{1n}$. We note that $\xi_{1,mod,n}$ preserves the property of being an asymptotically Gaussian random variable centered at zero, whereas $\hat{B}_{mod,n}$ drives the bias of the modified bootstrap statistic. Crucially, by (2.3.4) and (2.3.5) we have that:

$$\hat{B}_{mod,n} - B_n = Q_n B_{LP,n} - B_{AT,n} + Q_n \xi_{2n} + o_p(1) =: \xi_{2,mod,n} + o_p(1)$$

where this result is valid both for interior and boundary points. The asymptotic properties of $T_{mod,n}^*$ are summarized in the following proposition.

Proposition 2.3.6 *Let the conditions in Proposition 2.3.1, then* $\forall x \in \mathbb{S}_x$*: (i)*

$$(v_{1n}Q_n)^{-1}(T^*_{mod,n} - \hat{B}_{mod,n}) \xrightarrow{d^*}_p N(0,1); \tag{2.3.11}$$

and (ii)

$$\hat{B}_{mod,n} - B_n = \xi_{2,mod,n} + o_p(1); \tag{2.3.12}$$

(iii) moreover, by letting $\xi_{mod,n} := (\xi_{1n}, \xi_{2,mod,n})$

$$V^{-1/2}_{LP,mod,n}\xi_{mod,n} \xrightarrow{d} N(0, I_2) \tag{2.3.13}$$

where $V_{LP,mod,n} := \mathbb{V}[\xi_{mod,n}|\mathcal{X}_n]$*.*

The first part of Proposition 2.3.6 shows that the modified bootstrap statistic is asymptotically a standard normal when properly studentized and centered; the result follows directly from Proposition 2.3.2. The second part of the proposition formalizes the fact that the bootstrap bias is asymptotically centered at the limit of B_n when the proposed modification is applied, no matter the location of x relatively to the boundaries of its support. Finally, the third part shows that the joint convergence argument of ξ_{1n} and the "variance" component of the bootstrap bias is preserved after the modification.

Intuitively, the asymptotic covariance matrix of $\xi_{mod,n}$ is affected by the presence of Q_n. However, if x is an interior point, $Q_n = 1 + o_p(1)$ implies that $V_{LP,mod,n} = V_{LP} + o_p(1)$. If, instead, x is a boundary point:

$$V_{LP,mod,n} = \ddot{V}_{LP,mod} + o_p(1); \qquad \ddot{V}_{LP,mod} = (\ddot{v}_{ij,LP,mod}) = \text{diag}(1,Q) \cdot \ddot{V}_{LP} \cdot \text{diag}(1,Q) \tag{2.3.14}$$

Therefore, an adaptive estimator of the limit of $V_{n,mod}$ takes the form:

$$\hat{V}_{LP,mod,n} := \text{diag}(1,Q_n) \cdot \hat{V}_{LP,n} \cdot \text{diag}(1,Q_n) \tag{2.3.15}$$

where $\hat{V}_{LP,n}$ is defined in (2.3.3). Then, the consistency result $\hat{V}_{LP,mod,n} - V_{LP,mod,n} = o_p(1)$ follows directly from the fact that $\hat{V}_{LP,n} - V_{LP,n} = o_p(1)$.

By Proposition 2.3.6, one can intuitively obtain valid confidence intervals by applying "standard" prepivoting to the modified statistic $T^*_{mod,n}$. To this purpose, let $\hat{p}_{mod,n} := \mathbb{P}^*(T^*_{mod,n} \leq T_n)$, then the following proposition holds.

Proposition 2.3.7 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$\hat{p}_{mod,n} \xrightarrow{d} \Phi\left(m_{LP}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{2.3.16}$$

where m_{LP} is defined in Proposition 2.3.5; and (ii) if x is a boundary point,

$$\hat{p}_{mod,n} \xrightarrow{d} \Phi\left(\ddot{m}_{LP,mod}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{2.3.17}$$

where $\ddot{m}_{LP,mod} := \ddot{v}_{d,LP,mod}/\ddot{v}_{1,LP,mod} := \sqrt{\ddot{v}^2_{1,LP,mod} + \ddot{v}^2_{2,LP,mod} - 2\ddot{v}_{12,LP,mod}}/Q\ddot{v}_{1,LP}$.

Proposition 2.3.7 shows that, if one considers the modified p-value $\hat{p}_{mod,n}$, then the only source of invalidity of the bootstrap arises from the presence of m_{LP} and $\ddot{m}_{LP,mod}$, which are only functions of nuisance parameter not depending on higher order derivatives of g. The existence of a consistent estimator of $V_{LL,n}$, see (2.3.15), implies that a consistent estimator of m_{LP} and $\ddot{m}_{LP,mod}$ exists, such that it does not require ex-ante knowledge on the location of x. Therefore, if we let $H_{mod}(u)$ denote the limit of $\mathbb{P}\left(\hat{p}_{mod,n} \leq u\right)$, a uniformly consistent estimator of H_{mod} is

$$\hat{H}_{mod,n}(u) := \Phi\left(\hat{m}_{LP,mod,n}^{-1}\Phi^{-1}(u)\right) \tag{2.3.18}$$

where $\hat{m}_{LP,mod,n}^{2} := (\hat{v}_{1,mod,n}^{2} + \hat{v}_{2,mod,n}^{2} - 2\hat{v}_{12,mod,n})/Q_n^2\hat{v}_{1,n}^2$. And the LP bootstrap can provide asymptotically correct coverage both for interior and boundary points thanks to the following theorem.

Theorem 2.3.2 *Let the conditions of Proposition 2.3.1 and $x \in \mathbb{S}_x$, then*

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LP,mod}\right) \to 1-\alpha, \quad \alpha \in (0,1) \tag{2.3.19}$$

where $\widetilde{CI}_{LP,mod}$ is the prepivoted confidence interval in (2.2.6) with $\hat{H}_n = \hat{H}_{mod,n}$ and $\hat{L}_n$ the probability distribution (conditionally on the data) of the modified LP bootstrap statistic $T_{mod,n}^{*}$.

Remark 2.3.2 *Note that our results can also be extended by allowing the LP bootstrap DGP to be of the form $y_i^* = \breve{g}_n(x_i) + \varepsilon_i^*$ where $\breve{g}_n(x_i)$ is a local linear estimator adopting a different bandwidth with respect to h, say $\lambda = \lambda(n)$ with $\lambda \to 0$ as $n \to \infty$. By taking λ to be sufficiently larger than h, then ξ_{2n} would be asymptotically negligible. Standard bootstrap validity would then follow when x is an interior point, whereas a correction would still be needed when x is a boundary point. Hardle and Marron (1991) implemented a similar procedure, without the use of prepivoting, remarking significant distortions in finite samples. Prepivoting could then be relevantly applied to get better performances in finite samples through the presence of $\hat{m}_{mod,n}$, which would asymptotically, but not for small n, be equal to 1.*

2.4 Comparison with RBC methods

In this section, we compare the LP bootstrap-based prepivoted CIs presented in Section 3 to the RBC CIs proposed by Calonico et al. (2014, 2018). Specifically, in Section 4.1 we show that prepivoting can be applied to an alternative bootstrap DGP – which we label the fixed-local (FL) bootstrap – delivering CIs with asymptotically correct coverage under the same assumptions as those exploited in Section 3.1 and without the need for a correction specifically for boundary points. Note that Section 4.1 only includes the main technical result about the validity of the FL prepivoted CIs, whereas all the remaining details can be found in Appendix B. In Section 4.2, we show that the CIs presented in Section 4.1 are asymptotically equivalent to the RBC CIs. Finally, in Section 4.3, we compare the properties of the FL and LP bootstrap-based CIs in terms of asymptotic efficiency, showing that the LP bootstrap provides asymptotically shorter confidence lengths for all the most commonly used kernel functions.

2.4.1 FL bootstrap

Let us consider the alternative fixed-regressor wild bootstrap DGP

$$\tilde{g}_n(\tau) = \hat{\beta}_{0,n}(x) + \hat{\beta}_{1,n}(x)(\tau - x) + \frac{1}{2}\hat{\beta}_{2,n}(x)(\tau - x)^2$$

where $\hat{\beta}(x) := (\hat{\beta}_{0,n}(x), \hat{\beta}_{1,n}(x), \hat{\beta}_{2,n}(x))'$ are coefficients estimated via local quadratic regression at the fixed point x, i.e.

$$\hat{\beta}_n(x) = \underset{(b_0,b_1,b_2)\in\mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - b_0 - (x_i - x)b_1 - (x_i - x)^2 b_2\right) K\left(\frac{x_i - x}{h}\right)$$

Then, bootstrap data are generated as

$$y_i^* = \tilde{g}_n(x_i) + \varepsilon_i^* \tag{2.4.1}$$

where $\varepsilon_i^* := \tilde{\varepsilon}_i e_i^*$ where $\tilde{\varepsilon}_i$ are the leave-one-out residuals $\tilde{\varepsilon}_i := y_i - \tilde{g}_{n,-i}(x_i)$ and e_i^* is a iid random variable, conditionally on the original data, satisfying $\mathbb{E}^*[e_i^*] = 0$ and $\mathbb{E}^*[e_i^{*2}] = 1$. Similar bootstrap DGPs have been considered in the recent literature, specifically in the context of sharp and fuzzy RDDs in Bartalotti et al. (2017) and He and Bartalotti (2020).

Notice that (2.4.1) can be viewed as a "fixed" bootstrap DGP, in the sense that the bootstrap conditional expectation $E^*[y_i^*|x_i] = E^*[y_i^*]$ is based on a second-order Taylor approximation of the original conditional expectation $g(\tau) = E[y_i|x_i = \tau]$ around the fixed point x. This differs from the LP bootstrap, whose conditional expectation is based on different Taylor approximations of the original conditional expectations around each point x_i. Other than being computationally faster, this property of the FL bootstrap has the appealing advantage of not involving convolutions of the observed quantities, therefore bypassing the boundary issues shown in Section 3.2. Note, moreover, that we restrict here to the case in which $\tilde{g}$ is driven by coefficient estimators of local quadratic order for simplicity, but any order greater than that considered for deriving the test statistic can be applied if the latter is odd.

Let T_n be defined as in Section 3.1 and $\hat{g}_n^*(x)$ be a local linear estimator applied to the bootstrap sample generated as (2.4.1). The FL bootstrap analogue of T_n becomes $T_n^* = \sqrt{nh}(\hat{g}_n^*(x) - \tilde{g}_n(x))$, where

$$T_n^* = \tilde{B}_n + \xi_{1n}^*$$

such that $\tilde{B}_n := (nh)^{-1/2} \sum_{i=1}^n w(x_i) \left(\tilde{g}_n(x_i) - \tilde{g}_n(x)\right)$, and ξ_{1n}^* is the same as in Section 3.2 (with the only difference given by the residuals $\tilde{\varepsilon}_i$). It can be shown – see Proposition B.2.1 in Appendix B – that ξ_{1n}^* is asymptotically Gaussian with limit variance equivalent to that of ξ_{1n}.

We now consider the asymptotic behavior of $\tilde{B}_n - B_n$. By the definition of $\tilde{B}_n$, we have that:

$$\begin{aligned}\tilde{B}_n &:= \frac{1}{\sqrt{nh}} \sum_{i=1}^n w(x_i) \left(\tilde{g}_n(x_i) - \tilde{g}_n(x)\right) \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n w(x_i) \left(\hat{\beta}_1(x)(x_i - x) + \frac{1}{2}\hat{\beta}_2(x)(x_i - x)^2\right)\end{aligned}$$

where the second equality follows from $\sum_{i=1}^n w(x_i) = nh$. Moreover, since $\sum_{i=1}^n w(x_i)(x_i - x) = 0$,

$$\tilde{B}_n = \sqrt{nh^5}\frac{\hat{\beta}_{2,n}(x)}{2}\frac{1}{nh}\sum_{i=1}^n w(x_i)\left(\frac{x_i - x}{h}\right)^2 =: \sqrt{nh^5}\frac{\hat{\beta}_{2,n}(x)}{2}C_n \qquad (2.4.2)$$

We now aim at expanding this bootstrap bias. Specifically, we note that, since $\hat{\beta}_2(x)$ is not a consistent estimator of $g''(x)$, "standard" bootstrap validity will fail because $\tilde{B}_n - B_n \neq o_p(1)$. However, the FL bootstrap statistic will be such that $\tilde{B}_n - B_n$ is always asymptotically Gaussian and centered at zero, therefore allowing "standard" prepivoted CIs to deliver asymptotically correct coverage. Let $e_3' := (0, 0, 1)$, $y := (y_1, ...y_n)'$, $Z_{2x} := (Z_{2x1}, ..., Z_{2xn})'$ where $Z_{2xi} := (1, (x_i - x)/h), (x_i - x)/h)^2)'$. The equivalent kernel for the FL estimator becomes

$$l_i(x) := 2h^{-2}e_3'\left(\frac{Z_{2x}'W_xZ_{2x}}{n}\right)^{-1}Z_{2xi}K\left(\frac{x_i - x}{h}\right),$$

so that $\hat{\beta}_2(x) := (nh)^{-1}\sum_{i=1}^n l_i(x)y_i = (nh)^{-1}\sum_{i=1}^n l_i(x)(g(x_i) + \varepsilon_i)$. By a Taylor expansion of $g(x_i)$ around x, one can show that $(nh)^{-1}\sum_{i=1}^n l_i(x)g(x_i) = g''(x) + O_p(h)$, so that

$$\hat{\beta}_2(x) = g''(x) + \frac{1}{nh}\sum_{i=1}^n l_i(x)\varepsilon_i + o_p(1) \qquad (2.4.3)$$

By letting $\tilde{\xi}_{2n} := (nh)^{-1/2}\sum_{i=1}^n \tilde{l}(x_i)\varepsilon_i$, with $\tilde{l}_i(x) = h^2 l_i(x)/2$, the above implies that

$$\tilde{B}_n = B_{AT,n} + \tilde{\xi}_{2n} + o_p(1) \qquad (2.4.4)$$

so that $\tilde{B}_n - B_n$ is asymptotically centered at zero $\forall x \in \mathbb{S}_x$. Let $\tilde{\xi}_n := (\xi_{1n}, \tilde{\xi}_{2n})'$ and $V_{FL,n} := \mathbb{V}[\tilde{\xi}_n|\mathcal{X}_n]$, then we show in Proposition B.2.2, in Appendix B, that $\tilde{\xi}_n$ is asymptotically Gaussian with variance defined as the limit of $V_{FL,n}$. The fact that the limit of $v_{2,FL,n} := \mathbb{V}[\tilde{\xi}_{2n}|\mathcal{X}_n]$ is greater than 0 – both for interior and boundary points – is the only source of invalidity of the FL bootstrap "naive" (i.e., not prepivoted) CIs. Similarly than for the LP bootstrap, the limit of $V_{FL,n}$ can be estimated without ex-ante knowledge about the location of x via the consistent estimator

$$\hat{V}_{FL,n} := \frac{1}{nh}\sum_{i=1}^{n}\begin{pmatrix} w_i(x) & w_i(x)C_n\tilde{l}_i(x) \\ w_i(x)C_n\tilde{l}_i(x) & C_n^2\tilde{l}_i^2(x)\end{pmatrix}\tilde{\varepsilon}_i^2, \tag{2.4.5}$$

such that $\hat{V}_{FL,n} - V_{FL,n} = o_p(1)$. Note that (2.4.5) is equivalent to the $\hat{\sigma}^2_{us}$-HC3 formula for standard errors proposed in Calonico et al. (2018).

Since $\hat{V}_{FL,n} - V_{FL,n} = o_p(1)$, an estimator $\hat{m}_{FL,n}$ such that $\hat{m}_{FL,n} = m_{FL} + o_p(1)$ exists if x is an interior point; moreover, the same $\hat{m}_n$ is such that $\hat{m}_{FL,n} = \ddot{m}_{FL} + o_p(1)$ if x is a boundary point. As in Section 3.2; $\hat{m}_n$ is a plug-in estimator of the form

$$\hat{m}_{FL,n} := \frac{\sqrt{\hat{v}^2_{1,FL,n} + \hat{v}^2_{2,FL,n} - 2\hat{v}_{12,FL,n}}}{\hat{v}_{1,FL,n}} \tag{2.4.6}$$

which guarantees the presence of a uniformly consistent estimator of the cdf of $\hat{p}_n$, i.e.,

$$\hat{H}_{FL,n}(u) = \left(\hat{m}^{-1}_{FL,n}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{2.4.7}$$

The following theorem formalizes the validity of the FL bootstrap-based prepivoted confidence intervals.

THEOREM 2.4.1 *Let Assumptions 1-3 hold, then, $\forall x \in \mathbb{S}_x$,*

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{FL}\right) \to 1-\alpha, \quad \alpha \in (0,1) \tag{2.4.8}$$

where $\widetilde{CI}_{FL}$ is the prepivoted confidence interval in (2.2.6) with $\hat{H}_n = \hat{H}_{FL,n}$ and $\hat{L}_n$ the probability distribution (conditional on the data) of the FL bootstrap statistic T_n^*.

2.4.2 Asymptotic equivalence with RBC

We now show that the FL bootstrap-based prepivoted confidence intervals $\widetilde{CI}_{FL}$ are asymptotically equivalent to those proposed in Calonico et al. (2018).

The FL bootstrap-based prepivoted CIs are defined as:

$$\widetilde{CI}_{FL} := \left[\hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_{FL,n}\Phi^{-1}\left(1-\alpha/2\right)\right)\right), \hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_{FL,n}\Phi^{-1}\left(\alpha/2\right)\right)\right)\right]$$

Since $\tilde{B}_n$ is a measurable function of D_n and $\hat{L}_n$ is a probability distribution conditional on D_n, we have

$$\hat{L}_n^{-1}(u) = \tilde{B}_n + \hat{L}^{-1}_{\xi,n}(u) \tag{2.4.9}$$

where $\hat{L}_{\xi,n}(u) := \mathbb{P}^*\left(\xi^*_{1n} \le u\right)$. Moreover, by Proposition 2.3.2,

$$\hat{L}^{-1}_{\xi,n}(u) = v_1\Phi^{-1}(u) + o_p(1) \qquad \text{uniformly in } u \in (0,1) \tag{2.4.10}$$

therefore,

$$\widetilde{CI}_{FL} = \left[\left(\hat{g}_n(x) - (nh)^{-1/2}\tilde{B}_n\right) \pm (nh)^{-1/2}\hat{v}_{d,FL,n}\Phi^{-1}\left(1-\alpha/2\right)\right] + o_p((nh)^{-1/2}) \tag{2.4.11}$$

Equation (2.4.11) shows that the dominant part of $\widetilde{CI}_{FL}$ is equal to the CI proposed by Calonico et al. (2018), based on the "robust bias correction" (RBC) method, which we label CI_{AT}. To see this, note that $(nh)^{-1/2}\tilde{B}_n$ is exactly equal to their local quadratic bias estimator; moreover, $(nh)^{-1/2}\hat{v}_{d,FL}$ is equivalent to their studentization term, defined as a consistent estimator of the variance of $\hat{g}_n(x)-(nh)^{-1/2}\tilde{B}_n$. Since we focused on the leave-one-out residuals $\tilde{\varepsilon}_i$, our standard errors are equivalent to those Calonico et al. (2018) label "$\hat{\sigma}^2_{us}$-HC3", though implementation of their $\hat{\sigma}^2_{us}$-HCk method with $k=0,1,2,3$ is possible by appropriately changing the functional form of $\tilde{\varepsilon}_i$. .

2.4.3 Efficiency considerations

We now compare the efficiency properties of the two proposed confidence intervals. We start with the prepivoted CIs based on the LP bootstrap, which are defined as:

$$\widetilde{CI}_{LP} := \left[\hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_n\Phi^{-1}\left(1-\alpha/2\right)\right)\right), \hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_n\Phi^{-1}\left(\alpha/2\right)\right)\right)\right]$$

By the same arguments as those used in Section 4.2 and by exploiting Proposition 2.3.2, we have that

$$\widetilde{CI}_{LP} = \left[\left(\hat{g}_n(x) - (nh)^{-1/2}\tilde{B}_n\right) \pm (nh)^{-1/2}\hat{v}_{d,LP,n}\Phi^{-1}\left(1-\alpha/2\right)\right] + o_p((nh)^{-1/2}) \quad (2.4.12)$$

Our efficiency considerations will be based on comparisons on the dominant terms of the absolute length of the CIs. We let $\Delta(\widetilde{CI}_{LP})$ denote the absolute length of $\widetilde{CI}_{LP}$; then, by (2.4.9) and (2.4.10),

$$\Delta(\widetilde{CI}_{LP}) = \begin{cases} (nh)^{-1/2}v_{d,LP}|\Phi^{-1}(1-\alpha/2)-\Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is an interior point;} \\ (nh)^{-1/2}\ddot{v}_{d,LP}|\Phi^{-1}(1-\alpha/2)-\Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is a boundary point;} \end{cases}$$

By the same reasoning, by letting $\Delta(\widetilde{CI}_{FL})$ denote the length of the FL bootstrap-based prepivoted CI, we have

$$\Delta(\widetilde{CI}_{FL}) = \begin{cases} (nh)^{-1/2}v_{d,FL}|\Phi^{-1}(1-\alpha/2)-\Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is an interior point;} \\ (nh)^{-1/2}\ddot{v}_{d,FL}|\Phi^{-1}(1-\alpha/2)-\Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is a boundary point;} \end{cases}$$

Therefore, efficiency comparisons between $\widetilde{CI}_{LP}$ and $\widetilde{CI}_{FL}$ can be based on the difference between $v_{d,LP}$ and $v_{d,FL}$ if x is an interior point and between $\ddot{v}_{d,LP}$ and $\ddot{v}_{d,FL}$ if x is a boundary point. The following Proposition summarizes the properties of these quantities.

Proposition 2.4.1 *Let Assumptions 1-3 hold, then,*

$$\begin{pmatrix} v^2_{d,LP} \\ v^2_{d,FL} \end{pmatrix} = \frac{\sigma^2(x)}{f(x)}\begin{pmatrix} \mathcal{K}_{v_d,LP} \\ \mathcal{K}_{v_d,FL} \end{pmatrix}; \qquad \begin{pmatrix} \ddot{v}^2_{d,LP} \\ \ddot{v}^2_{d,FL} \end{pmatrix} = \frac{\sigma^2(0)}{f(0)}\begin{pmatrix} \ddot{\mathcal{K}}_{v_d,LP} \\ \ddot{\mathcal{K}}_{v_d,FL} \end{pmatrix};$$

where $\mathcal{K}_{v_d,LP}$, $\mathcal{K}_{v_d,FL}$, $\ddot{\mathcal{K}}_{v_d,LP}$ and $\ddot{\mathcal{K}}_{v_d,FL}$ are measurable functions of the kernel K, defined in Appendix B.

Proposition 2.4.1 shows that efficiency considerations can be reduced to a comparison of the known quantities $\mathcal{K}_{v_d,LP}$, $\mathcal{K}_{v_d,FL}$, $\ddot{\mathcal{K}}_{v_d,LP}$ and $\ddot{\mathcal{K}}_{v_d,FL}$. Table 1 shows the value of these quantities – computed via numerical integration – when the most commonly used kernel functions are adopted, showing that the LP bootstrap yields shorter confidence intervals under each considered scenario. This result will be confirmed by the Monte Carlo analysis shown in Section 5. Let us consider, for instance, the two most popular choices of K, i.e., the Epanechnikov's kernel when x is an interior point and the Triangular kernel when x is a boundary point. In the first scenario, the FL bootstrap provides about 21% larger confidence intervals, whereas in the second, the FL bootstrap displays a theoretical length which higher by about 20%.

Table 2.2: Comparison of the measurable components of v_d^2

	Interior		Boundary	
K	$\mathcal{K}_{v_d,LP}$	$\mathcal{K}_{v_d,FL}$	$\ddot{\mathcal{K}}_{v_d,FL}$	$\ddot{\mathcal{K}}_{v_d,FL}$
Triangular	0.95	1.33	7.18	10.29
Epanechnikov	0.85	1.25	6.80	9.82
Biweight	1.01	1.41	7.67	10.87
Triweight	1.15	1.55	8.54	11.87

Finally, since the dominant components in CI_{AT} and $\widetilde{CI}_{FL}$ are asymptotically equivalent, we can conclude that the dominant part of $\widetilde{CI}_{LP}$ is also smaller than that of CI_{AT}, providing theoretical justification for the numerical results shown in Section 5

2.5 Prepivoting in (Sharp) Regression-Discontinuity Design

As an application of our theory for local polynomial estimators, we now consider the relevant example of (sharp) regression-discontinuity. Specifically, let

$$y_i = g_0(x_i)\mathbb{I}_{\{x_i<c\}} + g_1(x_i)\mathbb{I}_{\{x_i\geq c\}} + \varepsilon_i, \qquad i = 1, \dots, n; \tag{2.5.1}$$

where $E(\varepsilon_i|x_i) = 0$, $\mathbb{V}(\varepsilon_i|x_i) =: \sigma^2(x_i)$, x_i is a random variable with bounded support $\mathbb{S}_x := [a, b]$, $(a, b) \in \mathbb{R}^2$, and pdf $f(x)$ such that $f : \mathbb{S}_x \to (0, +\infty)$, while $g_0 : [a, c] \to \mathbb{R}$ and $g_1 : [c, b] \to \mathbb{R}$. For simplicity of exposition and without loss of generality, we set $(a, b, c) = (-1, 1, 0)$. We are interested in estimating the difference on the conditional expectations at the right and at the left of the cutoff $c = 0$; therefore, our parameter of interest is

$$\tau_{srd} := g_1(0+) - g_0(0-). \tag{2.5.2}$$

In sharp RD, τ_{srd} identifies the average treatment effect at the threshold; see Hahn, Todd, and van der Klaauw (2001). Let $\hat{g}_{0,n}(0)$ and $\hat{g}_{1,n}(0)$ denote the local linear estimators (at the induced boundary $c = 0$) of $g_0(0+)$ and $g_0(0-)$, respectively; then, a natural estimator of τ_{srd} is

$$\hat{\tau}_{srd} := \hat{g}_{1,n}(0) - \hat{g}_{0,n}(0). \tag{2.5.3}$$

Being the difference of two local polynomial estimators, the bias of $\hat{\tau}_{srd}$ will be equivalent to the difference of biases of the two estimator. Indeed, if we let $\mathcal{T}_n := \sqrt{nh}(\hat{\tau}_n - \tau_{srd})$, we have that $\mathcal{T}_n = \mathcal{B}_n + \xi_{srd,1n}$. The asymptotic properties of $\xi_{srd,1n}$ and $\mathcal{B}_n$ immediately follow from the results in Section 3. Intuitively, $\mathcal{B}_n$ will converge in probability to a term which is proportional to the difference between the right and left derivative of g at 0, whereas $\xi_{srd,1n}$ is asymptotically Gaussian and centered at zero.

Let us now consider our prepivoted CIs in this context. The LP bootstrap DGP becomes:

$$y_i^* = \hat{g}_{0,n}(x_i)\mathbb{I}_{\{x_i<0\}} + \hat{g}_{1,n}(x_i)\mathbb{I}_{\{x_i\geq 0\}} + \varepsilon_i^* \tag{2.5.4}$$

where $\varepsilon_i^* := \hat{\varepsilon}_i e_i^*$ and $\hat{\varepsilon}_i$ the leave-one-out residuals $\hat{\varepsilon}_i := y_i - (\hat{g}_{0,-i,n}(x_i)\mathbb{I}_{\{x_i<0\}} + \hat{g}_{1,-i,n}(x_i)\mathbb{I}_{\{x_i\geq 0\}})$. On the other hand, if we let $\tilde{g}_{0,n}(x_i) := \hat{\beta}_{00} + \hat{\beta}_{01}x_i + \hat{\beta}_{02}x_i^2/2$ and $\tilde{g}_{1,n}(x_i) := \hat{\beta}_{10} + \hat{\beta}_{11}x_i + \hat{\beta}_{12}x_i^2/2$, where $\hat{\beta}_0 := (\hat{\beta}_{00}, \hat{\beta}_{01}, \hat{\beta}_{02})'$ and $\hat{\beta}_1 := (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12})'$ are the coefficient obtained through the usual local quadratic estimator at the left and at the right of the cutoff, respectively, we can define the FL bootstrap DGP as

$$y_i^* = \tilde{g}_{0,n}(x_i)\mathbb{I}_{\{x_i<0\}} + \tilde{g}_{1,n}(x_i)\mathbb{I}_{\{x_i\geq 0\}} + \varepsilon_i^* \tag{2.5.5}$$

By performing local linear estimation to both of these bootstrap DGP, one can obtain an estimator $\hat{\tau}_n^*$ analogously to (2.5.3) and the bootstrap test statistic $\mathcal{T}_n^*$. Such test statistic will be equal to $Q_n\sqrt{nh}(\hat{\tau}_n^* - \hat{\tau}_n)$ for the "modified" LP bootstrap and $\sqrt{nh}(\hat{\tau}_n^* - (\tilde{g}_{1n}(0) - \tilde{g}_{0n}(0)))$ for the FL bootstrap.

Remark 2.5.1 *Note that the modification for the LP bootstrap statistic is identical if the same kernel function is used to the right and to the left of the cutoff. Indeed, one could decompose the unmodified LP bootstrap bias into two terms: one considering the contribution of the bias arising from the observations to the left of the cutoff, one considering those arising from the observations to the right of the cutoff. By (2.3.5), the two contributions will be a product of a weighted convolution of $((x_i - x)/h)^2$ and the right- (or left-) second order derivative of g at the cutoff. Crucially, the limit of the weighted convolution of $((x_i - x)/h)^2$ is the same, no matter if only the contributions to the right or to the left of the cutoff are considered, if the same kernel K is considered. The modification can still be generalized to allow for different kernels to the right or to the left of the cutoff by decomposing the unmodified LP bootstrap statistic in two components (one considering the contributions at each side of the cutoff) and re-weighting each component according to the different kernel used.*

If we denote by $\hat{\mathcal{B}}_n$ the generic bias of the bootstrap test statistic (either "modified" LP or FL), then it follows from the results in Section 3.2 and 3.3 that $\hat{\mathcal{B}}_n - \mathcal{B}_n =:$

$\xi_{srd,2n} + o_p(1)$, where $\xi_{srd,2n}$ is asymptotically Gaussian and centered at zero, allowing for the constructions of the prepivoted CIs

$$\widetilde{CI}_{srd} := \left[\hat{\tau}_n - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_{srd,n}\Phi^{-1}(1-\alpha/2)\right)\right), \hat{\tau}_n - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_{srd,n}\Phi^{-1}(\alpha/2)\right)\right)\right]$$

where $\hat{L}_n := \mathbb{P}^*\left(\mathcal{T}_n^* \leq \mathcal{T}_n\right)$ and $\hat{m}_{srd,n}^2$ a consistent estimator of $m_{srd}^2 := \text{plim}_{n\to\infty}\{\mathbb{V}[\mathcal{T}_n - \hat{\mathcal{B}}_n|\mathcal{X}_n]\}/\text{plim}_{n\to\infty}\{\mathbb{V}^*[\mathcal{T}_n^* - \hat{\mathcal{B}}_n]\}$. As we have seen in Section 3.2 and 3.3, such estimator exists and can be based on leave-one-out residuals from the original model.

2.6 Monte Carlo

We now discuss the finite sample performance of the proposed CIs and compare them both with invalid bootstrap CI (i.e., not prepivoted), as well as with the RBC CIs proposed by Calonico et al. (2018), through the results of Monte Carlo simulations. Specifically, we focus on two simulation designs, which we label DGP1 and DGP2. Both DGP's take the form

$$y_i = g(x_i) + \varepsilon_i$$

where $\varepsilon_i \sim iidN(0, \sigma^2)$. In DGP1, $\sigma = 1$, $g(x) = g_1(x) = \sin(3\pi x/2)[1 + 18x^2(\text{sgn}(x) + 1)]^{-1}$ and $x_i \sim iidU_{[-1,1]}$; whereas in DGP2, $\sigma = 0.1295$, $g(x) = g_2(x) = 0.52 + 0.84x - 0.30x^2 + 2.397x^3 - 0.901x^4 + 3.56x^5$ and $x_i \sim iidU_{[0,1]}$. On the one hand, DGP1 is equivalent to a simulation setup previously considered in Berry, Carroll, and Ruppert (2001), Hall and Horowitz (2013) and Calonico et al. (2018). On the other hand, the conditional expectation $g_2(x)$ and the value of σ in DGP2 are taken from Model 3 in Calonico et al. (2014); specifically, $g_2(x)$ is equal to the conditional at the right of the cutoff in a sharp RD setup and arises from a modification of the estimated coefficients in Lee (2008). For both DGP's, we consider estimation for an interior and a boundary point. In DGP1, the evaluation points are $x = -1/3$ and $x = -1$, whereas for DGP2 those are $x = 0.5$ and $x = 0$. Under all the considered scenarios, we make use of the Epanechnikov's kernel and the MSE-optimal bandwidth. 5000 independent Monte Carlo draws are generated, with 999 bootstrap replications for each Monte Carlo draw. For all the wild bootstrap DGP's, $\{e_i^*\}$ is a sequence of iid random variables distributed as Rademacher on $[-1, 1]$.

The results of the Monte Carlo simulations are summarized in Table 2, where average empirical coverage and length of the CIs are shown. Other than the prepivoted LP, $\widetilde{CI}_{LP}$, modified LP, $\widetilde{CI}_{LP,mood}$, and FL, $\widetilde{CI}_{FL}$, bootstrap CIs, we also report results for the "naive", i.e. not prepivoted, CIs based on the LP bootstrap DGP, CI_{LP}, and FL bootstrap DGP, CI_{FL}, as well as those based on RBC, CI_{AT}. First of all, we detect significant undercoverage of both the "naive" (i.e., not prepivoted) bootstrap CIs, thus underlining the practical need of proper debiasing tecniques. Second, we observe that the prepivoted CIs show empirical coverage probabilities which are very close to the nominal levels – and comparable to RBC – under all considered scenarios. Moreover, asymptotic

Table 2.3: Coverage and length of 95% confidence intervals

n	h	CI_{LP}	CI_{FL}	$\widetilde{CI}_{LP}$	$\widetilde{CI}_{LP,mod}$	$\widetilde{CI}_{FL}$	CI_{AT}
DGP1: Interior Point							
		coverage					
250	0.189	89.1	82.4	93.9	94.1	93.2	94.8
500	0.165	89.1	82.4	93.5	93.7	93.7	95.0
750	0.152	89.7	82.1	94.5	94.4	94.2	95.1
1000	0.143	90.0	82.8	94.6	94.7	94.3	95.0
		length					
250	0.189	0.637	0.641	0.747	0.754	0.884	0.943
500	0.165	0.479	0.481	0.565	0.567	0.677	0.701
750	0.152	0.405	0.406	0.479	0.481	0.575	0.590
1000	0.143	0.361	0.361	0.427	0.428	0.514	0.525
DGP1: Boundary Point							
		coverage					
250	0.353	87.7	82.4	93.4	94.9	93.4	95.4
500	0.307	88.0	81.7	94.4	95.9	93.0	94.6
750	0.283	89.7	82.9	95.7	96.8	94.2	95.4
1000	0.267	90.0	82.3	95.6	96.7	94.2	95.3
		length					
250	0.353	1.272	1.365	1.558	1.775	1.853	2.152
500	0.307	0.963	0.998	1.194	1.335	1.394	1.522
750	0.283	0.815	0.835	1.017	1.130	1.187	1.264
1000	0.267	0.725	0.740	0.907	1.003	1.057	1.110
DGP2: Interior Point							
		coverage					
250	0.209	88.2	81.6	93.5	93.5	93.4	94.1
500	0.182	89.5	82.6	94.5	94.7	94.8	95.2
750	0.168	89.0	81.6	94.1	94.2	94.3	94.6
1000	0.158	88.8	81.8	94.5	94.5	94.4	94.5
		length					
250	0.209	0.055	0.055	0.065	0.066	0.078	0.080
500	0.182	0.042	0.042	0.049	0.050	0.059	0.060
750	0.168	0.035	0.035	0.042	0.042	0.050	0.051
1000	0.158	0.031	0.031	0.037	0.037	0.045	0.045
DGP2: Boundary Point							
		coverage					
250	0.574	69.4	87.8	82.5	92.2	96.3	97.1
500	0.500	85.2	84.8	93.5	97.4	96.0	96.0
750	0.461	88.2	82.7	94.7	96.8	94.7	95.0
1000	0.435	90.5	83.2	96.2	96.6	95.0	95.7
		length					
250	0.574	0.092	0.106	0.117	0.130	0.148	0.152
500	0.500	0.069	0.075	0.087	0.096	0.106	0.109
750	0.461	0.059	0.062	0.074	0.081	0.089	0.090

equivalence of $\widetilde{CI}_{FL}$ and CI_{AT}, as stated in Section 3.4, is confirmed by the numerical results, as the two methods behave very closely to each other both in terms of empirical coverage and average interval length. Finally, the efficiency results theoretically analyzed in Section 3.4 are confirmed by the numerical analysis, where CI_{AT} shows, for $n = 1000$, between 9%-22% larger confidence intervals with respect to $\widetilde{CI}_{LP,mod}$.

2.7 Conclusion

This paper advances the literature on nonparametric regression and RD designs by addressing a fundamental challenge: obtaining valid inference in the presence of asymptotic bias without resorting to undersmoothing or direct bias correction. We introduce two bootstrap methods – the LP and FL bootstraps – that restore validity and deliver asymptotically correct confidence intervals in a computationally practical manner via the use of prepivoting. While the FL bootstrap is asymptotically equivalent to RBC methods, the LP bootstrap offers higher efficiency, making it particularly advantageous in empirical applications. Importantly, our "modified" prepivoting approach ensures robustness of the widely-used LP bootstrap DGP even at boundary points, addressing a critical gap in existing methods. Monte Carlo simulations corroborate the theoretical advantages of our methods, showing empirical coverage close to the nominal levels and efficiency of the LP bootstrap across a variety of scenarios. Furthermore, we show that our methodology extends to RD designs, a cornerstone of applied econometrics. These results provide researchers with powerful tools for unbiased and efficient inference in nonparametric regression, promising to enhance the reliability of quasi-experimental analysis in economics and beyond.

Chapter 3

Parameters on the Boundary in Predictive Regression

(written with Giuseppe Cavaliere and Iliyan Georgiev)

3.1 Introduction

In this paper we revisit the well-known problem of bootstrap inference in regressions with parameter space defined by means of smooth inequality constraints. For instance, consider the setup of a regression $y_t = \alpha + \beta x_{t-1} + \varepsilon_t$ where the parameter space for (α, β) is defined by the constraint $\beta \geq 0$. This framework arises when only the possibilities $\beta = 0$ of no predictability (or no first-order Granger causality, generalizable to higher orders), and $\beta > 0$ of sign-restricted predictability, are entertained, and the model is estimated under the constraint $\beta \in [0, \infty)$. In applications, economic theory is often informative about the direction of predictability, and such information could be used to improve the efficiency of estimators and increase the power of hypotheses tests. A prominent example is provided by predictive regressions for financial returns; see, e.g., Phillips (2014) and the references therein. Interest can then be in testing the very hypothesis of no predictability (i.e., $\beta = 0$) by means of a one-sided test, or a special case of this hypothesis (e.g., $\alpha = \beta = 0$), or a hypothesis where the parameter vector may but need not lie on the boundary of the parameter space (e.g., $\alpha + \beta = 0$).

While in this context the bootstrap is potentially useful, its application is not straightforward if the parameter vector may lie on the boundary of the parameter space; see Andrews (2000). In particular, as we discuss in the following, even in a simple location model where the parameter space is a closed half-line, the cumulative distribution function [cdf] of the parametric bootstrap t-statistic, conditional on the original data, converges weakly to a random cdf, rather than to the target asymptotic distribution of the t-statistic computed from the original data.

Our first contribution is to show that in predictive regressions with parameter values on the boundary, the distribution of fixed regressor[1] bootstrap statistics, like the t-statistic for $\beta = 0$ in the regression above, may be random in the limit. Limiting randomness may arise in two ways. A first possible source of randomness in the limit bootstrap measure is in the non-stationarity of the regressor, which operates through the random limits of sample product moments. This is hardly surprising, see e.g. Georgiev et al. (2019). A second potential source of randomness is the location of the parameter vector on the boundary of the parameter space. Invalidity of standard bootstrap schemes when a parameter is on the boundary was initially discussed in Andrews (2000), where a simple location-model example was given; see also Chatterjee and Lahiri (2011). In the context of hypotheses tests in predictive regressions, we revisit Andrews' result and show that, for a general bootstrap scheme, the occurrence or non-occurrence of limiting bootstrap randomness due to the possible location of a parameter on the boundary of the parameter space depends on how well the bootstrap scheme approximates the mutual position of three objects: (i) the boundary, (ii) the parameter set identified by the null hypothesis, and (iii) the true parameter value. Standard bootstrap approximations of this mutual position may not be sufficiently precise, giving rise to complex conditioning in the limit bootstrap distribution, with ensuing bootstrap validity only for special types of statistics.

Our second contribution is to show that certain non-standard bootstrap schemes, designed to provide a better match with the geometric configuration in the original parameter space, give rise to limit bootstrap distributions where randomness, if present, is not attributable to the boundary value of the parameter vector. This fact allows us to establish bootstrap validity in an 'unconditional' sense; see Cavaliere and Georgiev (2020). That is, although randomness of the limiting bootstrap cdf prevents the possibility that the bootstrap could mimic the asymptotic distribution of the original statistic, we can show that in large samples bootstrap tests and asymptotic tests are correctly sized for essentially the same set of nominal sizes.

Formally, we make use of the following definition, which generalizes the definition of unconditional bootstrap validity given in Cavaliere and Georgiev (2020, p.2555). Let p_n and p_n^* be respectively the p-value of an asymptotic test and of its bootstrap analogue. Let also

$$C := \{q \in (0,1) : \lim_{n\to\infty} P(p_n \leq q) = q|\mathsf{H}_0\},$$

such that a test rejecting for $p_n \leq q$ (or for $p_n > q$) is correctly sized for nominal significance levels q (resp. $1-q$) with $q \in C$, as $n \to \infty$.[2] If, under the null hypothesis H_0,

$$P(p_n^* \leq q) \to q \text{ for all } q \in intC, \tag{3.1.1}$$

[1]We focus on 'fixed regressor' bootstrap schemes as they do not require knowledge on the regressor generating process. For instance, and in contrast to recursive-based schemes, they can be applied to both I(0) and I(1) settings.

[2]For $q \in \text{int}C$ it holds that $P(p_n = q) \to 0$ and rejections for $p_n \leq q$ (or $p_n > q$) are asymptotically equivalent to rejections for $p_n < q$ (or $p_n \geq q$).

where $intC$ denotes the interior of the set C, we say that the bootstrap test based on p_n^* is valid for H_0.[3] The meaning is that the bootstrap test and the asymptotic test are first-order asymptotically equivalent in terms of correct size control. In particular, bootstrap validity for simple hypotheses H_0 characterizes pointwise size control.

Notice that bootstrap validity as in (1.1) is implied by the classic definition of bootstrap consistency, namely that $\sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \to_p 0$ for a bootstrap statistic with cdf F_n^* conditionally on the data and an original test statistic with continuous asymptotic cdf F. The converse does not hold; that is, (3.1.1) does not imply classic bootstrap consistency, see the discussion in Cavaliere and Georgiev (2020).

For test statistics whose asymptotic distribution is continuous, it holds that $intC = (0, 1)$ and hence condition (3.1.1) should hold for all $q \in (0, 1)$ for the bootstrap to be valid. Unfortunately, parameter values on the boundary of the parameter space may induce discontinuities in the limiting cdf's, such that not even the exact p-values of the associated tests are asymptotically standard uniform on $[0, 1]$. This makes the above weaker version of the validity definition unavoidable.

Finally, we turn to the special case of one-sided tests for the null hypothesis that the parameter vector lies on the boundary of the parameter space, such that the boundary coincides with the parameter set identified by the null hypothesis. This case provides a transparent example of a limit bootstrap cdf which is random only on a subset of its domain. Then, if bootstrap validity is defined as in (3.1.1), in this case also some standard bootstrap schemes can be proved to be valid.

This paper is related to recent work by Fang and Santos (2019) and Hong and Li (2020). The latter two papers propose nonstandard bootstrap schemes – involving a tuning tool – which correct the inconsistency of 'classic' bootstrap methods in settings that cover parameters on the boundary as a special case. The main difference from the present contribution is that our theory applies to random limit bootstrap measures. Thus, Fang and Santos (2019) consider bootstrap inference in settings where the target asymptotic distribution, say that of a random element τ, can be thought of as a transformation φ of another random element τ', and both the distribution of τ' and the transformation φ need to be estimated; see also the related works by Dümbgen (1993), Hirano and Porter (2012), Fang (2014) and Chen and Fang (2019). Although Fang and Santos (2019) consider deterministic φ and the unconditional distribution of τ', such that their results are not directly applicable here, their way of conceptualizing the problem remains fruitful also in the case of random φ and random conditional distributions $\tau'|\tau''$ (for some random element τ''). We discuss this in Section 3.5.2.

Our contribution is also related to Hong and Li (2020), who propose a 'numerical bootstrap' which is valid in settings where a parameter space can be approximated locally

[3] Bootstrap unconditional validity as in Cavaliere and Georgiev (2020) is obtained as the special case $intC = (0, 1)$.

by a cone with vertex at the true value of the parameters; see Geyer (1994) for a detailed discussion of the approximation. Both the approaches in this paper and that by Hong and Li (2020) are connected to the large body of literature considering estimation and inference for constrained M-estimators; see, among others, Geyer (1994), Andrews (1999, 2000), and the references therein. In Section 3.5.2 we argue that, when applied to a restricted predictive regression, the 'numerical bootstrap' of Hong and Li (2020) performs a geometric approximation of the kind we propose, though at the cost of a slower-than-standard convergence rate for the resulting bootstrap estimator.

We present our main idea using first a simple location model for i.i.d. scalar data whose location parameter is constrained to be positive. This is done in Section 3.2. The predictive regression framework is presented in Section 3.3; in this section we also show that the bootstrap limit measure associated with standard fixed regressor wild bootstrap schemes is random. A new family of bootstrap algorithms and their validity are discussed in Section 3.4. Results on the validity of one-sided tests, connections to the previous literature, and uniform size control for the bootstrap tests are discussed in Section 3.5. Section 3.6 provides simulation evidence, whereas Section 3.7 concludes. Proofs are collected in the Appendix.

Notation and definitions

We use the following notation throughout. The spaces of càdlàg functions $[0,1] \to \mathbb{R}^n$, $[0,1] \to \mathbb{R}^{m\times n}$ and $\mathbb{R} \to \mathbb{R}$, all equipped with the respective Skorokhod J_1-topologies, are denoted by D_n, $D_{m\times n}$ and $D(\mathbb{R})$, respectively; see Kallenberg (1997, Appendix A2). For $n = 1$, the subscript in D_n is suppressed. $C_n(\mathbb{R}^n)$ is the space of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ equipped with the topology of uniform convergence on compacts. Integrals are over $[0,1]$ unless otherwise stated, Φ is the standard Gaussian cdf, $U_{[0,1]}$ is the uniform distribution on $[0,1]$ and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. If F is a cdf, possibly random, F^{-1} stands for the right-continuous generalized inverse, i.e., $F^{-1}(u) := \sup\{v \in \mathbb{R} : F(v) \leq u\}$, $u \in \mathbb{R}$. Unless differently specified, limits are for $n \to \infty$.

With (Z_n, Y_n) and (Z, Y) being random elements of the metric spaces $S_Z \times S_{Y_n}$ and $S_Z \times S_Y$ ($n \in \mathbb{N}$), and defined on a common probability space, we denote by '$Z_n|Y_n \xrightarrow{w}_p Z|Y$' (resp. '$Z_n|Y_n \xrightarrow{w}_{a.s.} Z|Y$') the fact that $E\{g(Z_n)|Y_n\} \to E\{g(Z)|Y\}$ in probability (resp. a.s.) for all bounded continuous functions $g : S_Z \to \mathbb{R}$. When Z_n is a bootstrap statistic and Y_n denotes the original data, we write '$Z_n \xrightarrow{w^*}_p Z|Y$' (resp. '$Z_n \xrightarrow{w^*}_{a.s.} Z|Y$'). Finally, with (Z_n, Y_n) and (Z, Y) possibly defined on different probability spaces, '$Z_n|Y_n \xrightarrow{w}_w Z|Y$' means that $E(g(Z_n)|Y_n) \xrightarrow{w} E(g(Z)|Y)$ for all bounded continuous functions $g : S_Z \to \mathbb{R}$, see Kallenberg (1997, 2017); we label this fact 'weak convergence in distribution'. For the special case of scalar random variables Z_n and Z, if the conditional distribution $Z|Y$ is diffuse (non-atomic), weak convergence in distribution is equivalent to the following weak convergence in $D(\mathbb{R})$:

$$F_n(\cdot|Y_n) := P(Z_n \leq \cdot|Y_n) \overset{w}{\to} P(Z \leq \cdot|Y) =: F(\cdot|Y). \tag{3.1.2}$$

When Z_n is a bootstrap statistic and conditioning is on the original data, we use the notation '$\overset{w^*}{\to}_w$'. For multivariate generalizations we refer to Cavaliere and Georgiev (2020, Appendix A).

3.2 Preview of the results in a location model

To illustrate the main arguments that will be proposed in the predictive regression framework later, consider as in Andrews (2000) and Cavaliere et al. (2017) the location model

$$y_t = \theta + \varepsilon_t \ (t = 1, ..., n)$$

where the ε_t's are i.i.d.$(0, 1)$ and the parameter space is $\Theta := \{\theta \in \mathbb{R} : \theta \geq 0\}$. Interest is in inference on the true value θ_0 of θ by using the Gaussian QMLE, $\hat{\theta}$. With $l_n(\theta) := -\frac{1}{2}\sum_{t=1}^{n}(y_t - \theta)^2$, we find $\hat{\theta} := \arg\max_{\theta\in\Theta} l_n(\theta) = \max\{0, \bar{y}_n\}$, $\bar{y}_n := n^{-1}\sum_{t=1}^{n} y_t$. If θ_0 is an interior point of Θ, i.e. $\theta_0 > 0$, then $n^{1/2}(\hat{\theta} - \theta_0) \overset{w}{\to} \xi$, $\xi \sim N(0, 1)$. In contrast, if θ_0 is on the boundary of Θ, i.e. $\theta_0 = 0$, the asymptotic distribution of $\hat{\theta}$ is

$$n^{1/2}(\hat{\theta} - \theta_0) = n^{1/2}\hat{\theta} \overset{w}{\to} \ell := \max\{0, \xi\} \tag{3.2.1}$$

again with $\xi \sim N(0, 1)$.

The first takeaway of this section is the fact that the location of a parameter on the boundary of the parameter space may induce limiting bootstrap randomness of a kind that invalidates bootstrap inference. To see this, consider in the context of the location model a standard Gaussian parametric bootstrap based on the bootstrap sample

$$y_t^* = \hat{\theta} + \varepsilon_t^*,$$

where the ε_t^*'s are i.i.d.$N(0, 1)$ independent of the original data. The bootstrap counterpart of $l_n(\theta)$ is $l_n^*(\theta) := -\frac{1}{2}\sum_{t=1}^{n}(y_t^* - \theta)^2$, and the usual bootstrap QMLE is $\hat{\theta}^* := \arg\max_{\theta\in\Theta} l_n^*(\theta) = \max\{0, \bar{y}_n^*\}$, $\bar{y}_n^* := \hat{\theta} + \bar{\varepsilon}_n^*$, $\bar{\varepsilon}_n^* := n^{-1}\sum_{t=1}^{n}\varepsilon_t^*$. Conditionally on the original sample, $\hat{\theta}^*$'s exact distribution is

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) = n^{1/2}\max\{-\hat{\theta}, \bar{\varepsilon}_n^*\} \sim \max\{-n^{1/2}\hat{\theta}, \xi^*\}|\hat{\theta},\ \xi^*|\hat{\theta} \sim \xi \sim N(0, 1), \tag{3.2.2}$$

with associated conditional cdf given by

$$P^*(n^{1/2}(\hat{\theta}^* - \hat{\theta}) \leq x) = \Phi(x)\,\mathbb{I}_{\{x \geq -n^{1/2}\hat{\theta}\}},\ x \in \mathbb{R}. \tag{3.2.3}$$

Now, when θ_0 is an interior point of Θ, $-n^{1/2}\hat{\theta}$ diverges to $-\infty$ in probability and the distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ given the data converges weakly in probability to the non-random distribution of ξ^*; the bootstrap therefore mimics the $N(0, 1)$ asymptotic distribution of the original statistic, the bootstrap distributional approximation is consistent and bootstrap inference is valid in the sense of (3.1.1), with $intC = (0,1)$. Conversely, when θ_0 is on the boundary of the parameter space, the cdf in (3.2.3) converges weakly in $D(\mathbb{R})$ to the random cdf $\Phi(x)\mathbb{I}_{\{x \geq -\ell\}}$. In terms of weak convergence in distribution,

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_w \ell^* | \ell,\ \ell^* := \max\{-\ell, \xi^*\}, \tag{3.2.4}$$

where ℓ is distributed as in (3.2.1) and is independent of ξ^*. The limit distribution in (3.2.4) is random, since its cdf is a stochastic process depending on the conditioning random variable ℓ. Thus, it is distinct from the limit distribution in (3.2.1), which is unconditional and hence characterized by a non-random cdf. Because the bootstrap limit distribution is random, the bootstrap approximation is not consistent for the limit in (3.2.1).

As we shall see in Section 3.4, limiting bootstrap randomness could be of two kinds: 'benign', thus not compromising the validity of bootstrap inference in the sense of (3.1.1), or 'malignant', thus invalidating bootstrap inference. In this example, a bootstrap test employing a bootstrap statistic $\tau_n^* := \phi(n^{1/2}(\hat{\theta}^* - \hat{\theta}))$ as the analogue of a statistic $\tau_n := \phi(n^{1/2}\hat{\theta})$, where ϕ is a real function, may not be valid in the sense of (3.1.1) under the null hypothesis $\mathsf{H}_0 : \theta_0 = 0$ even if the function ϕ is continuous, thus implying 'malignant' randomness.

To get some further insight into the source of limiting bootstrap randomness, which will be exploited in the next sections, it is useful to notice that the asymptotic distributions in (3.2.1) and (3.2.2) can be written as

$$\begin{aligned} \ell &= \max\{0, \xi\} = \arg\min_{\lambda \in \Lambda} |\lambda - \xi|,\ \Lambda := \{\lambda \in \mathbb{R} : \lambda \geq 0\} \\ \ell^* | \ell &= \max\{-\ell, \xi^*\} | \ell = \left(\arg\min_{\lambda \in \Lambda(\ell)} |\lambda - \xi^*|\right) | \ell,\ \Lambda(\ell) := \{\lambda \in \mathbb{R} : \lambda \geq -\ell\} = \Lambda - \ell, \end{aligned}$$

respectively. Hence, bootstrap randomness, and the implied bootstrap invalidity, can be attributed to the fact that in the bootstrap world the limit constraint set for the objective function $|\lambda - \xi^*|$ is the *random* half line $\Lambda(\ell)$ rather than the original fixed half line $\Lambda = \Lambda(0)$. That is, the chosen bootstrap scheme shifts the constraint set by the random variable $-\ell$, which is non-zero with probability $1/2$.

The second takeaway of this section is the fact that bootstrap validity could be restored by offsetting properly the previous shift of the limit constraint set. Specifically, this requires an ad hoc construction of a bootstrap parameter space intended to approximate well the mutual position of the true parameter value and the boundary of the original parameter space.

Consider a bootstrap scheme where the boundary of the bootstrap parameter space Θ^* is chosen in a data-driven way such that the mutual position of θ_0 and the boundary of Θ is well approximated irrespective of whether θ_0 belongs to $\partial\Theta$ or not. To this aim, introduce the half line $\Theta^* := \{\theta : \theta \geq g^*(\hat{\theta})\}$, where $g^*(\theta) := \theta - |\theta|^{1+\kappa}$, $\kappa > 0$, and the associated $\hat{\theta}^* := \arg\max_{\theta \in \Theta^*} l_n^*(\theta) = \max\{g^*(\hat{\theta}), \bar{y}_n^*\}$. The bootstrap QMLE statistic is then given by $n^{1/2}(\hat{\theta}^* - \hat{\theta}) = n^{1/2} \max\{g^*(\hat{\theta}) - \hat{\theta}, \bar{\varepsilon}_n^*\}$. Conditionally on the data, it is distributed as $\max\{n^{1/2}(g^*(\hat{\theta}) - \hat{\theta}), \xi^*\}|\hat{\theta}$, with $\xi^*|\hat{\theta} \sim N(0,1)$. If $\theta_0 = 0$, it then follows that $n^{1/2}(g^*(\hat{\theta}) - \hat{\theta}) = -n^{1/2}\hat{\theta}^{1+\kappa} \xrightarrow{p} 0$, and the bootstrap statistic conditionally on the data converges weakly in probability to ℓ of (3.2.1). Conversely, if $\theta_0 > 0$ then $n^{1/2}(g^*(\hat{\theta}) - \hat{\theta}) = -n^{1/2}\hat{\theta}^{1+\kappa} \xrightarrow{p} -\infty$ and the bootstrap statistic conditionally on the data converges weakly in probability to the $N(0,1)$ distribution. In both cases, the bootstrap mimics the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta_0)$ and bootstrap validity in the sense of (3.1.1) can be seen to be successfully restored.

Remark. In the location model, an appropriate choice of Θ^* simultaneously restores bootstrap validity and removes all the randomness from the limit bootstrap distribution. In the predictive regression framework we shall conclude that, in order to achieve bootstrap validity, it is essential to remove only the portion of limiting bootstrap randomness that is due to the location of the parameter vector on the boundary of the parameter space. As no other sources of limiting bootstrap randomness exist in the context of the location model, in this section the previous conclusion simplifies to eliminating all the limiting bootstrap randomness. □

Before moving on to predictive regressions, we notice that when a test of $\mathsf{H}_0 : \theta_0 = 0$ against $\mathsf{H}_1 : \theta_0 > 0$ is performed, employing $\tau_n^* := n^{1/2}(\hat{\theta}^* - \hat{\theta})$ as the bootstrap analogue of $\tau_n := n^{1/2}\hat{\theta}$, the standard parametric bootstrap with $\Theta^* = \Theta$ is valid in the sense of (3.1.1); see also Andrews (2000). Specifically, the bootstrap test rejects H_0 when the bootstrap p-value $\tilde{p}_n^* = 1 - p_n^*$ is small, with the following convergence satisfied under the null hypothesis:

$$p_n^* = P^*(\tau_n^* \leq \tau_n) = P^*(\xi^* \leq \tau_n) \xrightarrow{w} \Phi(\ell).$$

A similar convergence is satisfied by the p-value $\tilde{p}_n = 1 - p_n$ of the asymptotic test, with

$$p_n = P(\ell \leq u)|_{u=\tau_n} = \tfrac{1}{2}\mathbb{I}_{\{\tau_n = 0\}} + \Phi(\tau_n)\mathbb{I}_{\{\tau_n > 0\}} = \Phi(\tau_n) \xrightarrow{w} \Phi(\ell).$$

As ℓ is distributed like $\Phi^{-1}(U)\mathbb{I}_{\{U>1/2\}}$, $U \sim U_{[0,1]}$, it follows that $\Phi(\ell)$ is distributed like $\Phi(\Phi^{-1}(U)\mathbb{I}_{\{U>1/2\}})$. As a result, both the bootstrap and the asymptotic test are correctly sized for nominal levels below 1/2. This phenomenon, whose extensions to predictive regression are discussed in Section 3.5.1, does not generalize to hypotheses where one-sided tests are not appropriate or straightforward. Therefore, a remedy is necessary for the inference-invalidating limiting bootstrap randomness induced by the location of a parameter on the boundary.

3.3 The predictive regression setup

Consider the following predictive regression in a triangular array setup:

$$y_t = \theta_1 + \theta_2 x_{n,t-1} + \varepsilon_t, \qquad (t = 1, ..., n;\ n = 1, 2, ...), \tag{3.3.1}$$

where ε_t is a martingale difference sequence [mds] and $x_{n,t}$ is a non-stationary posited predicting variable satisfying the following assumption; see, e.g. Müller and Watson (2008) for references to primitive conditions.

Let $z_{n,t} := n^{-1/2} \sum_{s=1}^{t} \varepsilon_s$. Then:

(a) $\{\varepsilon_t\}$ is an mds w.r.t. some filtration to which $(x_{n,t}, z_{n,t})$ is adapted, with $E\varepsilon_t^2 = \omega_{zz} \in (0, \infty)$.

(b) a law of large numbers holds as $n \to \infty$:

$$\sum_{t=1}^{n} \begin{pmatrix} \Delta x_{n,t} \\ \Delta z_{n,t} \end{pmatrix} \begin{pmatrix} \Delta x_{n,t} & \Delta z_{n,t} \end{pmatrix} \xrightarrow{p} \Omega := \begin{pmatrix} \omega_{xx} & \omega_{xz} \\ \omega_{xz} & \omega_{zz} \end{pmatrix} > 0.$$

(c) an invariance principle holds in D_2 as $n \to \infty$:

$$(x_{n,\lfloor n\cdot \rfloor}, z_{n,\lfloor n\cdot \rfloor})' \xrightarrow{w} (X, Z)' \sim BM(0, \Omega),$$

a bivariate Brownian motion on $[0, 1]$.

Assumption 1 covers the specification $x_{n,t} = n^{-1/2} x_t$ for an $I(1)$ process x_t driven by an mds that could be contemporaneously correlated with ε_t.[4,5]

Assumption 1 implies that $\sum_{t=1}^{n} x_{n,t-1} \Delta z_{n,t} \xrightarrow{w} \int X dZ$, which need not have a mixed Gaussian distribution because X and Z need not be independent. Nevertheless, it holds that $\sum_{t=1}^{n} x_{n,t-1}(\Delta z_{n,t} - \omega_{xz}\omega_{xx}^{-1}\Delta x_{n,t}) \xrightarrow{w} \int X d(Z - \omega_{xz}\omega_{xx}^{-1} X)$, which is zero-mean mixed Gaussian with conditional variance $\sigma_e^2 \int X^2$, where $\sigma_e^2 := \omega_{zz} - \omega_{xz}^2 \omega_{xx}^{-1}$ is the variance of ε_t corrected for $\Delta x_{n,t}$. The bootstrap schemes discussed below all rely on the independence of the processes X and $Z - \omega_{xz}\omega_{xx}^{-1} X$.

Further, Assumption 1 imposes unconditional homoskedasticity for simplicity. As all the bootstrap schemes below are based on 'wild' bootstrap schemes, unconditional heteroskedasticity can be accommodated at only a notational cost.

The next assumption specifies the parameter space, say Θ, by means of a smooth inequality constraint.

The parameter space is $\Theta := \{\theta = (\theta_1, \theta_2)' \in \mathbb{R}^2 : g(\theta) \geq 0\}$, with non-empty boundary $\partial\Theta := \{\theta \in \mathbb{R}^2 : g(\theta) = 0\}$, where $g : \mathbb{R}^2 \to \mathbb{R}$ is continuously differentiable in

---

[4] As the bootstrap p-values discussed in the paper are invariant to rescaling of the regressor, the normalization of x_t by $n^{-1/2}$ has no practical implication. It is equivalent to specifying a local-to-zero regression coefficient, as is frequent in applications where y_t is a financial return and x_t is non-stationary.

[5] Results under two alternative stochastic specifications of $x_{n,t}$, as a near-unit root and as a stationary process, are given in the accompanying supplement, Section C.2.

some neighborhood of the true parameter value $\theta_0 := (\theta_{1,0}, \theta_{2,0})'$ with gradient $\frac{\partial}{\partial\theta'}g(\theta) \neq 0$ in that neighborhood.

In the following, $\dot{g}$ will denote the gradient of the function g evaluated at θ_0.

Assumption 2 generalizes the leading example of the parameter space $\Theta = \mathbb{R} \times [0, \infty)$ obtained by setting $g(\theta) = (0, 1)\theta = \theta_2$. The boundary of Θ then corresponds to the case $\theta_2 = 0$ of no predictability of y_t by $x_{n,t-1}$ whereas the interior of Θ corresponds to the case of sign-restricted predictability.

Interest is in bootstrap inference on a null hypothesis H_0 identifying a set of parameter values that has a non-empty intersection with the boundary of the parameter space. In particular, we consider the following mutual positions of the boundary, the parameter set identified by H_0 and the true value θ_0:

- G_1. H_0 is the hypothesis that θ_0 belongs to the boundary: $\mathsf{H}_0 : g(\theta_0) = 0$;
- G_2. H_0 is a simple null hypothesis on the boundary: $\mathsf{H}_0 : \theta_0 = \bar{\theta}$, $g(\bar{\theta}) = 0$;
- G_3. $\mathsf{H}_0 : h(\theta_0) = 0$, where $\{\theta \in \mathbb{R}^2 : h(\theta) = 0\}$ is not a subset of the boundary $\partial\Theta$, but meets $\partial\Theta$ at a singleton set.

For example, let again $g(\theta) = \theta_2$, such that the parameter space is $\mathbb{R} \times [0, \infty)$ with boundary $\partial\Theta = \mathbb{R} \times \{0\}$. Then the hypothesis of no predictability $\mathsf{H}_0 : \theta_{2,0} = 0$ falls under G_1. The hypothesis $\mathsf{H}_0 : \theta_0 = \bar{\theta} = (0, 0)'$ that y_t is unpredictable with zero mean falls under G_2. Finally, the hypothesis $\mathsf{H}_0 : (1, 1)\theta_0 = \theta_{1,0} + \theta_{2,0} = 0$ falls under G_3 by setting $h(\theta) := (1, 1)\theta$; in this case, the intersection point of the boundary and the parameter set identified by H_0 is $(0, 0)'$ which might, but need not, be the true value under H_0.

3.3.1 Asymptotic distributions

Let $\hat{\theta}$ be the OLS estimator of $(\theta_1, \theta_2)'$ in the equation

$$y_t = \theta_1 + \theta_2 x_{n,t-1} + \delta\Delta x_{n,t} + e_t \tag{3.3.2}$$

subject to the constraint $\hat{\theta} \in \Theta$, i.e. $g(\hat{\theta}) \geq 0$, and where the role of the regressor $\Delta x_{n,t}$ is to ensure that the residuals are asymptotically uncorrelated with the innovations driving $x_{n,t}$, a convenient prerequisite for the bootstrap implementations. The existence, with probability approaching one, of a measurable minimizer of the residual sum of squares (3.3.2) over the set Θ can be established in a similar but simpler way than that of its bootstrap counterpart in our detailed proof of Theorem 3.4.1. Moreover, any two such minimizers are first-order asymptotically equivalent, explaining our usage of 'the' associated with the constrained OLS estimator. Specifically, any such minimizer $\hat{\theta}$ satisfies $n^{1/2}(\hat{\theta} - \theta_0) \stackrel{w}{\to} \ell(\theta_0)$, with $\ell(\theta_0)$ depending on the position of θ_0 relative to the boundary $\partial\Theta$. Thus, $\ell(\theta_0) = \tilde{\ell} := M^{-1/2}\xi$ if $\theta_0 \in int\Theta := \Theta \setminus \partial\Theta$, where $M := \int \tilde{X}\tilde{X}'$, $\tilde{X} := (1, X)'$, $\xi \sim N(0, \sigma_e^2 I_2)$ is independent of X, and $\sigma_e^2 > 0$ is the variance of ε_t corrected for $\Delta x_{n,t}$, whereas

$$n^{1/2}(\hat{\theta}-\theta_0) \xrightarrow{w} \ell(\theta_0) = \ell := \arg\min_{\lambda\in\Lambda} ||\lambda - M^{-1/2}\xi||_M,\ \Lambda := \{\lambda \in \mathbb{R}^2 : \dot{g}'\lambda \geq 0\} \quad (3.3.3)$$

if $g(\theta_0) = 0$, with $||x||_M := (x'Mx)^{1/2}$ for $x \in \mathbb{R}^2$; see Section 12 in the working paper version of Andrews (1999) or the proof of Theorem 3.4.1 for the bootstrap counterpart.

The previous asymptotic result is sufficient in order to see that the possibility of having θ_0 at the boundary of the parameter space Θ induces a dichotomy in the limit distribution of $n^{1/2}(\hat{\theta}-\theta_0)$ similar to the dichotomy established in the introductory location-model example. Replicating the constraint set in the limit distribution by means of a bootstrap scheme will be our main concern in what follows.

3.3.2 Standard bootstrap invalidity

Consider first a fixed-regressor wild bootstrap sample generated as

$$y_t^* = \hat{\theta}_1 + \hat{\theta}_2 x_{n,t-1} + \varepsilon_t^*, \quad (3.3.4)$$

where $\varepsilon_t^* = \hat{e}_t w_t^*$, $t = 1, ..., n$, with $\hat{e}_t$ the residuals of (3.3.2) and w_t^* i.i.d. $N(0,1)$, independent of the original data.[6] Then the distribution of $n^{1/2}(\hat{\theta}-\theta_0)$ could be tentatively approximated by the distribution of $n^{1/2}(\hat{\theta}^*-\hat{\theta})$ conditional on the original data, where $\hat{\theta}^*$ is obtained by regressing y_t^* on $(1, x_{n,t-1})'$ under the constraint $\hat{\theta}^* \in \Theta^* = \Theta$, i.e., $g(\hat{\theta}^*) \geq 0$ as for the original estimator; see Andrews (2000)[7].

To motivate the analysis in the next section, it is useful to anticipate some asymptotic properties of $\hat{\theta}^*$ which obtain by specializing Theorem 3.4.1 below to the fixed-regressor wild bootstrap scheme. For $\theta_0 \in int\Theta$, it turns out that the bootstrap distribution converges to a conditional version of the limit distribution of $n^{1/2}(\hat{\theta}-\theta_0)$ found earlier:

$$n^{1/2}(\hat{\theta}^*-\hat{\theta}) = n^{1/2}(\tilde{\theta}^*-\hat{\theta}) + o_p(1) \xrightarrow{w^*}_w \tilde{\ell}|M, \quad (3.3.5)$$

where $\tilde{\theta}^*$ denotes the unconstrained OLS estimator from the bootstrap sample. The limit bootstrap distribution is, therefore, random. The vehicle of limiting bootstrap randomness is the random matrix M, such that limiting bootstrap randomness is fully attributable to the stochastic properties of the regressor. Due to the fact that the bootstrap replicates a conditional version of the limit distribution of the original estimator $\hat{\theta}$, bootstrap inference is not invalidated. Rigorous statements in this sense will be provided in Corollary 3.4.1.

[6]The conclusions do not change if another zero-mean unit-variance distribution with a finite fourth moment is used instead of the standard Gaussian distribution.

[7]Note that the term $\Delta x_{n,t}$ is no longer necessary because $x_{n,t-1}$ and ε_t^* are independent conditionally on the data.

On the other hand, if $\theta_0 \in \partial\Theta$ the bootstrap statistic converges as follows:

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_w \ell^*|(M,\ell) \tag{3.3.6}$$
$$\ell^* := \arg\min_{\lambda\in\Lambda_\ell^*} ||\lambda - M^{-1/2}\xi^*||_M,\ \Lambda_\ell^* := \{\lambda \in \mathbb{R}^2 : \dot{g}'\lambda \geq -\dot{g}'\ell\}$$

where $\xi^* \sim N(0, \sigma_e^2 I_2)$ is independent of (M, ℓ). In contrast with the case $\theta_0 \in int\Theta$ and additionally to the random matrix M, in (3.3.6) also the random vector ℓ appears as a vehicle of limiting bootstrap randomness. Moreover, the limit in (3.3.6) is not a conditional version of the limit of $n^{1/2}(\hat{\theta} - \theta_0)$, inasmuch as Λ_ℓ^* in (3.3.6) is a random half-plane, rather than the original admissible set Λ of (3.3.3). The kind of limiting bootstrap randomness introduced by ℓ is similar to the one established in the introductory location model and, in general, it invalidates bootstrap inference. The reason for the discrepancy between Λ and Λ_ℓ^* is that the parameter space of the standard fixed-regressor wild bootstrap does not approximate well the original mutual position of the true value θ_0 and the boundary, unless $g(\hat{\theta}) = 0$. Other, non-standard bootstrap schemes may be designed in order to provide better approximations, at least under the null hypothesis. Under these schemes the possible boundary position of θ_0 is no longer a vehicle of limiting bootstrap randomness, while the role of the random matrix M in the limit bootstrap distribution is maintained. This topic is analyzed in the next section.

3.4 Asymptotically valid bootstrap schemes

In order to unify the discussion of several bootstrap schemes for inference on H_0 under the three cases G_1, G_2 and G_3, consider a bootstrap sample generated as in (3.3.4) and, more generally than before, a bootstrap OLS estimator $\hat{\theta}^*$ constrained to belong to a bootstrap parameter space Θ^* satisfying the following assumption.

The bootstrap parameter space is $\Theta^* := \{\theta \in \mathbb{R}^2 : g(\theta) \geq g^*(\hat{\theta})\}$ for some function $g^* : \mathbb{R}^2 \to \mathbb{R}$ which is continuously differentiable in a neighborhood of θ_0 and satisfies $g^*(\theta) \leq g(\theta)$ for $\theta \in \Theta$.

The standard bootstrap considered in Section 3.3 obtains by setting $g^* = 0$, such that $\Theta^* = \Theta$, the original parameter space. Alternatively, setting $g^* = g$ restricts the bootstrap true value $\hat{\theta}$ to lie on the boundary of the bootstrap parameter space Θ^*.[8] Finally, setting $g^* = g - |g|^{1+\kappa}$ for some $\kappa > 0$ introduces a correction, in the spirit of an alternative to the standard bootstrap mentioned in Andrews (2000, p.403, Method two), Fang and Santos (2019, Example 2.1) and Cavaliere et al. (2022), where the bootstrap true value either shrinks to the boundary of the bootstrap parameter space at a proper rate or remains bounded away from this boundary, according to whether θ_0 belongs to the

[8] As $\hat{\theta} \xrightarrow{p} \theta_0$ and $\dot{g}(\theta_0) \neq 0$, it follows by continuity that $P(\dot{g}(\hat{\theta}) \neq 0) \to 1$, such that, with probability approaching one, $\hat{\theta}$ is not a stationary point of g. In particular, with probability approaching one, $\hat{\theta}$ is not a local minimizer of g, implying that $\hat{\theta} \in \partial\Theta^*$ under $\Theta^* = \{\theta \in R^2 : g(\theta) \geq g(\hat{\theta})\}$.

original boundary $\partial\Theta$ or not. Other choices of g^* with the same implication are discussed in Sections 3.5.2 and 3.6.3.

To formulate the next theorem, recall M and $\ell(\theta_0)$ introduced in Section 3.3.1, and let $\xi^*|(M,\ell(\theta_0)) \sim N(0,\sigma_e^2 I_2)$ as in Section 3.3.2. Let also $D_n = \{y_t, x_{n,t-1}\}_{t=1}^n$ denote the original data. Finally, call a convergence in distribution $Z_n \xrightarrow{w} Z$ and a weak convergence of random distributions $Z_n^*|D_n \xrightarrow{w}_w Z^*|Y$ joint, denoted as $(Z_n,(Z_n^*|D_n)) \xrightarrow{w}_w (Z,(Z^*|Y))$, if $(Z_n, E\{g(Z_n^*)|D_n\}) \xrightarrow{w} (Z, E\{g(Z)|Y\})$ for all continuous and bounded real functions g with matching domain.

Theorem 3.4.1 *Under a null hypothesis H_0 as in G_1–G_3 and under Assumptions 1-3, the bootstrap estimator $\hat\theta^*$ obtained by regressing y_t^* of (3.3.4) on $(1, x_{n,t-1})'$ under the constraint $\hat\theta^* \in \Theta^*$, satisfies*

$$(n^{1/2}(\hat\theta-\theta_0),(n^{1/2}(\hat\theta^*-\hat\theta)|D_n)) \xrightarrow{w}_w (\ell(\theta_0),(\ell^*(\theta_0)|(M,\ell(\theta_0))))\,,$$

where in the case $g^*(\theta_0) < g(\theta_0)$,

$$\ell^*(\theta_0) = \tilde\ell^* := M^{-1/2}\xi^* \text{ with } \tilde\ell^*|(M,\ell(\theta_0)) = \tilde\ell|M \tag{3.4.1}$$

in the sense of a.s. equality of conditional distributions, whereas in the case $g^*(\theta_0) = g(\theta_0)$,

$$\ell^*(\theta_0) = \ell^* := \arg\min_{\lambda\in\Lambda_\ell^*} ||\lambda - M^{-1/2}\xi^*||_M,\ \Lambda_\ell^* := \{\lambda\in\mathbb{R}^2 : \dot g'\lambda \geq (\dot g^* - \dot g)'\ell(\theta_0)\}. \tag{3.4.2}$$

The following conclusions could be drawn.

(i) Consider first configurations G_1 and G_2 under H_0, such that $g(\theta_0) = 0$. Consider the magnitude order, in probability, of the distance between the bootstrap 'true' value $\hat\theta$ and the bootstrap boundary $\partial\Theta^*$ as a measure of how precisely the bootstrap approximates the geometry of G_1 and G_2. As seen previously, the standard bootstrap corresponds to $g^* = 0$ and approximates the geometry up to an exact magnitude order of $n^{-1/2}$, resulting in a situation where the belonging of θ_0 to the boundary contributes to the randomness of the limit bootstrap distribution given by (3.3.6) and (3.4.2) via conditioning on the r.v. $\ell(\theta_0) = \ell$. Conversely, bootstrap schemes employing $g^*(\theta_0) = g(\theta_0)$ and $\dot g^* = \dot g$, such that the bootstrap boundary is tangent to the original boundary at θ_0, give rise to approximations of order $o_p(n^{-1/2})$ and all the randomness in the bootstrap limit is due to the properties of the stochastic regressor via the random variable M, as now $\ell^*|(M,\ell) = \ell|M$ in the sense of a.s. equality of random distributions; see (3.3.3) and (3.4.2). Moreover, for such schemes the bootstrap mimics a conditional version of the asymptotic distribution of the original estimator: $n^{1/2}(\hat\theta^* - \hat\theta) \xrightarrow{w^*}_w \ell|M$. Examples are the 'restricted' bootstrap based on $g^* = g$, which replicates the geometry of the original data under H_0

by putting $\hat{\theta}$ on the bootstrap boundary, and the choices $g^* = g - |g|^{1+\kappa}$ for some $\kappa > 0$. In general, the limit distribution of the resulting bootstrap estimator is random, with randomness depending on both the stochastic regressor and the position of θ_0 relative to the boundary.

(ii) Consider now the case in G_3, such that $g(\theta_0) = 0$ need not, but may hold under H_0. Among the bootstraps considered in (i), the standard one would fail to mimic a conditional version of the original limit distribution if $g(\theta_0) = 0$, while the 'restricted' one would fail if $g(\theta_0) > 0$. As an alternative, consider the bootstrap based on $g^* = g - |g|^{1+\kappa}$ for some $\kappa > 0$. If $\theta_0 \in \partial\Theta$, then this choice puts the bootstrap true value $\hat{\theta}$ at the asymptotically negligible distance of $o_p(n^{-1/2})$ from the bootstrap boundary, whereas if $\theta_0 \in int\Theta$, then $\hat{\theta}$ is bounded away from the bootstrap boundary, in probability. This guarantees bootstrap validity under some regularity conditions, see (ii) in Corollary 3.4.1 below.

In general, bootstrap validity in the sense of (3.1.1) can be evaluated through the following corollary of Theorem 3.4.1 above.

Corollary 3.4.1 *Under the assumptions of Theorem 3.4.1, a necessary and sufficient condition for the convergence*

$$(n^{1/2}(\hat{\theta} - \theta_0), (n^{1/2}(\hat{\theta}^* - \hat{\theta})|D_n)) \xrightarrow{w}_w (\ell(\theta_0), (\ell(\theta_0)|M)) \tag{3.4.3}$$

is that: (i) under G_1 and G_2, $g(\theta_0) = g^*(\theta_0)$ and $\dot{g} = \dot{g}^*$; (ii) under G_3, either $g(\theta_0) = g^*(\theta_0)$ and $\dot{g} = \dot{g}^*$, or $g(\theta_0) > \max\{0, g^*(\theta_0)\}$.

Moreover, under (3.4.3) the bootstrap is valid in the sense of (3.1.1) for any pair of statistics τ_n, τ_n^* such that, under H_0, $\tau_n = \phi(n^{1/2}(\hat{\theta} - \theta_0)) + o_p(1)$ and $\tau_n^* = \phi(n^{1/2}(\hat{\theta}^* - \hat{\theta})) + o_p(1)$ for some continuous real function ϕ such that $\phi(\ell(\theta_0))$ is well-defined a.s.

The class of functions $g^* = g - |g|^{1+\kappa}$ for $\kappa > 0$ satisfies both conditions (i) and (ii) of Corollary 3.4.1; hence, the ensuing bootstrap inference is valid under all of G_1-G_3. In contrast, the standard bootstrap violates condition (i) and, in general, is asymptotically invalid if $g(\theta_0) = 0$. An exception is when the discrepancy between the original and the bootstrap geometry is offset by the use of a test statistic that takes into account the geometric position of the null hypothesis in the original parameter space. Section 3.5.1 focuses on this setup.

Remark. The practical implications of Corollary 3.4.1 depend on the choice of the statistic τ_n and the respective function ϕ, which will typically be a linear $\phi(l) = l' \frac{\partial r}{\partial \theta'}(\theta_0)$ arising from the delta method, with $l \in \mathbb{R}^2$, $\frac{\partial r}{\partial \theta'}(\theta_0) \neq 0$. For instance, if $g(\theta_0) = 0$ and $\phi(\ell)$ depends on ℓ only through $\dot{g}'\ell = \max\{0, \dot{g}' M^{-1/2}\xi\}$, then the cdf of $\phi(\ell)$ will not be continuous. Still, the bootstrap will be valid in the sense of (3.1.1), meaning that the largest open subset of $[0,1]$ on which the bootstrap test is correctly sized as $n \to \infty$ coincides with the analogous set for the asymptotic test. This set will be smaller than $(0,1)$, however. An example is $\tau_n = n^{1/2}g(\hat{\theta})$, $\tau_n^* = n^{1/2}(g(\hat{\theta}^*) - g(\hat{\theta}))$ with $\phi(\ell) = \dot{g}'\ell$, corresponding to a right-sided test of $\mathsf{H}_0 : g(\theta_0) = 0$.

Remark. Bootstrap validity extends readily to statistics where $n^{1/2}(\hat{\theta}-\theta_0)$ is normalized by some $\hat{\Sigma} = \Sigma(M_n) + o_p(1)$ for a function $\Sigma : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ which is continuous on the set of positive definite matrices. Specifically, bootstrap validity holds if, under H_0, $\tau_n = \phi(n^{1/2}\hat{\Sigma}(\hat{\theta} - \theta_0)) + o_p(1)$ and $\tau_n^* = \phi(n^{1/2}\hat{\Sigma}(\hat{\theta}^* - \hat{\theta})) + o_p(1)$, where ϕ is a continuous real function such that $\phi(\Sigma(M)\ell(\theta_0))$ is a.s. well-defined. □

3.5 Discussion and extensions

In this section we address the following three issues: (i) the validity of one-sided bootstrap tests; (ii) a discussion of the bootstrap schemes from Corollary 3.4.1 within the paradigm of some previous works – specifically, Fang and Santos (2019) and Hong and Li (2020); and (iii) uniform bootstrap validity.

3.5.1 Validity of one-sided standard bootstrap tests

Under case G_1, consider testing $\mathsf{H}_0 : g(\theta_0) = 0$ against the alternative $H_1 : g(\theta_0) > 0$ using a one-sided test and the standard bootstrap, i.e., with $g^* = 0$. For a test statistic of the form $\tau_n := n^{1/2}g(\hat{\theta})$, a bootstrap counterpart is given by $\tau_n^* := n^{1/2}(g(\hat{\theta}^*) - g(\hat{\theta}))$ and the associated one-sided bootstrap test rejects for large values of the bootstrap p-value $p_n^* := P^*(\tau_n^* \le \tau_n)$; equivalently, for small values of $\tilde{p}_n^* := 1 - p_n^*$. As for $\hat{\theta}^*$, also τ_n^* is affected in the limit by extra randomness due to θ_0 being on the boundary. From (3.4.1), which reduces to (3.3.3) and (3.3.6), it follows by the delta method that

$$(\tau_n, (\tau_n^*|D_n)) \xrightarrow{w}_w (\dot{g}'\ell, (\dot{g}'\ell^*|(M,\ell))) = (\dot{g}'\ell, (\max\{-\dot{g}'\ell, \dot{g}'\tilde{\ell}^*\}|(M,\ell))), \tag{3.5.1}$$

with ℓ, ℓ^* and $\tilde{\ell}^*$ as previously defined. For τ_n^*, however, the randomness induced by conditioning on ℓ affects the sample paths of the associated random cdf on the negative half-line alone, because $\dot{g}'\ell \ge 0$, and is thus irrelevant for bootstrap tests with nominal size in $(0, \frac{1}{2})$. Put differently, the bootstrap p-values $\tilde{p}_n^*$ are asymptotically uniformly distributed below $\frac{1}{2}$. This follows rigorously from the next generalization of Theorem 3.1 in Cavaliere and Georgiev (2020), the proof being analogous, where conditions for bootstrap validity restricted to a subset of nominal testing levels are formulated.

Theorem 3.5.1 *Let there exist a random variable τ and a random element X, both defined on the same probability space, such that the support of τ_n is contained in a finite or infinite closed interval T, and $(\tau_n, F_n^*) \xrightarrow{w} (\tau, F)$ in $\mathbb{R} \times D(T)$ for $F_n^*(u) := P(\tau_n^* \le u|D_n)$ and $F(u) := P(\tau \le u|X)$, $u \in T$. If the possibly random cdf F is sample-path continuous* *on T, then the bootstrap p-value $p_n^* := F_n^*(\tau_n)$ satisfies*

$$P(p_n^* \le q) \to q$$

for q such that $q \in F(T)$ a.s.

By Theorem 3.5.1 with $T = [0, \infty)$, which corresponds to the support of τ_n and $\tau := \dot{g}'\ell$, it follows that the standard bootstrap applied to the one-sided statistic τ_n is asymptotically correctly sized for nominal test sizes in $(0, \frac{1}{2})$.

3.5.2 Fang and Santos (2019) and Hong and Li (2020)

In this section we put the geometric considerations of Section 3.4 in the perspective of Fang and Santos (2019), and of the numerical bootstrap of Hong and Li (2020). The discussion is often specialized to the case of an affine constraint.

Consider the constrained OLS estimator $\hat{\theta}$ of Section 3.3.1. Its limit distribution, see (3.3.3), is the distribution of $\ell(\theta_0) = \varphi_{\theta_0}(M^{-1/2}\xi)$ with

$$\varphi_{\theta_0}(u) = \begin{cases} u & \text{if } g(\theta_0) > 0 \\ \dot{g}_\perp(\dot{g}_\perp' M \dot{g}_\perp)^{-1}\dot{g}_\perp' M u + M^{-1}\dot{g}(\dot{g}' M^{-1}\dot{g})^{-1}\max\{0, \dot{g}'u\} & \text{if } g(\theta_0) = 0 \end{cases},$$

with $u \in \mathbb{R}^2$. By a projection identity, the expression in the second line of the previous display collapses to u whenever $\dot{g}'u \ge 0$. Note that the distribution of $M^{-1/2}\xi$ conditional on M can be estimated consistently by the distribution of the unconstrained bootstrap OLS estimator conditional on the data; that is,

$$n^{1/2}(\tilde{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_w M^{-1/2}\xi | M.$$

One can then ask what properties of an estimator $\hat{\varphi}_n$ of φ_{θ_0} are sufficient for $\hat{\varphi}_n(n^{1/2}(\tilde{\theta}^* - \hat{\theta})) \xrightarrow{w^*}_w \varphi_{\theta_0}(M^{-1/2}\xi)|M$ to hold. Fang and Santos (2019) address this question in the setup of deterministic transformations of non-random limit distributions, instead of the random transformation φ_{θ_0} of the random distribution $M^{-1/2}\xi|M$. Although not directly applicable here, Theorem 3.2 of Fang and Santos (2019) provides the key insight: there should be sufficient uniformity in the convergence of $\hat{\varphi}_n$ to φ_{θ_0}. Consider for instance

$$\begin{aligned} \hat{\varphi}_n(u) &= \hat{\dot{g}}_\perp(\hat{\dot{g}}_\perp' M_n \hat{\dot{g}}_\perp)^{-1}\hat{\dot{g}}_\perp' M_n u \\ &\quad + M_n^{-1}\hat{\dot{g}}(\hat{\dot{g}}' M_n^{-1}\hat{\dot{g}})^{-1}\max\{-n^{1/2}|g(\hat{\theta})|^{1+\kappa}, \hat{\dot{g}}'u\},\ u \in \mathbb{R}^2, \end{aligned} \tag{3.5.2}$$

where $\hat{\dot{g}} = \frac{\partial}{\partial\theta'}g(\hat{\theta})$, $M_n = n^{-1}\sum_{t=1}^n \tilde{x}_t\tilde{x}_t'$ with $\tilde{x}_t = (1, x_{n,t-1})'$, and $\kappa > 0$. Given that $M_n \xrightarrow{w} M$, $\hat{\dot{g}} \xrightarrow{p} \dot{g}$ and $n^{1/2}|g(\hat{\theta})|^{1+\kappa} \xrightarrow{p} \infty\mathbb{I}_{\{g(\theta_0)>0\}}$, it is easily checked that $\hat{\varphi}_n \xrightarrow{w} \varphi_{\theta_0}$ on $C_2(\mathbb{R}^2)$, and the convergence of $\hat{\varphi}_n$ is joint with that of $n^{1/2}(\hat{\theta} - \theta_0)$ and $n^{1/2}(\tilde{\theta}^* - \hat{\theta})$, the latter one given the data. These facts are sufficient to ensure that

$$(n^{1/2}(\hat{\theta} - \theta_0), (\hat{\varphi}_n(n^{1/2}(\tilde{\theta}^* - \hat{\theta}))|D_n)) \xrightarrow{w}_w (\ell(\theta_0), (\ell(\theta_0)|M))$$

on $\mathbb{R}^4$, essentially as a consequence of the continuous mapping theorem (CMT) and the continuity of the evaluation map from $C_2(\mathbb{R}^2) \times \mathbb{R}^2$ to $\mathbb{R}^2$. As the previous limit is the same as in Corollary 3.4.1, it follows that bootstrap inference based on the distribution of $\hat{\varphi}_n(n^{1/2}(\tilde{\theta}^* - \hat{\theta}))$ conditional on the data is valid. Moreover, for the valid bootstrap schemes obtained from Corollary 3.4.1 with $g^* = g - |g|^{1+\kappa}$, $\kappa > 0$, the bootstrap estimator $\hat{\theta}^*$ satisfies $n^{1/2}(\hat{\theta}^* - \hat{\theta}) = \hat{\varphi}_n(n^{1/2}(\tilde{\theta}^* - \hat{\theta}))$ for affine functions g. It can be concluded that $\hat{\varphi}_n$ of (3.5.2) implicitly performs the geometric approximation proposed in Section 3.4, and so does any other estimator of φ_{θ_0} that converges like $\hat{\varphi}_n$.

We now argue that such an estimator of φ_{θ_0} is embedded in the numerical bootstrap of Hong and Li (2020). This ensures the validity of the numerical bootstrap for the predictive regression of interest here, though at the cost of a slower consistency rate of the bootstrap estimator than in Corollary 3.4.1. Let $s_n \to \infty$ be a sequence such that $n^{-1/2}s_n \to 0$. Hong and Li (2020) propose in their eq. (4.9) a bootstrap estimator $\hat{\theta}^*_{nb}$ where the constraint set of our $\ell(\theta_0)$ (i.e., $\mathbb{R}^2$ if $\theta_0 \in \operatorname{int}\Theta$ and the half-plane Λ if $\theta_0 \in \partial\Theta$), would be estimated by $\Lambda^*_{nb} = \{\lambda \in \mathbb{R}^2 : g(\hat{\theta} + s_n^{-1}\lambda) \geq 0\}$, the implied bootstrap parameter space being $\Theta^*_{nb} = \hat{\theta} + s_n^{-1}\Lambda^*_{nb} = \Theta$. The bootstrap estimator itself, adapted to our setup, could be written as

$$\hat{\theta}^*_{nb} = \arg\min_{g(\theta) \geq 0} ||s_n(\theta - \hat{\theta}) - M_n^{-1/2}\xi_n^*||_{M_n},$$

where ξ_n^* is a bootstrap variable such that $\xi_n^* \xrightarrow{w^*}_p N(0, I_2)$; e.g., $\xi_n^* = n^{1/2}M_n^{1/2}(\tilde{\theta}^* - \hat{\theta})$. In the simple case of an affine g we find the explicit expression

$$s_n(\hat{\theta}^*_{nb} - \hat{\theta}) = \bar{\varphi}_n(M_n^{-1/2}\xi_n^*)$$

for $\bar{\varphi}_n$ defined similarly to $\hat{\varphi}_n$, with the only difference that in (3.5.2) the term $n^{1/2}|g(\hat{\theta})|^{1+\kappa}$ is replaced by $s_n g(\hat{\theta})$. As $s_n g(\hat{\theta}) \xrightarrow{p} \infty \mathbb{I}_{\{g(\theta_0)>0\}}$ similarly to $n^{1/2}|g(\hat{\theta})|^{1+\kappa}$, $\kappa > 0$, it follows that $\bar{\varphi}_n$ converges similarly to $\hat{\varphi}_n$. As a result,

$$(n^{1/2}(\hat{\theta} - \theta_0), (s_n(\hat{\theta}^*_{nb} - \hat{\theta})|D_n)) \xrightarrow{w}_w (\ell(\theta_0), (\ell(\theta_0)|M)),$$

ensuring the validity of the numerical bootstrap, though the consistency rate of $\hat{\theta}^*_{nb}$ is $s_n = o(n^{1/2})$ instead of $n^{1/2}$. In contrast, the rate of $n^{1/2}$ would be achieved by our proposed bootstrap estimator, with $n^{1/2}(\hat{\theta}^* - \hat{\theta}) = \bar{\varphi}_n(M_n^{-1/2}\xi_n^*)$, if $\Theta^* = \{\theta \in \mathbb{R}^2 : g(\theta) \geq g_n^*(\hat{\theta})\}$ with $g_n^* = g - n^{-1/2}s_n|g|$ is specified in Assumption 3.

3.5.3 Uniformity considerations

In agreement with Chatterjee and Lahiri (2011), Remark 3, the focus in this paper is on pointwise bootstrap validity. For situations where uniform bootstrap validity is of interest, our key takeaways are similar to the literature on non-random limiting bootstrap measures. First, for the null hypothesis G_1 that the true parameter value lies on the boundary of the parameter space, the pointwise-valid bootstrap schemes outlined in Corollary 3.4.1 display asymptotic rejection probabilities matching the local power of the bootstrap test whenever the true parameter value varies along a sequence that is local to the boundary at the $n^{-1/2}$ rate. This fact is associated with rejection frequencies above the nominal test size along local-to-the-boundary parameter sequences (cf. Fang and Santos, 2019, Remark 3.6). Second, if conservative bootstrap inference along such parameter sequences is desired, it can be achieved for hypotheses G_1–G_3 by adapting the approach of Doko Tchatoka and Wang (2021), and Cavaliere et al. (2024), at the cost of a potential decrease in power.

To illustrate these points, consider a sequence of true parameter values $\theta_n = \theta_0 + n^{-1/2}\vartheta$ such that $g(\theta_0) = 0$ and $\dot{g}'\vartheta = c > 0$ with $g(\theta_n) = n^{-1/2}c + o(n^{-1/2})$. Moreover, let

$$\ell(\vartheta, c) := \vartheta + \arg\min_{\lambda \in \Lambda^c} ||\lambda - M^{-1/2}\xi||_M, \ \Lambda^c = \{\lambda \in \mathbb{R}^2 : \dot{g}'\lambda + c \geq 0\}, \tag{3.5.3}$$

and $\ell(0, 0) = \ell$ of eq. (3.3.3). Then, the joint convergence result

$$(n^{1/2}(\hat{\theta} - \theta_0), (n^{1/2}(\hat{\theta}^* - \hat{\theta})|D_n)) \xrightarrow{w}_w (\ell(\vartheta, c), (\ell(0, 0)|M)) \tag{3.5.4}$$

holds for the bootstrap schemes satisfying conditions (i) and (ii) of Corollary 3.4.1. For a function $r : \mathbb{R}^2 \to \mathbb{R}$ which is continuously differentiable close to θ_0, consider the statistics $\tau_n = n^{1/2}r(\hat{\theta})$ and $\tau_n^* = n^{1/2}(r(\hat{\theta}^*) - r(\hat{\theta}))$, and distinguish among the extreme possibilities $\dot{r} = \alpha\dot{g}$ with $\alpha > 0$, and $\dot{r} = \alpha\dot{g}_\perp$ with $\alpha \neq 0$, where $\dot{r} = \frac{\partial r}{\partial \theta'}(\theta_0)$. The former possibility arises in testing the null hypothesis that θ_0 lies on the boundary (e.g., with $r = g$), whereas the latter one arises when the null is orthogonal to the boundary (e.g., with $r(\theta) = \theta_1$, $\mathsf{H}_0 : \theta_1 = 0$ and $\Omega = \mathbb{R} \times [0, \infty)$). If $\dot{r} = \dot{g}$ and, without loss of generality, $\alpha = 1$, the delta method yields

$$(\tau_n, (\tau_n^*|D_n)) \xrightarrow{w}_w (\max\{0, \dot{g}'M^{-1/2}\xi + c\}, (\max\{0, \dot{g}'M^{-1/2}\xi\}|M)).$$

With $\gamma_M := (\dot{g}'M^{-1}\dot{g})^{-1/2}$, it follows that

$$\begin{aligned} P^*(\tau_n^* \leq \tau_n) \xrightarrow{w} \pi(c; M, \xi) &:= \tfrac{1}{2}\mathbb{I}_{\{\dot{g}'M^{-1/2}\xi + c < 0\}} + \Phi(\gamma_M(\dot{g}'M^{-1/2}\xi + c))\mathbb{I}_{\{\dot{g}'M^{-1/2}\xi + c \geq 0\}} \\ &> \pi(0; M, \xi), \end{aligned}$$

where $\pi(0; M, \xi) \overset{d}{=} \frac{1}{2}\mathbb{I}_{\{U<0.5\}} + U\mathbb{I}_{\{U \geq 0.5\}}$, $U \sim U_{[0,1]}$, represents the limit distribution of the bootstrap p-value under the null. The inequality above implies that bootstrap tests rejecting for large bootstrap p-values will exhibit rejection frequencies above the nominal test size.

On the other hand, if $\dot{r} = \alpha\dot{g}_\perp$, it holds that

$$(\tau_n, (\tau_n^*|D_n)) \xrightarrow{w}_w (\dot{r}'\gamma_M^\perp M^{1/2}\xi + \dot{r}'\gamma_M^\perp M\vartheta, (\dot{r}'\gamma_M^\perp M^{1/2}\xi|M)),$$

with $\gamma_M^\perp := \dot{g}_\perp\dot{g}_\perp'(\dot{g}_\perp' M\dot{g}_\perp)^{-1}$, such that the boundary is asymptotically irrelevant. Bootstrap tests of the null that $r(\theta_0) = 0$ could be conservative or liberal according to the sign of $\dot{r}'\dot{g}_\perp \dot{g}_\perp' M\vartheta$. Similar considerations apply whenever $\dot{r}'\dot{g}_\perp \neq 0$.

For situations where liberal tests are not desirable, a possible remedy is suggested next. It involves a continuum of boundaries for the bootstrap parameter space and its implementation requires a discretization of that continuum.

Let $\tilde{\theta}$ be the unrestricted OLS estimator of θ in regression (3.3.2). For every $s \in I_n := [-|g(\tilde{\theta})|^{1-\mu}, g(\hat{\theta})]$, let $\hat{\theta}_s^*$ be the bootstrap estimator over the parameter space $\Theta_s^* := \{\theta \in \mathbb{R}^2 : g(\theta) \geq s - g(\hat{\theta})^{1+\kappa}\}$, where $\mu \in (0,1)$ and $\kappa > 0$ are fixed. For a continuously differentiable function r, let $p_n^*(s)$ be the p-value of a test based on $\tau_n = n^{1/2} r(\hat{\theta})$ and $\tau_n^* = n^{1/2}(r(\hat{\theta}^*) - r(\hat{\theta}))$. Then

$$\limsup_{n\to\infty} P(\sup_{s\in I_n} p_n^*(s) \leq q) \leq q$$

for all $q \in intC$, where C is the set from display (3.1.1) for the benchmark asymptotic test based on the unfeasible statistic $n^{1/2}(r(\hat{\theta}) - r(\theta_n))$ and the simple null hypothesis that θ_n is the true parameter value. This conservative generalization of the validity property (3.1.1) holds irrespective of the values of the drift parameter c. Specifically, the role of $-|g(\tilde{\theta})|^{1-\mu}$ in the definition of I_n is to guarantee that $g(\hat{\theta}) - cn^{-1/2} \in I_n$ with probability approaching one. Conservative size control then follows from the fact that $\hat{\theta}_s^*$ with $s = g(\hat{\theta}) - cn^{-1/2}$ satisfies

$$(n^{1/2}(\hat{\theta} - \theta_n), (n^{1/2}(\hat{\theta}_s^* - \hat{\theta})|D_n)) \xrightarrow{w} (\ell(0,c), (\ell(0,c)|M))\,;$$

see eqs. (3.5.3)–(3.5.4).

3.6 Numerical results and choice of the tuning parameters

In this section we analyze the finite sample performance of the proposed bootstrap methodology by means of numerical simulations. The purpose is twofold: first, to investigate the practical advantage of our methodology over standard bootstrap methods; second, to provide some practical guidance on how to choose the functions g^* and the tuning parameter κ in the definition of the bootstrap parameter space. Simulations are based on setup $\mathcal{G}_3$ of Section 3.3, as it covers the general case of a true parameter value that could, but need not, lie on the boundary of the parameter space under the null hypothesis. This section is organized as follows. In Section 3.6.1 we describe the data generating processes, the null hypotheses and the adopted bootstrap schemes. In Section 3.6.2 we discuss the performance of the tests both under the null and under local alternatives. Section 3.6.3 deals with the choice of g^* and κ. Additional numerical results are provided in the accompanying supplement, Section C.3.

3.6.1 Monte Carlo design

We consider the same data generating process (DGP) as in (3.3.1), where $x_{n,t} = n^{-1/2}x_t$, $x_t := \sum_{i=1}^{t} \varepsilon_{x,i}$, $\varepsilon_{x,t} \sim iid\ N(0,1)$, with the following specifications of ε_t:

1. $\varepsilon_t \sim iid\ N(0,1)$;
2. $\varepsilon_t = \sigma_t \nu_t$, where $\sigma_t^2 = 0.7 + 0.3\varepsilon_{t-1}^2$ and $\nu_t \sim iid\ N(0,1)$;
3. $\varepsilon_t = \sqrt{0.5}\varepsilon_{x,t} + \sqrt{0.5}\eta_t$, where $\eta_t \sim iid\ N(0,1)$.

In each case, $\{\varepsilon_{x,t}\}$ is independent of, respectively, $\{\varepsilon_t\}$, $\{\nu_t\}$ and $\{\eta_t\}$. In Case 1, the regression errors are independent and Gaussian, while in Case 2 they exhibit ARCH-type conditional heteroskedasticity. Case 3 allows for correlation between ε_t and the regressor's innovation $\varepsilon_{x,t}$.

The parameter space is specified as $\Theta := \{\theta \in \mathbb{R}^2 : g(\theta) \geq 0\}$ where $g(\theta) = \theta_2$. That is, $\Theta := \mathbb{R} \times [0, \infty)$ – such that its boundary is given by $\partial\Theta = \mathbb{R} \times \{0\}$. For all parameter values, we test the null hypothesis $\mathsf{H}_0 : h(\theta_0) = 0$, with $h(\theta) = \theta_1 + \theta_2$, against the two-sided alternative $h(\theta_0) \neq 0$. To do so, we employ the test statistics $\tau_n = \phi(\sqrt{n}h(\hat{\theta}))$ and $\tau_n^* = \phi(\sqrt{n}(h(\hat{\theta}^*) - h(\hat{\theta})))$, where $\phi(x) = x^2$, while $\hat{\theta}$ and $\hat{\theta}^*$ denote the original and bootstrap constrained LS estimators, respectively. In order to analyze size control and power of the proposed tests, we consider both empirical rejection probabilities [ERPs] under the null and under local alternatives. For tests performed under the null, we consider three different choices of the true value θ_0, one located on $\partial\Theta$ and two located on $\Theta\backslash\partial\Theta$; specifically, $\theta_0 \in \{(0,0)', (-0.75, 0.75)', (-1.5, 1.5)'\}$. Under H_1, we employ a local alternative of the form $\theta_0 = a_0 n^{-1/2}$, $a_0 \in \mathbb{R}^2$, such that $h(\theta_0) \neq 0$ unless $a = (0,0)'$.

Tests are based on p-values obtained using a 'standard' – i.e., with $\Theta^* = \Theta$ – fixed-regressor Gaussian wild bootstrap and the proposed 'corrected' bootstrap scheme. For the latter, the bootstrap parameter space is set to $\Theta^* = \mathbb{R} \times [g^*(\hat{\theta}_2), \infty)$, where the function g^* satisfies the assumptions of Corollary 3.4.1, see also Section 3.6.3. In order to assess the impact of the tuning parameter κ, we consider a grid of possible values for κ. Numerical results are based on 50,000 Monte Carlo simulations, each involving $B = 999$ bootstrap repetitions. Sample sizes are set to $n \in \{100, 200, 400, 800, 1600\}$.

3.6.2 Empirical rejection probabilities

We now discuss the ERPs of the bootstrap tests. Specifically, the Monte Carlo results in Table 1 and 2 refer to the case in which data are generated under the null and under local alternatives, respectively. The proposed modified bootstrap parameter space is based on the function $g^* = g - |g|^{1+\kappa}$ for several values of $\kappa > 0$.

Table 1 shows that the 'standard' bootstrap scheme typically under-rejects the true null hypothesis when the parameter lies on the boundary of the parameter space Θ whereas, as expected, its ERPs are closer to the nominal level when θ_0 is in the interior of Θ. Our proposed bootstrap performs similarly to the 'standard' bootstrap for very small values of κ, with the impact of the correction becoming more relevant as κ increases. If the parameter is on the boundary of the parameter space ($\theta_0 \in \partial\Theta$), our proposed bootstrap scheme gives rise to smaller absolute size distortions than the 'standard' bootstrap, for all the considered DGPs and all values of κ. When $\theta_0 \in \text{int}\Theta$, we observe very little variability in the ERPs across the different bootstrap methods, at least for reasonably small values of κ.

Table 2 reports the ERPs of the tests when data are generated under local alternatives $\theta_0 = a_0 n^{-1/2}$, $a_0 \in \{(-3,0)', (3,0)', (5,0)'\}$, such that the true parameter values lie on the boundary of the parameter space. Results show that both bootstrap schemes have power under local alternatives, with the 'corrected' bootstrap generally showing higher ERPs than the 'standard' bootstrap, in line with the results obtained under the null. Finally, we notice that the sign of the deviations from the null hypothesis matters, with positive deviations showing higher ERPs. This finding can be explained by the fact that the limit distribution of $n^{1/2}(h(\hat{\theta}) - h(\theta_0))$ is asymmetric when θ_0 lie on the boundary of Θ. Results about local alternatives such that θ_0 are $n^{-1/2}$-local to the boundary are substantially similar and are reported in Section S.2 of the supplement.

3.6.3 Choice of g^* and κ

We now consider the practical issue of choosing the function g^* and the tuning parameter κ used to construct the modified bootstrap parameter space Θ^*.

Regarding g^*, in Section 4 we discussed the functions $g^*_{(1)} := g - |g|^{1+\kappa}$, $\kappa > 0$, which satisfy the assumptions of Corollary 3.4.1 and were employed in the simulations so far, whereas in Section 3.5.2 we considered also $g^*_{(2)} := g - n^{-\kappa}|g|$, $\kappa \in (0, 1/2)$, corresponding to $s_n = n^{1/2-\kappa}$ in the concluding paragraph of Section 3.5.2. Numerical results in Table 1 and 2 and in the accompanying Supplement, Section S.2, show that both choices of g^* deliver good test performance, both under the null and under local alternatives. The most salient difference between $g^*_{(1)}$ and $g^*_{(2)}$ is that tests based on $g^*_{(1)}$ tend to be more robust to the choice of κ when $g(\theta_0) \geq 1$.

Concerning the choice of the tuning parameter κ, we focus on $g^* = g^*_{(1)}$. Preliminary considerations point at a possible trade-off between the cases of a boundary and an interior location of the true parameter θ_0. Thus, for $\theta_0 \in \partial\Theta$, larger values of κ accelerate the convergence of $g(\hat{\theta})^{1+\kappa}$ to zero, which can be expected to favor bootstrap performance as the bootstrap true value $\hat{\theta}$ is put at a smaller distance from the bootstrap boundary. On the other hand, if $\theta_0 \in \text{int}(\Theta)$ and $g(\theta_0) \in (0,1)$, in small samples large values of κ may put $\hat{\theta}$ too close to the bootstrap boundary, yielding inferior bootstrap performance.

Our Monte Carlo study indeed confirms that small values of κ are preferable when $\theta_0 \in \text{int}(\Theta)$ and $g(\theta_0) \in (0,1)$; however, it also shows that the proposed correction quickly provides satisfactory size control for small values of κ even when $\theta_0 \in \partial\Theta$. Finally, we notice that when $\theta_0 \in \text{int}(\Theta)$ and $g(\theta_0) \geq 1$ the choice of κ has negligible impact on the

ERPs. Overall, our numerical analysis suggests that choices of κ close to 0.5 provide quite satisfactory size control across all the considered scenarios.

Remark. The above guideline about the choice of κ is based on numerical evidence; it delivers a reasonable simple choice which can be easily implemented. It is not optimal in any sense, and indeed alternative methods could be employed to obtain data-driven choices of κ. For instance, the unrestricted parameter estimates could be used to assess how far the true parameter value θ_0 is from the boundary of the parameter space, and then calibrate the choice of κ accordingly. This approach would be in the spirit of Romano, Shaikh and Wolf (2014), who suggest to improve the power of tests of moment inequalities by introducing a first step, where a confidence region for the moments is constructed using their unrestricted estimates. Although this approach may improve the finite sample properties of our tests, it would require a preliminary choice of further tuning parameters, such as β in Romano et al. (2014), hence introducing an extra layer of complexity. □

3.7 Conclusions

In this paper we analyzed the problem of bootstrap hypotheses tests on the parameters (α, β) of a predictive regression $y_t = \alpha + \beta x_{t-1} + \varepsilon_t$, generalizable to higher dimensions, when the parameter space is defined by means of a smooth constraint $g(\alpha, \beta) \geq 0$ and the true parameter vector under the null hypothesis may lie on the boundary of the parameter space. In the framework of constrained parameter estimation, implementation of the bootstrap is not straightforward, as the presence of a parameter on the boundary of the parameter space makes the bootstrap measure random in the limit.

We discussed possible solutions to this inference problem. Specifically, we presented some modifications of standard bootstrap schemes where the bootstrap parameter space is shifted by a data-dependent function, thus allowing us to regain control over the boundary as a source of limiting bootstrap randomness. We also proved validity of the associated bootstrap inference in the cases where the posited predicting variable is I(1).

Our contribution is novel in the framework of predictive regression, in that the existing literature has not analyzed the bootstrap in contexts combining non-stationarity of the posited predictor with a priori knowledge about the possible form of predictability, represented by a restricted parameter space. The value of our work is to provide valid bootstrap implementations in this setting.

Chapter 4

When did the Phillips Curve Become Flat? A Time-varying Estimate of Structural Parameters

(written with Claudio Lissona and Antonio Marsi)

4.1 Introduction

The disconnect between the observed fluctuations in unemployment and inflation in the last decades is a well-known empirical fact in the macroeconomic literature (Stock and Watson, 2020; Ball and Mazumder, 2020; Bobeica et al., 2021). As shown by Stock and Watson (2020), a set of simple OLS regressions of core inflation on a measure on the unemployment gap, for the US, reveals a declining pattern. The estimated OLS coefficient is -0.48 for the 1960-83 sample, -0.26 for the 1984-99 sample and -0.03 (and not statistically different from zero) for the 2000-2019 sample. However, these are changes in reduced form correlations, which can be the result of different underlying structural changes. As shown extensively in Del Negro et al. (2020); McLeay and Tenreyro (2020); Bergholt et al. (2023), among others, a decreasing correlation between inflation and unemployment can be caused by at least three different phenomena: (1) a decreasing value of the structural coefficient relating changes in economic slack and inflation; (2) an increase in the strength of the reaction of the central bank to business cycle shocks; (3) an increase in the proportion of aggregate fluctuations driven by supply-type shocks as compared to demand-type shocks. Other structural explanations are based on some departure from rational expectations models (see e.g. Coibion et al., 2018) or stems from the involved network structure of the economy (see e.g. Rubbo, 2022). Disentangling among these different explanations essentially boils down to being able to identify the structural Phillips Curve (PC), rather than a reduced form version (Mavroeidis et al., 2014). Furthermore, different explanations would lead to opposite monetary policy considerations. In fact, a decrease in the structural slope of the PC (i.e. explanation (1) above), would imply that the central bank's ability to control inflation is impaired, since it is based on the idea that by setting the interest rate the central bank can affect economic slack and thus inflation via the PC. On the other hand, it is clear that explanation (2) above is much more comforting for central bankers. As we explain in the related literature section, evidence is still mixed and conclusions crucially depend on the empirical strategy used to identify the structural equation. The PC can be formulated in many different ways. Here we refer to its hybrid New- Keynesian formulation, following Barnichon and Mesters (2020a) and many others:

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda x_t + \eta_t \tag{4.1.1}$$

where π_t is some measure of inflation and γ_b, γ_f are the parameters governing the stickiness of inflation and the role played by expectations about future inflation, respectively. x_t is some measure of economic slack and the unemployment gap can be conveniently used for this purpose (Galì et al., 2011; Galì, 2011). Notice that π_t should be intended as a deviation from the long run trend, as we will discuss later. η_t is a cost-push (supply-type) shock.

Over recent decades, an extensive portion of the literature has focused on PC estimation, given its major implications, in primis for monetary policy (Mavroeidis et al., 2014). However, apart from few exceptions such as Inoue et al. (2022) and Galí and Gambetti (2019), the literature is still missing a precise estimate of the evolution over time of λ, where the coefficient is allowed to be fully time-varying. In this paper we contribute to filling this gap, using an innovative approach. We believe that tracking the value of λ over time is extremely useful, as it enables not only to show whether λ has in fact decreased or not, but also when exactly did the structural change take place. Understanding when the change happened can help to shed light on the ultimate forces behind the phenomenon.

To do this, we first model the dynamics of aggregate macro variables, in a time-varying fashion. We do this by using a non-parametric specification of a stochastic time-varying VAR model, following Giraitis et al. (2018). Details of this model are presented in Section 2. This choice has several useful advantages compared to competing alternatives. Simple splitting of the sample, as in Del Negro et al. (2020), fails to provide evidence on the exact timing of the occurrence of any structural change and is subject to the arbitrary choice of the splitting date. Rolling window analysis would reduce sample size in an inefficient way. Competing time-varying VAR models, such as the bayesian TVC-VAR (Primiceri, 2005; Del Negro and Primiceri, 2015) - besides being much more demanding from a computational viewpoint - impose ex-ante parametric restrictions on the process driving the evolution of time-varying coefficients.

Notice also that the approach we use enables us to circumvent the usual problem of cleaning for the long run trend in both π_t and x_t. The stochastic time-varying VAR model embeds the estimation of a *random attractor* which can be interpreted as an estimate of the long-run inflation trend and the natural unemployment rate (see the details in the next section). This permits to avoid using methods such as the Hodrick-Prescott filter to estimate long run components of the time series.

We identify an aggregate demand shock in the time-varying VAR model, by using the Excess Bond Premium (EBP) by Gilchrist and Zakrajšek (2012) and ordering it last in a recursive-ordering (Cholesky) identification strategy. This is the same identification strategy used in Del Negro et al. (2020) and it is a convenient choice in our setup for several reasons that we will detail later. At this stage we are therefore able to estimate a set of time-varying impulse response functions to the identified aggregate demand shock. Furthermore, an alternative identification strategy, based on sign-restrictions, is explored in Appendix D.2.

Lewis and Mertens (2022) show that a structural macro equation like (4.1.1) can be conveniently and consistently estimated by computing the impulse responses of the variables in the equation (namely π_t and x_t) to a properly identified structural shock (for the PC, a demand shock orthogonal to the supply shock η_t), and by then regressing impulse responses one on each other. Armed with our time-varying impulse responses, we can in fact get a time-varying estimate of the NKPC by applying this idea. This methodology, called the SP-IV estimator in Lewis and Mertens (2022), is an extremely smart way to solve many issues practically faced by researchers when trying to estimate structural equations like the PC. In a nutshell, the endogeneity of $E_t\pi_{t+1}$ and x_t is the main challenge here. Let us consider equation (4.1.1) and forget for simplicity about past and future inflation components and focus only of x_t. Equation (4.1.1) is only one of a large set of structural equations driving macroeconomic fluctuations, hence $E(\eta_t x_t) \neq 0$, and (4.1.1) cannot be estimated by OLS. The traditional solution is to use past values of macro variables as external instruments, under the assumption that $E(x_{t-1}\eta_t) = 0$. However, it is enough for η_t to exhibit some persistence for the assumption not to be valid anymore. Using further lags could solve the problem, but at the strong cost of reducing instrument's relevance. Barnichon and Mesters (2020a) proposes then to use a sequence of past and present (properly identified) shocks, e.g. a sequence of past monetary surprises, to instrument x_t. This approach however requires the strongest assumption of contemporaneous and lag exogeneity of the instrument. The SP-IV method by Lewis and Mertens (2022) conveniently addresses all these issues and we refer to their paper for all the details. In our case, the most convenient feature of SP-IV is that the IRFs to the demand shock can be identified using any valid identification scheme, hence it enables us to employ the Cholesky ordering strategy of Del Negro et al. (2020).

Using this approach we are able to trace out the evolution over time of the parameters of a time-varying NKPC, which can be framed within a DSGE model with exogenous variation in the structural parameters (Canova and Sala, 2009; Castelnuovo, 2012; Galvão et al., 2016). We apply our methodology to a sample of monthly US macroeconomic variables and estimate a strong reduction of the slope of the NKPC. Specifically, the estimated slope goes sharply to zero already in the 80's and remain zero from 1990 onward. At the same time, we show that the γ_f coefficient, measuring the importance of inflation expectations, grows over time. Notice that a combination of high γ_f and low λ is, from a theoretical point of view, a dangerous situation where the central bank's ability to control inflation is impaired while expectations are at the same time quite important, hence any news shock to expectations can significantly influence the inflation rate, possibly leading to explosive behaviors.

By comparing the estimated IRFs of the interest rate to the demand shock over time, we rule out the hypothesis of a stronger reaction of the central bank as a driver of the flattening of the PC (i.e. explanation (2) above). If anything, the estimated response of the interest rate seems to become more muted over time.

To show that our approach provides valid estimates of the structural parameters of interest, we conduct a simulation study where we analyze the performance of our methodology by generating data from a time-varying-parameters version of a simple NK model à la Gali and Gertler (1999).

The remainder of the paper proceeds as follows. First, we briefly review the related literature. In Section 4.2 we explain in details the methodology used. In Section 4.3 we present the data used and how we specify the time-varying VAR model. In Section 4.4 results from the baseline specification are presented. In Section **??** the procedure we use to conduct inference on the parameters of interest is discussed. In Section 4.5 the evidence from the simulation analysis is presented, while Section 4.6 provides some concluding remarks. Appendix D.1 shows results when using variables at quarterly frequency. Appendix D.2 presents results using the sign-restrictions identification strategy. Appendix D.3 shows the results for the Euro Area.

4.1.1 Related literature

The empirical literature on the PC is huge and a exhaustive survey is not the aim of this paper (see Mavroeidis et al., 2014, for a survey of older contributions. We limit ourselves to list some relevant papers that study the evolution over time of PC coefficients.

Barnichon and Mesters (2020a), for instance, estimates λ for the US, from two overlapping samples: for the 1969-2007 sample, they estimate a value of -0.42, while focusing on the 1990-2007 sample yields an estimate of -0.24, pointing to a reduction of the structural slope of the NKPC, λ. However, the external instruments used for identification in the two samples are different: the Romer and Romer (2004) narrative measure of monetary policy shocks in the first sample, high-frequency monetary surprises à la Kuttner

(2001) in the second sample

Del Negro et al. (2020) estimates a bayesian-VAR on US data, splitting the sample in two parts: 1973-1989 and 1990-2019 and identify a structural shock associated to the EBP, as we did in this paper. The document a significantly smaller reaction of the inflation rate to the shock in the second sample, even when conditioning for the path of unemployment. There are few studies in the literature trying to estimate a fully time-varying version of the PC. A notable example is Giraitis et al. (2021), who use a nonparametric time-varying stochastic IV estimator to directly estimate a time-varying version of (1). Differently from us, they estimate the PC coefficients by using a set of lags of unemployment and inflation as external instruments. This identification strategy presents all the issues briefly mentioned above and discussed extensively in Barnichon and Mesters (2020a) and Lewis and Mertens (2022). Their analysis on US data points a decline in the slope of the PC, which is estimated to be significantly different from zero only in the early 80's. Another study very much related to this paper is Galí and Gambetti (2019). They analyze the evolution of the wage-NKPC over time for the US, documenting a decreasing slope. However, we differ from Galí and Gambetti (2019), along several dimensions. First, to model time-varying dynamics of the macroeconomic variables, they use the methodology in Del Negro and Primiceri (2015): a bayesian TVC-VAR. This model assumes time-varying coefficients to follow a random-walk, while the methodology we use is consistent with a much wider class of dynamic models, as explained in section 2. Second, they focus on the wage Phillips Curve while we focus on the price one. Third, their structural identification strategy is based on a combination of long-run and short-run sign-restrictions applied to the set of structural shock,[1] while we rely on the Excess Bond Premium to identify the demand shock, as in Del Negro et al. (2020). Fourth, they use their estimates of the time-varying impulse response functions to construct time-varying "Phillips Multipliers" (Barnichon and Mesters, 2020b). In short, it consists in computing the trade-off between inflation and unemployment, at different horizons, by taking the ratio of the cumulative impulse reponses of the two variables. However, compared to the Lewis and Mertens's (2022) SP-IV method, this methodology has important shortcomings: (i) the estimated trade-off between inflation and unemployment is horizon specific; (ii) it does not allow to control for (and include in the estimated time-varying coefficients) the lagged and forward inflation components of the Phillips Curve.

Berger et al. (2016) and Fu (2020) estimate a series of reduced form models for the PC, allowing for time variation in the coefficients. While Berger et al. (2016) find evidence of stability of reduced form PC parameters over time, Fu (2020) concludes that it is important to account for time variation. We differ from these studies as we focus on

[1]In their framework, they have the following set of structural shocks: technology, demand (non-monetary), monetary policy, price markup, wage markup.

the estimation of structural parameters, rather than reduced form ones.

4.2 Methodology

In this paper, the estimation of time-varying NKPC parameters is achieved by combining two advanced econometric techniques that have been recently proposed. We use a nonparametric stochastic time-varying VAR model (Giraitis et al., 2014, 2018) to obtain time-varying impulse response functions to a structural aggregate demand shock. We then exploit the novel methodology proposed by Lewis and Mertens (2022) to estimate the structural parameters of the NKPC over the sample considered.

4.2.1 Estimation of the structural NKPC

Given the structural NKPC:

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda x_t + \eta_t \tag{4.2.1}$$

Barnichon and Mesters (2020a) and Lewis and Mertens (2022) show that its parameters can be estimated using a "regression in impulse responses" approach. Denote by $Irf_h(\pi), Irf_h(x)$ the impulse responses of variables π_t and x_t, after h periods, to a structural shock v_t orthogonal to η_t:

$$\begin{aligned} Irf_h(\pi) &= E(\pi_{t+h}/\mathcal{I}_{t-1}, v_t = 1) - E(\pi_{t+h}/\mathcal{I}_{t-1}) \\ Irf_h(\pi) &= E(x_{t+h}/\mathcal{I}_{t-1}, v_t = 1) - E(x_{t+h}/\mathcal{I}_{t-1}) \end{aligned} \tag{4.2.2}$$

NKPC parameters can then be estimated by:

$$\left[\widehat{\lambda}\ \widehat{\gamma}_b\ \widehat{\gamma}_f\right]' = \left(\Theta_X'\Theta_X\right)^{-1}\Theta_X'\Theta_Y \tag{4.2.3}$$

where Θ_X, Θ_Y collect the estimated response of variables across horizons:

$$\Theta_Y = \begin{bmatrix} \widehat{Irf}_0(\pi) \\ \widehat{Irf}_1(\pi) \\ \vdots \\ \widehat{Irf}_H(\pi) \end{bmatrix} \qquad \Theta_X = \begin{bmatrix} \widehat{Irf}_0(x) & 0 & \widehat{Irf}_1(\pi) \\ \widehat{Irf}_1(x) & \widehat{Irf}_0(\pi) & \widehat{Irf}_2(\pi) \\ \vdots & \vdots & \vdots \\ \widehat{Irf}_H(x) & \widehat{Irf}_{H-1}(\pi) & \widehat{Irf}_{H+1}(\pi) \end{bmatrix} \tag{4.2.4}$$

Lewis and Mertens (2022) show that the necessary impulse responses can be obtained with any valid forecasting model and identification scheme, hence a VAR can be used to estimate Θ_X, Θ_Y, as long as π_t and x_t are part of the vector of endogenous variables.

4.2.2 Time-varying NKPC

The NKPC should be seen as one of the equations of small- and large-scale structural macroeconomic models, namely DSGE models. Introducing time-variation of the parameters in a DSGE model is not straightforward (see, for a discussion, Kapetanios et al., 2019). One way to model time-variation of the parameters is to define a stochastic process for the parameters (or a subset of them) and assume that agents know how parameters evolve over time and take this into account when forming expectations; this assumption is made for instance in Justiniano and Primiceri (2008). This approach makes the time-varying DSGE model much more complicated and requires to augment the set of shocks with the innovations to parameters' value, which implies that only a subset of the parameters can be let vary over time. We use a different approach, following Canova and Sala (2009); Castelnuovo (2012); Galvão et al. (2016), and assume that agents take parameter variation as exogenous when forming expectations about the future. This assumption keeps the model tractable and simple. At each point in time agents take parameters' value as given and think they will stay at the same value forever. In the next period agents learn about the changes in parameters and adjust their equations, but do not use this change to forecast future changes in the parameters. A similar result would be obtained by assuming that parameters follow a random walk process, hence the best guess about their future value would still be the current value.

Under these assumptions about time-variation of the "deep" parameters, we can rewrite the NKPC, in a time-varying fashion:

$$\pi_t = \gamma_{b,t}\pi_{t-1} + \gamma_{f,t}E_t\pi_{t+1} + \lambda_t x_t + \eta_t \tag{4.2.5}$$

We can use the methodology outlined above to estimate $\lambda_t, \gamma_{b,t}$ and $\gamma_{f,t}$ over time, as long as we are equipped with time-varying impulse response functions:

$$\begin{aligned} Irf_{t,h}(\pi) &= E\left(\pi_{t+h}/\mathcal{I}_{t-1}, v_t = 1\right) - E\left(\pi_{t+h}/\mathcal{I}_{t-1}\right) \\ Irf_{t,h}(x) &= E\left(x_{t+h}/\mathcal{I}_{t-1}, v_t = 1\right) - E\left(x_{t+h}/\mathcal{I}_{t-1}\right) \end{aligned} \tag{4.2.6}$$

$$\left[\widehat{\lambda}_t \; \widehat{\gamma}_{b,t} \; \widehat{\gamma}_{f,t}\right]' = \left(\Theta_{X,t}'\Theta_{X,t}\right)^{-1}\Theta_{X,t}'\Theta_{Y,t} \tag{4.2.7}$$

$$\Theta_{Y,t} = \begin{bmatrix} \widehat{Irf}_{0,t}(\pi) \\ \widehat{Irf}_{1,t}(\pi) \\ \vdots \\ \widehat{Irf}_{H,t}(\pi) \end{bmatrix} \qquad \Theta_{X,t} = \begin{bmatrix} \widehat{Irf}_{0,t}(x) & 0 & \widehat{Irf}_{1,t}(\pi) \\ \widehat{Irf}_{1,t}(x) & \widehat{Irf}_{0,t}(\pi) & \widehat{Irf}_{2,t}(\pi) \\ \vdots & \vdots & \vdots \\ \widehat{Irf}_{H,t}(x) & \widehat{Irf}_{H-1,t}(\pi) & \widehat{Irf}_{H+1,t}(\pi) \end{bmatrix} \tag{4.2.8}$$

where H is the maximum horizon included in the regression, chose a priori.

4.2.3 Time-varying VAR

To estimate time-varying impulse responses we follow Giraitis et al. (2018) and specify a time-varying-parameters VAR, given by:

$$y_t - \mu_t = \Psi_t\left(y_{t-1} - \mu_{t-1}\right) + u_t, \quad t = 1, ..., T \tag{4.2.9}$$

where y_t and u_t a $n-$dimensional vectors and μ_t is a persistent stochastic attractor. The VAR innovations u_t have zero-mean and conditional variance-covariance $\Sigma_t = E\left(u_t u_t' | I_{t-1}\right)$ and $E\left(u_t u_s'\right) = 0$ for $t \neq s$. Innovations u_t are related to n structural shocks by $u_t = P_t \varepsilon_t$, where P_t is the Cholesky decomposition of Σ_t, so that $P_t P_t' = \Sigma_t$, and $E\left(\varepsilon_t \varepsilon_t'\right) = I_n$.

Notice that we do not assume any parametric structure for the time variation of both the autoregressive parameters Ψ_t and the volatility process Σ_t as we only require either them to be bounded in probability or their expectation to be bounded. In particular, by letting $||\cdot||_{sp}$ be the spectral norm, we assume $\sup_{s:|s-t|\leq k} ||\Psi_t - \Psi_s||_{sp}^2 = O_p\left(k/t\right)$ and $\sup_{s:|s-t|\leq k} \mathbb{E}||P_t - P_s||_{sp}^2 = O\left(k/t\right)$. Finally, unit and explosive roots are bounded away by assuming that $||\Psi_t||_{sp} < 1, \forall t \in [0, T]$.

The dynamic structural impulse response functions at time t, for horizon h are given by:

$$\Phi_{t,h} = \Psi_t^h P_t \tag{4.2.10}$$

We estimate the above model, following **?**, by a nonparametric kernel estimator. Let[2] $K(x)$ be a bounded, nonnegative kernel function with piecewise-bounded derivative such that $\int K(x)dx = 1$. We estimate μ_t as follows:

$$\hat{\mu}_t = K_t^{-1} \sum_{j=1}^{T} k_{tj} y_j \tag{4.2.11}$$

where $k_{tj} = K\left(\left(t-j\right)/H_\Psi\right)$, $K_t = \sum_{j=1}^{T} k_{tj}$ and H_Ψ is a bandwith parameter such that $H_\Psi \to \infty$. The formula above basically corresponds to a weighted average of y_t, with weights defined by the kernel function and the bandwidth parameter.

Define $\hat{y}_t = y_t - \hat{\mu}_t$, we estimate Ψ_t as follows:

$$\hat{\Psi}_t = \left(\sum_{j=1}^{T} k_{tj} \hat{y}_{t,j} \hat{y}_{t,j-1}'\right)\left(\sum_{j=1}^{T} k_{tj} \hat{y}_{t,j-1} \hat{y}_{t,j-1}'\right)^{-1} \tag{4.2.12}$$

We then estimate Σ_t, based on the variance-covariance of the residuals $\hat{u}_j = \hat{y}_{t,j} - \hat{\Psi}_t \hat{y}_{t,j-1}$, as follows:

$$\hat{\Sigma}_t = L_t^{-1} \sum_{j=1}^{T} l_{ij} \hat{u}_j \hat{u}_j' \tag{4.2.13}$$

where $\quad l_{tj} := K\left(\frac{t-j}{H_h}\right), \quad L_t := \sum_{j=1}^{T} l_{tj}$ and H_h is another bandwidth parameter satisfying $H_h \to \infty$.

Finally we estimate time-varying impulse response functions by:

$$\hat{\Phi}_{t,h} = \hat{\Psi}_t^h \hat{P}_t \tag{4.2.14}$$

[2] Notice that these conditions are satisfied by many of the most widely used kernel functions, including kernels with both finite and infinite support. Examples of valid kernels in this setting are the *flat kernel*, the Epanechnikov's kernel and the Gaussian kernel, see **?**.

where $\hat{P}_t$ is the Cholesky factorization of $\hat{\Sigma}_t$.

Following the methodology in Giraitis et al. (2018), in our empirical study we set $H_\Psi = T^{0.6}, H_h = T^{0.4}$ for the Gaussian kernel function:

$$K(x) = \left(\frac{1}{\sqrt{2\pi}}\right) e^{-x^2/2} \tag{4.2.15}$$

4.3 Data and VAR specification

We apply our methodology on a sample of US macroeconomic variables observed at monthly frequency. In Appendix D.1 we repeat the analysis on variables at quarterly frequency. The time-varying-parameters VAR (TVP-VAR) includes a set of seven endogenous variables: (1) the unemployment rate; (2) core inflation, measured by the annualized monthly growth rate of the core CPI; (3) the 12-month growth rate of the PPI for all commodities; (4) the log change of industrial production; (5) the 10-year treasury rate; (6) the 3-month T-bill rate and (7) the Excess Bond Premium, by Gilchrist and Zakrajšek (2012). The set of variables mimics the one used by Lewis and Mertens (2022). The monthly data sample covers the period from 1973M1 to 2019M12, for a total of 564 observations.

We order the Excess Bond Premium last and identify the structural shocks by a Cholesky decomposition. The EBP shock is our structural shock of interest, i.e. the demand-type shock. Let us recall that the Phillips Curve is the aggregate-supply curve in the New-Keynesian models, hence a demand shock should be used to properly identify it. The EBP is ordered last, which means that it is assumed have zero-instantaneous effect on all the other variables in the VAR, while it is allowed to adjust instantaneously to all the other structural shocks. This is the same identification strategy employed by Del Negro et al. (2020). The idea is that the EBP is a measure of credit/financial frictions, hence an exogenous increase in its value should be interpreted as an exogenous worsening of financial conditions which should ultimately affect the economy via a demand channel (e.g. by a reduction of the access to bank loans for firms).

Our choice for an identification strategy based on a simple Cholesky ordering is justified by the need of keeping it as simple as possible, as other more involved strategies may not be the best choice in the complex fully time-varying model we propose. For instance, it is typical in the literature to use external proxy variables such as monetary surprises Barnichon and Mesters (2020a) or other proxies such as the unemployment shock of Angeletos et al. (2020), used e.g. by Lewis and Mertens (2022). However, using an external instrument approach in our time-varying context is not trivial as the relevance condition, which is hardly met in simpler context with fixed-over-time parameters, would need to be met at each point in time. Nonetheless, in Appendix D.2 we explore a different identification strategy based on sign restrictions which provide additional interest insights.

The number of lags to include in the VARs is a tricky subject in our time-varying context and a rule to follow in this regard is still missing in the literature. We decide to include 2 lags in the monthly VAR, for several reasons. First, since we employ a time-varying dynamic model, with a time-varying attractor, we certainly need a smaller number of lags to achieve a good fit of the data, compared to a VAR with fixed coefficients. Furthermore, the BIC criterion applied to a fixed-coefficients VAR indicates one single lag as the best choice for the quarterly data sample and two lags for the monthly one. The BIC criterion is known to be a very parsimonious indication, but it can be used as a rough rule of thumb in our context.

4.4 Results, monthly VAR

Figure 4.1 shows the estimated time-varying IRFs to the EBP shock, over the sample analyzed. The responses have been normalized, for each t, such that the instantaneous response of the EBP to the EBP shock is equal to 1. We observe a stronger but shorter-lasting response of the unemployment rate at the beginning of the sample. In fact around 1980 the unemployment response reverts in the log run, turning into the negative territory. The estimated peak unemployment response drops significantly in the 90's and then increase again after 2000, then steadily decreasing again toward the end of the sample.

The estimated response of core inflation looks strong and negative only until mid 80's, while it is very muted further on. This provides a first evidence of a flattening of the PC, consistent with the results found by Del Negro et al. (2020).

It is interesting to look at the estimated responses of the interest rate to the EBP shock as they provide a partial answer to the question of what caused the documented flattening. As explained above, a flatter PC would be observed if the central bank becomes more committed to stabilize inflation and thus reacts more strongly to demand shocks hitting the economy. In that case, we would observe a muted response of inflation combined with an increase response of the interest rate. Figure 4.1 shows that the estimated response of the interest rate has, if anything, become softer along the years considered. More in details, the decrease in the interest rate observed after an EBP shock is very strong in the 1973-1983 period and in the 1995-2005 period. Outside of these two windows the reaction of interest rate looks more modest.

To give a more meaningful answer to the main question of this paper, namely the value of the NKPC parameters over time, we apply the methodology described in the previous section. We specify a NKPC as in Lewis and Mertens (2022), but in a time-varying-parameters fashion:

$$\pi_t^{1q} = (1 - \gamma_{f,t})\, \pi_{t-3}^{1y} + \gamma_{f,t} \pi_{t+12}^{1y} + \lambda_t x_t + \eta_t \tag{4.4.1}$$

where π_t^{1q} is the annualized percent change in the Core CPI from a quarter ago in month t, π_t^{1y} is the percent change in the Core CPI over the preceding year in month t, and x_t is the unemployment rate in month t. The variable definitions in terms of quarterly and annual inflation rates follows Barnichon and Mesters (2020a). Since our IRFs are estimated by using the annualized monthly inflation rate, IRFs have been adjusted accordingly to implement the regression shown in equation (4.2.7). Furthermore, as in Lewis and Mertens (2022), we include only horizons $h = 0, 3, 6, \ldots, 33$ to construct the impulse response vectors. Results are shown in Figure 4.2, withe the estimated values of $\lambda_t, \gamma_{b,t}$ and $\gamma_{f,t}$ over time.

To construct confidence intervals for the estimated parameters over time, we use the fixed-regressor wild bootstrap Zanelli (2023).

Regarding λ_t we observe a very clear pattern: the NKPC becomes almost totally flat already at the end of the 80's and remain flat from that moment onward. This result is new in the literature and shows that practically all the structural change from a steep to a flat PC took place in the 80's.

As for the $\gamma_{b,t}$ and $\gamma_{f,t}$ coefficients, a clear pattern emerges: past inflation matters less and less as we move forward along the sample, while the coefficient associated to expected inflation increases, accordingly (recall that we impose $\gamma_{b,t} + \gamma_{f,t} = 1$). This suggests a growing importance over time of expectations, together with a decreasing role of past inflation. Notice that this result is in line with the findings of arnichon and Mesters (2020a), despite their identification strategy is based on monetary policy shocks rather than on the EBP. While the observed trend is that of a shift from past to future inflation, a big and temporary jump in the $\gamma_{b,t}$ coefficient is estimated after 2000, which brings its value back to the 1973 one.

4.5 Monte Carlo simulations

To verify the performance of our methodology in the estimation of a time-varying NKPC, we perform some simulation exercises. We simulate data from a simple textbook New Keynesian (NK) model (Woodford, 2003; Gali, 2008) which features indexation to past inflation as in Gali and Gertler (1999). We follow Bergholt et al. (2023) for the exposition and specification of the most part of the model.

We let some of the parameters to be time-varying. Introducing time-variation of the parameters in a DSGE model is not straightforward see Kapetanios et al., 2019, for a discussion). One way to model time-variation of the parameters is to define a stochastic process for the parameters (or a subset of them) and assume that agents know how parameters evolve over time and take this into account when forming expectations; this assumption is made for instance in Justiniano and Primiceri (2008). This approach makes the time-varying DSGE model much more complicated and requires to augment the set of shocks with the innovations to parameters' value, which implies that only a subset of the parameters can be let vary over time. We follow a different approach, following Canova and Sala (2009); Castelnuovo (2012); Galvão et al. (2016), and assume that agents take parameter variation as exogenous when forming expectations about the future. This

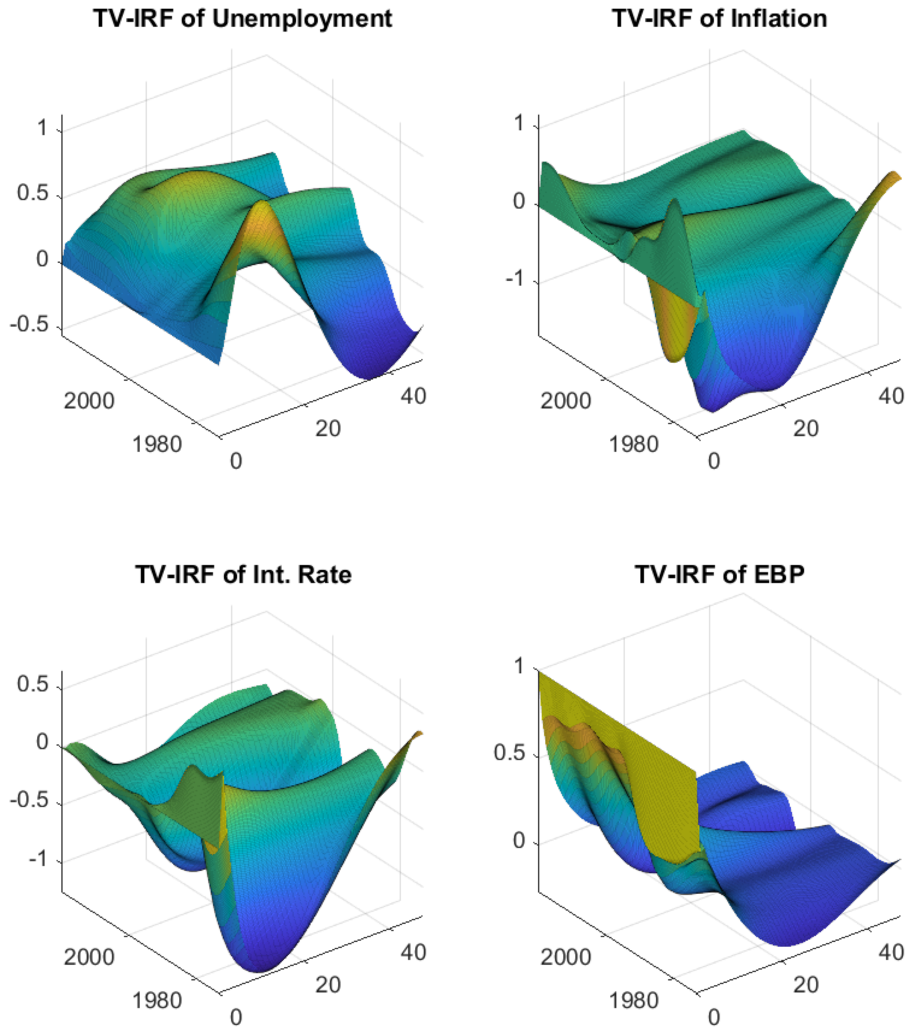


Figure 4.1: Time-varying impulse response functions to an EBP shock, for a selected set of variables. VAR at monthly frequency.

assumption keeps the model tractable and simple. At each point in time agents take parameters' value as given and think they will stay at the same value forever. In the next period agents learn about the changes in parameters and adjust their equations, but do not use this change to forecast future changes in the parameters. A similar result would be obtained by assuming that parameters follow a random walk process, hence the best guess about their future value would still be the current value. Three deep structural parameters are allowed to be time-varying: θ_t, the Calvo parameter setting the degree of price stickyness; $\phi_{\pi,t}$, the strength of reaction of the central bank to changes in inflation; $\sigma_{u,t}$, the variance of the demand shock.

Letting these three specific deep parameters to be time varying is a grounded choice. The time-variation of θ_t implies a time-variation of the PC coefficients; specifically, it

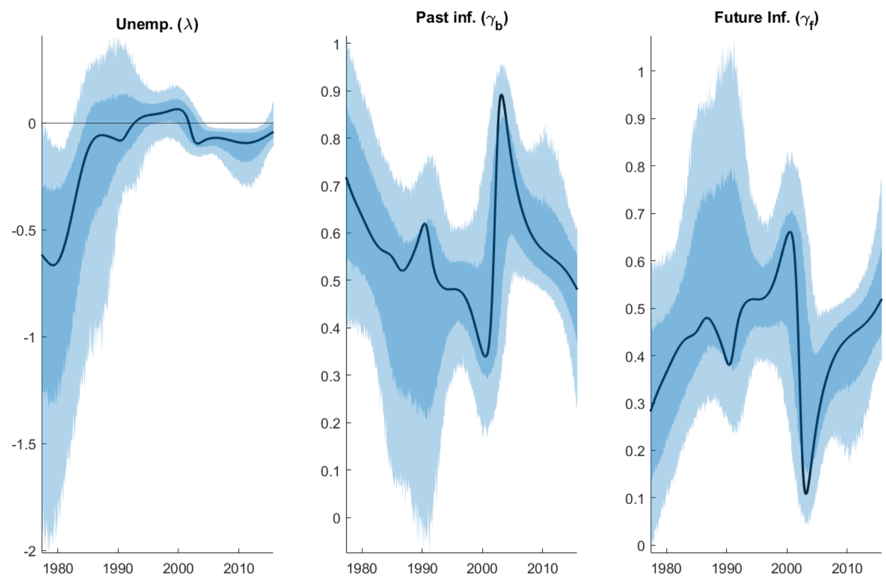


Figure 4.2: Time-varying estimates of the NKPC parameters, from the regression in impulse responses estimated on monthly data. Blue areas show 68% and 90% wild bootstrap confidence intervals.

implies an increasing over time slope λ_t, a decreasing $\gamma_{b,t}$ and an increasing $\gamma_{f,t}$. Time-variation of the variance of demand shocks $\sigma_{u,t}$ and the responsiveness of monetary policy $\phi_{\pi,t}$ have on the contrary no effect on structural NKPC parameters, although they affect the reduced form correlation between inflation and the output gap. Indeed, these two forces are often invoked as possible explanations of the observed trends in reduced form correlations, alternative to the structural change hypothesis. In this way we simulate a world in which we estimate the structural NKPC over time with possible confounding factors in action.

The model in log-linearized form is summarized by the following equations:

$$y_t = \mathbb{E}_t y_{t+1} - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t \pi_{t+1} - u_t\right) \tag{4.5.1}$$

$$y_t = a_t + n_t \tag{4.5.2}$$

$$w_t = \psi_t + \sigma y_t + \varphi n_t \tag{4.5.3}$$

$$mc_t = w_t - a_t \tag{4.5.4}$$

$$\pi_t = \gamma_{b,t}\pi_{t-1} + \gamma_{f,t}\mathbb{E}_t \pi_{t+1} + \lambda_t mc_t + z_t \tag{4.5.5}$$

which feature five endogenous variables, which are denoted as log-deviations from the steady state: the output gap y_t; hours worked n_t; the real wage w_t; real marginal costs mc_t; price inflation π_t. There are four exogenous shocks: demand shocks u_t; productivity shocks a_t; labor supply shocks ψ_t; cost-push shocks z_t. Equation (4.5.5) is the Phillips

Curve formulation of Gali and Gertler (1999), where

$$\begin{aligned} \lambda_t &\equiv (1-\omega)(1-\theta_t)(1-\beta\theta_t)\phi_t^{-1} \\ \gamma_{f,t} &\equiv \beta\theta_t\phi_t^{-1} \\ \gamma_{b,t} &\equiv \omega\phi_t^{-1} \end{aligned} \tag{4.5.6}$$

with $\phi_t \equiv \theta_t + \omega[1-\theta_t(1-\beta)]$. θ_t denotes the Calvo parameter, i.e. the fractions of firms which are not able to adjust the price at any given period. Following Gali and Gertler (1999), we assume that a fraction $1-\omega$ of firms behave as in the standard Calvo pricing setting: they set their price optimally, according to their expectations of future marginal costs. The remaining fraction ω use a simple rule of thumb and index their price to prices in $t-1$. The model is closed by a simple Taylor monetary policy rule:

$$i_t = \phi_{pi,t}\pi_t + \phi_y y_t + m_t \tag{4.5.7}$$

where m_t is a monetary policy shock.

Equations (1)-(6) can be combined to arrive at a two-equations specification:

$$y_t = \frac{1}{\sigma + \phi_y}\left(\sigma \mathbb{E}_t y_{t+1} - \phi_{\pi,t}\pi_t + \mathbb{E}_t\pi_{t+1} + d_t\right) \tag{4.5.8}$$

$$\pi_t = \gamma_{b,t}\pi_{t-1} + \gamma_{f,t}\mathbb{E}_t\pi_{t+1} + \kappa_t y_t + s_t \tag{4.5.9}$$

Equation (4.5.8) denotes the IS curve, where $d_t = u_t - m_t$ collects the two demand shocks. Equation (4.5.9) is the NKPC, with $\kappa_t = \lambda_t(\sigma + \varphi)$ and $s_t = z_t + \lambda_t\psi_t - \lambda(1+\varphi)a_t$ collects the three supply-side shocks.

We further assume that all the five shocks follow an AR(1) process and that all demand (supply) shocks feature the same autoregressive parameter ρ_d (ρ_s). Hence we can write:

$$d_t = \rho_d d_{t-1} + \varepsilon_{d,t} \qquad s_t = \rho_s s_{t-1} + \varepsilon_{s,t} \tag{4.5.10}$$

where $\varepsilon_{d,t} = \varepsilon_{u,t} - \varepsilon_{m,t}$ and $\varepsilon_{s,t} = \varepsilon_{z,t} + \lambda_t\varepsilon_{\psi,t} - \lambda(1+\varphi)\varepsilon_{a,t}$.

We assume that innovations are normally distributed, with time-varying variance:

$$\varepsilon_{d,t} \sim \mathcal{N}(0, \sigma_{d,t}^2) \qquad \varepsilon_{s,t} \sim \mathcal{N}(0, \sigma_{s,t}^2) \tag{4.5.11}$$

with $\sigma_{d,t}^2 = \sigma_{u,t}^2 + \sigma_m^2$ and $\sigma_{s,t}^2 = \sigma_z^2 + \lambda_t^2\sigma_\psi^2 + \lambda^2(1+\varphi)^2\sigma_a^2$.

We conduct four different simulation experiments, which differ in the assumption about the process followed by time-varying parameters. In the first scenario, we assume that all parameters are fixed over time. Table 4.1 show the value chosen for all the parameters of the model, which are standard values used in the literature.

In the second scenario, we let the three time-varying parameters to vary over time in a linear fashion. Namely we assume that θ_t follows a linearly increasing pattern, starting from 0.6 at the beginning of the sample and reaching value 0.9 at the end of

Parameter	Value	Parameter	Value
θ	0.75	σ	1
ϕ_π	1.5	ϕ_y	0.125
φ	2	β	1
ρ_d	0.75	ρ_s	0.75
σ_u	1	σ_m	0
σ_a	0.2	σ_ψ	0.2
σ_z	0.05	ω	0.25

Table 4.1: Chosen values for the parameters of the NK model, when they are set to be fixed over time.

the sample, i.e. $\theta_t = 0.6 + (0.9 - 0.6)\frac{t-1}{T-1}$. Similarly, we set $\phi_{\pi,t} = 1 + (2 - 1)\frac{t-1}{T-1}$ and $\sigma_{u,t} = 1.5 + (0.5 - 1.5)\frac{t-1}{T-1}$.

In the third scenario we assume that the three parameters follow a deterministic sinusoidal pattern over time, differing from sample to sample. We set:

$$\theta_t = 0.75 + 0.15 \sin\left[2\pi\left(\mu_{theta,s} + \frac{t}{fT}\right)\right] \tag{4.5.12}$$

$$\phi_{\pi,t} = 1.5 + 0.5 \sin\left[2\pi\left(\mu_{\phi,s} + \frac{t}{fT}\right)\right] \tag{4.5.13}$$

$$\sigma_{u,t} = 1 + 0.5 \sin\left[2\pi\left(\mu_{\sigma,s} + \frac{t}{fT}\right)\right] \tag{4.5.14}$$

Notice that the processes above are such that the three parameters are bounded above and below by the same values used in the linear processes used in the second scenario. Furthermore, the three parameters $\mu_{\theta,s}, \mu_{\phi,s}, \mu_{\sigma,s}$ are drawn randomly in each sample from a Uniform$(0, 2\pi)$ distribution, shifting randomly the staring point of the process in each sample. Finally, the parameter f governs the frequency of the sinusoidal pattern, hence the degree of time variation in the parameter. We conduct three simulations exercises with, respectively, $f = 1$ (Sin1), $f = 2$ (Sin2) and $f = 3$ (Sin3).

Finally in the fourth scenario the three parameters are allowed to follow bounded stochastic processes. We set:

$$\theta_t = 0.75 + 0.15\frac{a_{\theta,t}}{\max_{0\le j\le t}|a_{\theta,j}|} \tag{4.5.15}$$

$$\phi_{\pi,t} = 1.5 + 0.5\frac{a_{\phi,t}}{\max_{0\le j\le t}|a_{\phi,j}|} \tag{4.5.16}$$

$$\sigma_{u,t} = 1 + 0.5\frac{a_{\sigma,t}}{\max_{0\le j\le t}|a_{\sigma,j}|} \tag{4.5.17}$$

where:

$$a_{\theta,t} = a_{\theta,t-1} + v_{\theta,t} \qquad v_{\theta,t} = \alpha v_{\theta,t-1} + \varepsilon_{\theta,t} \tag{4.5.18}$$

$$a_{\phi,t} = a_{\phi,t-1} + v_{\phi,t} \qquad v_{\phi,t} = \alpha v_{\phi,t-1} + \varepsilon_{\phi,t} \tag{4.5.19}$$

$$a_{\sigma,t} = a_{\sigma,t-1} + v_{\sigma,t} \qquad v_{\sigma,t} = \alpha v_{\sigma,t-1} + \varepsilon_{\sigma,t} \tag{4.5.20}$$

$a_{\theta,t}, a_{\phi,t}, a_{\sigma,t}$ are random walks with persistent error terms $v_{\theta,t}, v_{\phi,t}, v_{\sigma,t}$. The parameter α governs the degree of persistence in the error terms, hence the overall smoothness in the variation of the parameters, with higher α implying greater smoothness. Again, we conduct three simulations exercises setting, respectively, $\alpha = 0.5$ (Rw1), $\alpha = 0.8$ (Rw2) and $\alpha = 0.99$ (Rw3).

For each experiment, we simulate 500 samples from the NK model solution, under the eight different specification explained above. In each sample we first estimate the coefficients of the NKPC using a fixed-parameters version of the methodology used in the paper. In this case, we estimate a single VAR model for the full sample and use Lewis and Mertens (2022) methodology to recover the NKPC parameters. Structural impulse responses to the demand shock are recovered by using the demand shock series as an external instrument. We do this by simply adding the demand shock the vector of estimated residuals and then compute the Cholesky decomposition of the resulting variance-covariance matrix, as suggested by Plagborg-møller and Wolf (2021). Notice that we are assuming the shock is fully observed, while typically we observe only a proxy for it, with some measurement error left. We make this assumption to see the performance of our methodology in the simplest and best possible scenario. After the fixed-parameters estimation, we apply our time-varying methodology. By comparing the two estimation, we can appreciate the extent to which our methodology capture the time variation of the NKPC parameters. The Appendix provides other interest insights by exploring other identification strategies and extending the analysis to the Euro Area.

Table 4.2 shows the resulting bias and mean squared error for the three NKPC parameters . For the NKPC slope λ, we can see that the time-varying estimation always results in a sharp drop in the bias and in the MSE, for all specifications but for the fixed parameter one. For the γ_b and γ_f coefficients this is not always the case and the fixed parameter estimation performs better than the time-varying one in some cases. This is due to the fact that the three deep parameters we allow to vary over time (and the bounds we impose for them) imply only a rather small variation in these two parameters, making them almost constant over time. Notice also that since we impose the usual constraint that the two coefficients some to one, the bias and MSE of the two coefficients are deterministically related.

Figure 4.3 provides a visualization of the performance of our methodology in the specification where deep parameters are allowed to vary linearly. We can see how our methodology is able to track quite precisely the evolution of NKPC coefficients over time.

Bias ×1000						
	Fixed Coef. VAR			T.V. Coef. VAR		
	λ	γ_b	γ_f	λ	γ_b	γ_f
Fixed	1.4016	1.1238	-1.1238	5.5171	2.9949	-2.9949
Linear	-35.6856	-13.9714	13.9714	6.7828	3.5507	-3.5507
Sin1	-61.8254	-20.7109	20.7109	6.6713	3.6371	-3.6371
Sin2	-31.8063	-12.7159	12.7159	7.081	2.4202	-2.4202
Sin3	-15.9856	-3.02508	3.02508	8.7397	7.1829	-7.1829
Rw1	-14.7815	-3.3333	3.3333	7.2565	4.4142	-4.4142
Rw2	-16.9236	-5.30863	5.30863	6.6709	2.8273	-2.8273
Rw3	-20.8632	-6.29557	6.29557	7.7829	1.1187	-1.1187
MSE ×1000						
	Fixed Coef- VAR			T.V. Coef. VAR		
	λ	γ_b	γ_f	λ	γ_b	γ_f
Fixed	0.13071	0.54061	0.54061	0.72497	2.4577	2.4577
Linear	10.7359	1.10001	1.10001	1.176	2.5209	2.5209
Sin1	25.4701	2.03697	2.03697	2.3328	3.335	3.335
Sin2	13.0518	1.42993	1.42993	2.0583	3.1808	3.1808
Sin3	6.6213	0.91404	0.91404	2.1404	3.0606	3.0606
Rw1	6.5545	0.87429	0.87429	2.6711	3.3484	3.3484
Rw2	6.8627	0.96797	0.96797	2.327	3.4049	3.4049
Rw3	8.9312	1.2783	1.2783	2.8211	4.3974	4.3974

Table 4.2: Bias and mean squared error for the three NKPC coefficients in the different simulation experiments.

4.6 Conclusion and further developments

In this paper we estimated a time-varying structural NKPC for the US, by combining the SP-IV method by Lewis and Mertens (2022) with a nonparametric estimate of time-varying impulse response functions. Results show a drastic decline in the slope of the PC in the years 1973-1990 and a flat PC from 1990 onward. Furthermore, the coefficient related to inflation expectations is estimated to increase over time. These results, in line with other studies, support the hypothesis of a decrease in the structural slope of the PC as the main force behind the observed decline in the reduced form correlation between unemployment and inflation. Other possible explanations, such as as increased responsiveness of monetary policy to demand shocks, are not supported by our analysis. Furthermore, our study allows to place the time at which the NKPC becomes practically flat somewhere between years 1985 and 1990. A simulation analysis from a simple DSGE model confirms the validity of our methodology.

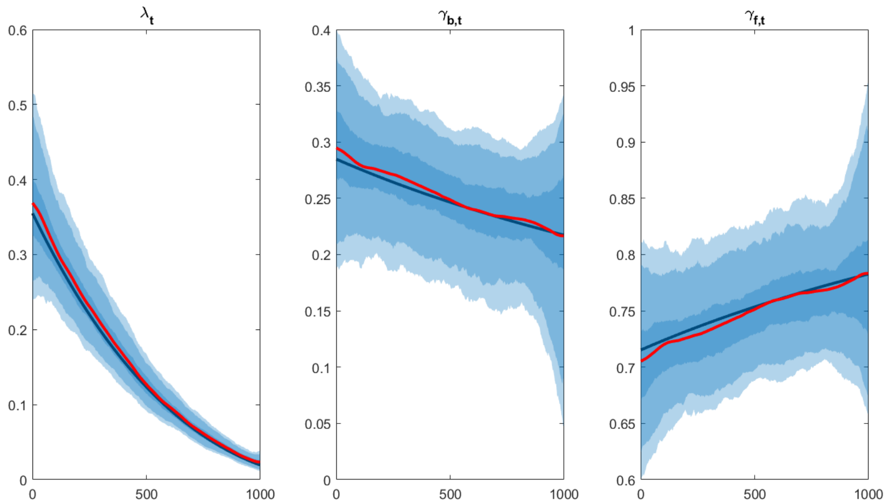


Figure 4.3: Simulation results in the specification with linearly changing parameters. Black lines show the true value of NKPC parameters over time. Red lines show the average estimate. Blue areas show 5-10-32-68-90-95 percentiles of the empirical distribution of the estimates.

Chapter 5

A Note on Robust Inference on Stochastic Time-Varying Coefficients

5.1 Introduction

Models with time-varying coefficients have recently gained considerable interest in parallel with the growing need to model structural changes of macroeconomic and financial variables. The business cycle, rare disasters (e.g., pandemic outbreaks) and regime changes in monetary policy represent some of the possible reasons why models with time-constant parameters might not be representative of the observed economic phenomena; see, among others, Cogley and Sargent (2001); Stock and Watson (2002); Kapetanios and Tzavalis (2010); Demetrescu et al. (2020). Nonparametric techniques have the appealing advantage – with respect to competing methods, e.g., Bayesian methods – of being computationally efficient, and have shown to provide the desired properties of consistency and asymptotic normality when estimating persistence and volatility of such processes - see Giraitis et al. (2014, 2018).

In this note, we focus on a simple time-varying coefficient model in which the time-varying slope coefficient evolves as a random walk with bounded variation. The random walk specification has been widely adopted in the literature (see, e.g., Stock and Watson, 2002, Primiceri, 2005, Fu et al., 2022), but there is no contribution in the literature focusing on the behavior of nonparametric methods in this setup when a "large" bandwidth is adopted. We show that the standard local-constant kernel estimator of the time-varying coefficient preserves its properties of consistency and convergence to a mean-zero Gaussian distribution when the bandwidth is "large", but the variance of such limit distribution is larger. Hence – differently than in the case of a deterministic time-varying coefficient – the use of a "large" bandwidth does not affect the center of the limit distribution of the estimator, but only its variance.

In this note, we provide novel results in the literature of stochastic time-varying coefficient by: i) deriving the value of the MSE-optimal bandwidth when the coefficient is a random walk with local-to-zero variance; ii) showing that the confidence intervals (CIs) proposed by Giraitis et al. (2014, 2018) are invalid in this setup, whereas a different studentization based on the variability of the stochastic time-varying coefficient is necessary to deliver CIs asymptotically correct coverage. The practical importance of the method is shown via numerical simulations. Moreover, we discuss the important issue of delivering correct intervals without the need to estimate the local-to-zero variance of the stochastic time-varying coefficient via the bootstrap.

The rest of the note is structured as follows. In Section 5.2, we present the model and the estimator. In Section 5.3 we show the main theoretical and numerical results, as well as a discussion on possible implementations of the bootstrap. Section 5.4 concludes.

5.2 The model and the estimator

We consider a DGP of the form:

$$y_t = \beta_t x_t + \varepsilon_t$$
$$\beta_t = \beta_{t-1} + \lambda_n \nu_t$$

where ε_t is a *mds* and ν_t is (for simplicity) assumed to be a *iid* process with $\mathbb{E}[\nu_t] = 0$ and $\mathbb{E}[\nu_t^2] = 1$. Moreover, we let

$$\lambda_n := n^{-1/2}\sigma_\nu \tag{5.2.1}$$

so that the variance of the time-varying parameter is local-to-zero. In order to derive a central limit theorem for $\hat{\beta}_t$, a crucial condition to bound the "bias" term of the test statistic associated to the estimator is

$$\sup_{l:|t-l|<k} |\beta_t - \beta_l| = O_p\left(\sqrt{\frac{k}{n}}\right) \tag{5.2.2}$$

Remark 5.2.1 *Condition (5.2.2) is satisfied under the considered scenario because:*

$$\sup_{l:|t-l|<k} |\beta_t - \beta_l| = \lambda_n \sup_{l:|t-l|<k} \left|\sum_{j=1}^{l} \nu_j\right| = O(n^{-1/2})O_p(k^{1/2})$$

where the last equality follows from the fact that

$$(l\tau)^{-1/2} \sum_{j=1}^{[l\tau]} \nu_j \to_{D[0,1]} W(\tau), \qquad l \to \infty \tag{5.2.3}$$

where $\tau \in [0,1]$ and W is a standard Brownian motion.

Even though the analysis can be extended to any class of local-polynomial estimators of β_t for a fixed t, we here focus, for simplicity, on the asymptotic properties of the local constant kernel estimator:

$$\hat{\beta}_t := \frac{\sum_{j=1}^n k_{tj} y_j x_j}{\sum_{j=1}^n k_{tj} x_j^2} \tag{5.2.4}$$

where $k_{tj} := K((t-j)/H)$ such that $K : \mathbb{R} \to [0, +\infty)$ is a standard (truncated) kernel function and H is a bandwidth satisfying $H \to \infty$ and $H/T \to 0$. Let us define $\hat{\sigma}_{x,n}^2 := \sum_{j=1}^n k_{tj} x_j^2$ and $\hat{\sigma}_{x\varepsilon,n}^2 := \sum_{j=1}^n k_{tj}^2 x_j^2 \varepsilon_j^2$; then we can write:

$$\frac{\hat{\sigma}_{x,n}^2}{\hat{\sigma}_{x\varepsilon,n}} \left(\hat{\beta}_t - \beta_t\right) = \underbrace{\frac{\sum_{j=1}^n k_{tj}(\beta_j - \beta_t) x_j^2}{\sqrt{\sum_{j=1}^n k_{tj}^2 x_j^2 \varepsilon_j^2}}}_{=:B_n} + \underbrace{\frac{\sum_{j=1}^n k_{tj} x_j \varepsilon_j}{\sqrt{\sum_{j=1}^n k_{tj}^2 x_j^2 \varepsilon_j^2}}}_{=:\xi_{1n}}$$

Then the following proposition follows directly from Theorem 2.3 in Giraitis et al. (2014).

Proposition 1. *(i) Let the assumptions above hold, then, as $H \to \infty$, we have:*

$$\frac{\hat{\sigma}_{x\varepsilon,n}}{\hat{\sigma}_{x,n}^2} B_n = O_p\left(\sqrt{\frac{H}{n}}\right) \tag{5.2.5}$$

$$\frac{\hat{\sigma}_{x\varepsilon,n}}{\hat{\sigma}_{x,n}^2} \xi_{1n} = O_p\left(\frac{1}{\sqrt{H}}\right) \tag{5.2.6}$$

and

$$\xi_{1n} \xrightarrow{d} N(0,1) \tag{5.2.7}$$

(ii) If additionally $H = o(\sqrt{n})$,

$$\frac{\hat{\sigma}_{x,n}^2}{\hat{\sigma}_{x\varepsilon,n}} \left(\hat{\beta}_t - \beta_t\right) \xrightarrow{d} \xi_1 \tag{5.2.8}$$

From Proposition 1 we can see that confidence intervals based on the standard errors $\hat{\sigma}_{x,n}^2/\hat{\sigma}_{x\varepsilon,n}$ are only valid if the condition $H = o(\sqrt{n})$ is satisfied. The intuition behind this result is that setting $H = o(\sqrt{n})$ is equivalent to setting an "undersmoothing" bandwidth in nonparametric regression, which makes the term B_n asymptotically negligible. This can be seen from the fact that $\hat{\sigma}_{x,n}^2/\hat{\sigma}_{x\varepsilon,n} = O_p(\sqrt{H})$, implying – from (5.2.5) – that:

$$B_n = O_p\left(\frac{H}{\sqrt{n}}\right)$$

where $O_p\left(\frac{H}{\sqrt{n}}\right) = o_p(1)$ only if $H = o(\sqrt{n})$.

Remark 5.2.2 *A choice of H of the form $H = o(\sqrt{n})$ poses two main issues.*
(i) It can be shown that a choice of the bandwidth of the form $H = o(\sqrt{n})$ is MSE-suboptimal. To see this, one can note that, on the grounds of MSE-optimality, one should equate the rates of convergence of the squared bias and the variance of the centered esti-

mator. Intiuitively, if we have:

$$\mathbb{E}\left[\left(\frac{\hat{\sigma}^2_{x,n}}{\hat{\sigma}_{x\varepsilon,n}}B_n\right)^2\right] = c_1\frac{H}{n} + o(1) \qquad \mathbb{E}\left[\left(\frac{\hat{\sigma}^2_{x,n}}{\hat{\sigma}_{x\varepsilon,n}}\xi_{1n}\right)^2\right] = c_2\frac{1}{H} + o(1) \tag{5.2.9}$$

where c_1 and c_2 are constants defined in Section 5.3. Then, by equating the two dominant terms terms one obtains:

$$H_{MSE} = \sqrt{c_2 n/c_1} = O\left(\sqrt{n}\right) \tag{5.2.10}$$

where which clearly is not consistent with the condition $H = o(\sqrt{n})$.
(ii) One could reasonably pick a bandwidth of the form $H = o(\sqrt{n})$ on the grounds of easier tractability of the bias term. However, this poses the question of which bandwidth satisfying $H = o(\sqrt{n})$ is better to choose, as infinitely many such choices (and infinitely many rates of convergence) exist.

Issue (i) above shows that an MSE-optimal bandwidth should be of the form $H = O(\sqrt{n})$. If this is the case, then confidence intervals based on the standard errors $\hat{\sigma}^2_{x,n}/\hat{\sigma}_{x\varepsilon,n}$ (those suggested by Giraitis et al., 2014) are not valid, as they do not account for the additional variability given by B_n. Therefore, CIs based on $\hat{\sigma}^2_{x,n}/\hat{\sigma}_{x\varepsilon,n}$ will show undercoverage for the true value of β_t. Since the variance of B_n directly proportional to the value of σ_ν, what we expect to see is that, for a fixed n and a fixed bandwidth of the form $H = O(\sqrt{n})$, the undercoverage would be more accentuated as σ_ν increases. This is confirmed by the results in Section 5.3.

5.3 Main Results

This section is divided in the parts. In Section 5.3.1, we show derive the MSE-optimal bandwidth for the local-constant kernel estimator in this setup, and a proper studentization for the demeaned estimator that is valid when a "large" bandwidth is selected. In Section 5.3.2, we show numerical evidence in support of the use of the proposed standard errors. Finally, in Section 5.3.3, we give preliminary intuition on how the bootstrap could be used to deliver asymptotically valid confidence intervals in this setup.

5.3.1 Theoretical Results

We now move to formalizing the theoretical results. To do so, we impose the following set of assumptions.

Assumption 1. *(i) x_t is a covariance stationary stochastic process with $\mathbb{E}|x_t|^r < C_1$ for some $C_1 \in \mathbb{R}^+$ and some $r \geq 4$ and satisfying $\sum_{k=1}^{L_n} |\gamma(k)| = o(L_n)$ for $\gamma(k) := Cov(x_t, x_{t+k})$ and some $L_n \to \infty$; (ii) ε_t is a mds with $\mathbb{E}[\varepsilon_t|x_t] = 0$ and $\mathbb{E}|\varepsilon_t|^{r'} < C_2$ for some $C_2 \in \mathbb{R}^+$ and some $r' \geq 4$.*

Assumption 2. *β_t is a random walk process $\beta_t = \beta_{t-1} + \lambda_n \nu_t$, independent with $(x_t, \varepsilon_t)'$,* *with* $\lambda_n := n^{-1/2}\sigma_\nu$*, with* $\sigma_\nu > 0$*,* $\mathbb{E}|\nu_t|^2 = 1$ *and* $\mathbb{E}|\nu_t|^\delta < C_3$ *for some* $\delta \geq 4$*.*

Assumption 3. $K : \mathbb{R} \to [0, +\infty)$ *is a second order kernel function such that* $K(x) = 0$ *if* $|x| \geq 1$ *and* $K(x) > 0$ *otherwise.*

Assumption 1 is a standard regularity condition on the regressor and the error term, widely adopted in the literature (see, for instance, Fu et al., 2022); Assumption 2 defines the main properties of the stochastic time-varying coefficient; whereas Assumption 3 characterizes the kernel function; note that Assumption 3 allows for all the most widely adopted truncated kernel functions, e.g., the Uniform or the Epanechnikov's kernel.

Theorem 5.3.1 *Let Assumptions 1-3 hold, then:*

$$MSE[\hat{\beta}_t] = \frac{H}{n}\sigma_\nu^2\sigma_x^2 \int K^2(u)u^2 du + \frac{1}{H}\sigma^2 \int K^2(u)du \tag{5.3.1}$$

and the MSE-optimal bandwidth is:

$$H_{MSE} := \sqrt{\frac{n\sigma^2 \int K^2(u)du}{\sigma_\nu^2\sigma_x^2 \int K^2(u)u^2 du}} \tag{5.3.2}$$

Theorem 5.3.1 provides the value of the MSE-optimal bandwidth in the considered scenario. As expected, such value is inversely proportional to the value of σ_ν: the larger the variability of the stochastic time-varying coefficient, the shorter should be the window of observations considered for estimating β_t at a fixed time point.

Theorem 5.3.2 *Let Assumptions 1-3 hold, then, if* $H = O(\sqrt{n})$*:*

$$\frac{\hat{\sigma}_{x,n}^2}{\sqrt{\hat{\sigma}_{x,n}^2 + \kappa\sigma_\nu^2\hat{\sigma}_x^2 \int K^2(u)u^2 du}} \left(\hat{\beta}_t - \beta_t\right) \xrightarrow{d} \xi_1 \tag{5.3.3}$$

where $\kappa = \lim_{n\to\infty} H/\sqrt{n}$*.*

Theorem 5.3.2 shows that the standard errors for the local constant estimator of $\hat{\beta}_t$ at a time fixed point changes when "large" bandwidth is adopted in place of an "undersmooothing" bandwidth. The practical relevance of Theorem 5.3.2 is analyzed in the following section.

5.3.2 Numerical Results

We now show that the results in Section 5.3.1 have practical relevance via Monte Carlo simulations. The aim is twofold: first, we show that CIs based on the standard errors proposed by Giraitis et al. (2014) suffer from undercoverage, even in large samples, when a "large" bandwidth is selected; second, we show that the modified standard errors based on Theorem 2 are able to correct this undercoverage. Specifically, we consider simulations for the model

$$y_t = \beta_t x_t + \varepsilon_t$$
$$\beta_t = \beta_{t-1} + \lambda_n \nu_t$$

where $x_t = \rho x_{t-1} + u_t$ and $(\varepsilon_t, \nu_t, u_t)' \sim N(0_{3\times 1}, \Omega)$ where $\Omega = \text{diag}(\sigma_\varepsilon^2, \sigma_\nu^2, \sigma_u^2)$. We set $(\rho, \sigma_\varepsilon, \sigma_u) = (0.3, 1, 1)$ and show results for a grid of values of σ_ν. Simulations are based on 5,000 Monte Carlo replications of the above DGP.

	n				
σ_ν	100	250	500	750	1000
1	83.3	86.0	89.3	90.0	88.8
1.5	81.6	85.1	85.1	87.4	88.4
2	77.7	80.4	82.0	82.3	80.7
2.5	75.2	80.7	78.1	78.2	78.5
3	73.2	76.1	73.0	74.9	75.6

Table 5.1: *Coverage probabilities of 95% confidence intervals based on Giraitis et al. (2014) method with $H = \sqrt{n}$.*

Table 1 and 2 report average empirical coverage probabilities (ECPs) of CIs for the stochastic time-varying parameter at a fixed time point $\tau = \lfloor 0.5n \rfloor$ for the values $n \in \{100, 250, 500, 750, 1000\}$ and $\sigma_\nu \in \{1, 1.5, 2, 2.5, 3\}$. Specifically, Table 1 shows ECPs of CIs based on Giraitis et al.'s (2014) standard errors for the choice $H = \sqrt{n}$, i.e., that suggested in Giraitis et al. (2014, 2018). It is clear from Table 1 that the CIs are undercovering, with such effect being more accentuated as σ_ν increases. On the other hand, Table 2 compares the same method to the CIs relying on the standard errors based on Theorem 2, for the choice $H = H_{MSE}$. Two main conclusions can be drawn from Table 2. First, CIs based on Theorem 2 are inevitably wider to a level that suffices to capture the additional variability of the estimator given by the fact that $B_n \neq o_p(1)$, therefore delivering ECPs very close to the nominal lever under all considered scenarios. Second, the choice $H = H_{MSE}$ seems to have some also for Giraitis et al.'s (2014) CIs;

5.3.3 Bootstrap Inference

The main limitation of the proposed procedure is that standard errors based on Theorem 2 are unfeasible, as they depend on $\sigma_\varepsilon, \sigma_\nu$ and σ_x. While σ_ε and σ_x are relatively easily to estimate, the fact that σ_ν is a local-to-zero variance of a latent random walk process makes its estimation not straightforward. We here show how the bootstrap could be a potential solution of the problem.

Consider a fixed regressor wild bootstrap DGP of the form

$$y_t^* = \hat{\beta}_t x_t + \varepsilon_t^*, \tag{5.3.4}$$

where $\varepsilon_t^* := \eta_t^* \hat{\varepsilon}_t$, such that η_t^* is an independent and identically distributed (conditionally

		Coverage					Av. Length				
		n					n				
σ_ν	Method	100	250	500	750	1000	100	250	500	750	1000
1	GKY14	76.8	77.9	77.7	78.1	78.4	0.584	0.460	0.384	0.347	0.322
	Z24	94.2	95.1	94.9	95.3	94.9	0.935	0.734	0.613	0.554	0.514
1.5	GKY14	76.3	77.2	77.0	78.2	78.2	0.724	0.570	0.475	0.427	0.396
	Z24	94.7	94.8	94.6	95.1	95.2	1.169	0.909	0.758	0.681	0.632
2	GKY14	75.2	76.8	78.1	77.0	77.9	0.840	0.663	0.553	0.496	0.460
	Z24	94.0	94.3	95.3	94.7	94.5	1.368	1.064	0.883	0.792	0.734
2.5	GKY14	74.0	76.0	76.3	77.6	77.5	0.950	0.746	0.622	0.557	0.515
	Z24	93.5	94.3	94.5	95.3	94.8	1.554	1.203	0.992	0.890	0.823
3	GKY14	73.4	75.9	76.6	77.4	76.8	1.043	0.825	0.684	0.614	0.571
	Z24	93.3	93.9	94.2	94.7	94.0	1.744	1.331	1.096	0.982	0.910

Table 5.2: *Coverage probabilities of 95% confidence intervals based on Giraitis et al. (2014) method (GKY14) - therefore based on the standard errors in (1.8) - and the method of this paper (Z24) - therefore based on the (unfeasible) standard errors in (2.3) - both using* $H = H_{MSE}$.

on the original data) random sequence with mean zero and unit variance and $\hat{\varepsilon}_t$ are the residuals from the original model. As typical in the bootstrap literature, the above bootstrap method would be valid in the standard sense if the condition

$$\frac{\hat{\sigma}^2_{x,n}}{\sqrt{\hat{\sigma}^2_{x,n} + \kappa\sigma^2_\nu\hat{\sigma}^2_x \int K^2(u)u^2du}} \left(\hat{\beta}^*_t - \hat{\beta}_t\right) \xrightarrow{d^*}_p \xi_1 \tag{5.3.5}$$

Preliminary numerical evidence shows that condition (5.3.5) is not satisfied in our setup. Preliminary numerical results suggest that the source of invalidity of the bootstrap seems to be twofold: first, the center of the bootstrap distribution seems to be random, suggesting that the distribution of the bootstrap test statistic distribution is random in the limit (see Cavaliere and Georgiev, 2020); second, the variance of such random in the limit distribution seems to be different from that of the asymptotic test statistic.

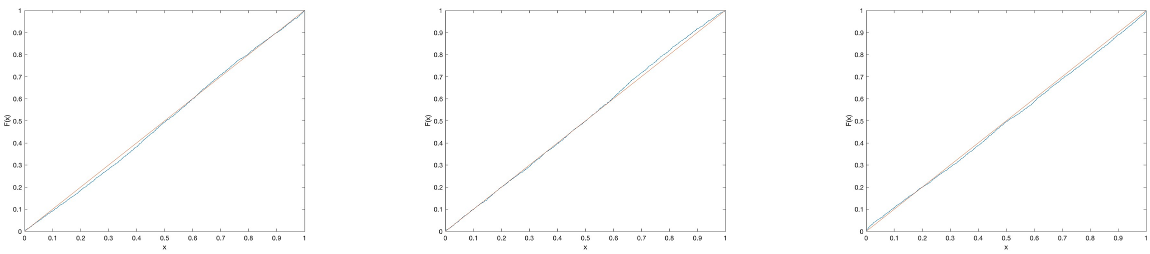
As there is evidence in support of the fact that the bootstrap is not valid in a standard sense, we here propose an alternative bootstrap method, based on the same fixed-regressor wild bootstrap DGP, which has the potential to be able to remove both sources of invalidity of the bootstrap in a standard sense.

Let us define $T_n := \sqrt{H}(\hat{\beta}_t - \beta_t)$ and $T^*_n := \sqrt{H}(\hat{\beta}^*_t - \hat{\beta}_t)$; moreover, let us denote with v_1^2 the limit variance of T_n and with v_2^2 the limit variance of T^*_n. The main idea behind our procedure is that, even if T^*_n is random in the limit, if $v_1 = v_2$, then there are conditions under which the bootstrap p-value $\hat{p}_n := \mathbb{P}^*(T^*_n \leq T_n)$ is valid, see Cavaliere and Georgiev (2020). Since $v_1 \neq v_2$, then one could think about restoring validity by considering the modified p-value

$$\hat{p}^{\texttt{mod}}_n := \mathbb{P}^*(QT^*_n \leq T_n) \tag{5.3.6}$$

where $Q = (v_1/v_2)$, which would be able to correct both sources of invalidity of the "standard" bootstrap method. Even if v_1 and v_2 are both infeasible, as they both depend on the value of σ_ν, we show evidence to the fact that Q does not. Specifically, we performed the following simulation exercise: (i) we simulate the same DGP used in Section 5.3.2 with $\sigma_\nu = 1$; (ii) we compute the sample variance (over Monte Carlo replications) of T_n and T_n^*; (iii) we run a new Monte Carlo simulation in which, at each replication, we compute the value of $\hat{p}_n^{\texttt{mod}}$, using the sample values of v_1 and v_2 obtained in point (ii); (iv) we compute the empirical distribution function; (v) we replicate points (iii)-(v) for different values of σ_ν, keeping the same value of Q obtained in point (ii). Figure 5.1 shows the empirical distribution functions of $\hat{p}_n^{\texttt{mod}}$, where we have evidence in favor of the uniformity of the proposed modified p-value.

The derivation of the distribution of T_n^* and Q and, in general, the proof of the validity of the proposed bootstrap method, is current work in progress of the author.



(a) $\sigma_\nu = 1, \sigma_x = 1$ (b) $\sigma_\nu = 1, \sigma_x = 2$ (c) $\sigma_\nu = 2.5, \sigma_x = 1$

Figure 5.1: Empirical distribution function of $\hat{p}_n^{\texttt{mod}}$.

5.4 Conclusion

In this note, we have analyzed the properties of nonparametric estimation methods for time-varying coefficient models under the assumption that the coefficient follows a random walk with local-to-zero variance.

We derived the MSE-optimal bandwidth for the random walk specification and demonstrated that it is proportional to the square root of the sample size. This finding challenges the common practice of using "undersmoothing" bandwidths to simplify bias analysis, as such choices are MSE-suboptimal and result in inefficient confidence intervals (CIs). Furthermore, we showed that the CIs proposed by Giraitis et al. (2014) are invalid for large bandwidths because they fail to account for the additional variability introduced by the stochastic nature of the time-varying coefficient. We proposed a modified standardization procedure, which corrects this issue and ensures valid CIs with asymptotically correct coverage.Our theoretical contributions are complemented by numerical simulations, which confirm the practical relevance of our results.

In conclusion, this note highlights the importance of adapting nonparametric inference methods to account for the stochastic nature of time-varying coefficients. By addressing the limitations of existing approaches, our results provide a more accurate framework for modeling and inference in time-varying systems, paving the way for future advancements in this growing field of research.

Appendices

Appendix A

Appendix to Chapter 1

A.1 Special case: T_n is asymptotically Gaussian

In this section, we specialize Assumptions 1.3.1, 1.3.1, and 1.3.2 to the case where $T_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ is a normalized parameter estimator whose limiting distribution is normal. We consider the following special case of Assumption 1.3.1.

It holds that $T_n - B_n \to_d N(0, v^2)$, where $v^2 > 0$.

Assumption A.1 covers statistics T_n based on asymptotically biased estimators: when $B_n \to_p B$, we have $T_n \to_d N(B, v^2)$, in which case B is the asymptotic bias of $\hat{\theta}_n$. More generally, we can interpret B_n as a bias term that approximates $E(\sqrt{n}(\hat{\theta}_n - \theta_0))$ although B_n does not need to have a limit. Note that Assumption A.1 obtains from Assumption 1.3.1 when we let $\xi_1 \sim N(0, v^2)$ and $G_\gamma(u) = \Phi(u/v)$, in which case $\gamma = v$.

Let D_n^* denote a bootstrap sample from D_n and let $\hat{\theta}_n^*$ be a bootstrap version of $\hat{\theta}_n$. The bootstrap analogue of T_n is $T_n^* = \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$.

It holds that (i) $T_n^* - \hat{B}_n \overset{d^*}{\to}_p N(0, v^2)$, and (ii)

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} \overset{d}{\to} N(0, V), \quad V := (v_{ij}), \quad i, j = 1, 2,$$

where $v_d^2 := v_{11} + v_{22} - 2v_{12} > 0$ with $v_{11} := v^2 > 0$.

Assumption A.1(i) requires the bootstrap statistic $T_n^* - \hat{B}_n$ to mimic the asymptotic distribution of $T_n - B_n$, as in Assumption 1.3.1(i). However, and contrary to Assumption 1.3.1(i), here this limiting distribution is the zero mean Gaussian distribution (i.e. $G_\gamma(u) = \Phi(u/v)$), which means that we can interpret $\hat{B}_n$ as a bootstrap bias correction term; i.e., $\hat{B}_n = E^*(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n))$. Assumption A.1(ii) assumes that $\hat{B}_n - B_n$ is also asymptotically distributed as a zero mean Gaussian random variable (jointly with $T_n - B_n$).[1] An implication of this assumption is that

$$T_n - \hat{B}_n = (T_n - B_n) - (\hat{B}_n - B_n) \overset{d}{\to} N(0, v_d^2), \tag{A.1.1}$$

[1] In terms of Assumption 1.3.1, Assumption A.1 corresponds to the case where the vector $\xi = (\xi_1, \xi_2)'$ is a multivariate normal distribution with covariance matrix V.

where $v_d^2 := v_{11} + v_{22} - 2v_{12}$. We do not require V to be positive definite; for instance, $v_{22} = 0$ whenever $\hat{B}_n - B_n = o_p(1)$, and in fact V can be rank deficient even when $v_{22} > 0$. However, we do impose the restriction that $v_d^2 > 0$. This ensures that the limiting distribution function of $T_n - \hat{B}_n$, given by $F_\phi(u) = \Phi(u/v_d)$, is well-defined and continuous, as assumed in Assumption 1.3.1(ii). Note that we can let $\phi = V$ in this case, or simply set $\phi = v_d$.

Let $\hat{p}_n$ denote the standard bootstrap p-value as defined in Section 1.3. We then obtain the following.

Corollary A.1.1 *Under Assumptions A.1 and A.1,* $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$*, where* $m^2 := v_d^2/v^2$.

Corollary A.1.1 follows immediately from Theorem 1.3.1 when we let $G_\gamma(u) = \Phi(u/v)$ and $F_\phi(u) = \Phi(u/v_d)$. It shows that the asymptotic distribution of $\hat{p}_n$ is uniform only when $m = 1$, or equivalently when $v_d^2 = v^2$. In this case, the difference $\hat{B}_n - B_n$ is $o_p(1)$. When $v_d^2 \neq v^2$, $\hat{B}_n - B_n$ is random even in the limit, implying that the limiting bootstrap distribution function of T_n^* is conditionally random. Although random limit bootstrap measures do not necessarily invalidate bootstrap inference, as discussed by Cavaliere and Georgiev (2020), this is not the case here. However, we can solve the problem of bootstrap invalidity by applying the prepivoting approach or by modifying the test statistic from T_n to $T_n - \hat{B}_n$.

To describe the prepivoting approach, note that the limiting distribution of $\hat{p}_n$ is given by

$$H(u) := \lim P(\hat{p}_n \leq u) = \Phi(m^{-1}\Phi^{-1}(u)).$$

Hence, a plug-in approach amounts to estimating $m^2 := v_d^2/v^2$, where v^2 and v_d^2 are defined in Assumption A.1. Suppose that $\hat{v}_n^2$ and $\hat{v}_{d,n}^2$ are consistent estimators of v^2 and v_d^2 (i.e., assume that $(\hat{v}_n^2, \hat{v}_{d,n}^2) \to_p (v^2, v_d^2))$ and let $\hat{m}_n^2 := \hat{v}_{d,n}^2/\hat{v}_n^2$. Then, by Corollary 1.3.2, it immediately follows that

$$\tilde{p}_n = \Phi(\hat{m}_n^{-1}\Phi^{-1}(\hat{p}_n)) \xrightarrow{d} U_{[0,1]}$$

under Assumptions A.1 and A.1. For brevity, we do not formalize this result here.

To describe the double bootstrap modified p-value, $\tilde{p}_n := \hat{H}_n(\hat{p}_n) = P^*(\hat{p}_n^* \leq \hat{p}_n)$, when applied to the special case where T_n satisfies Assumption A.1, we now introduce Assumption A.1.

Let $T_n^{**} = \sqrt{n}(\hat{\theta}_n^{**} - \hat{\theta}_n^*)$ and suppose that (i) $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0, v^2)$, in probability, and (ii) $T_n^* - \hat{B}_n^* \xrightarrow{d^*}_p N(0, v_d^2)$, where v_d^2 is as defined in Assumption A.1(ii).

Under Assumption A.1(i), the double bootstrap distribution of $T_n^{**} - \hat{B}_n^*$ mimics the distribution of $T_n^* - \hat{B}_n$, where the double bootstrap bias term $\hat{B}_n^* = E^{**}(\sqrt{n}(\hat{\theta}_n^{**} - \hat{\theta}_n^*))$ is asymptotically centered at $\hat{B}_n$ under Assumption A.1(ii). When $v_d^2 \neq v^2$, the double bootstrap bias is not a consistent estimator of $\hat{B}_n$, but that is not needed for the asymptotic validity of the modified double bootstrap p-value $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$ defined in Section 1.3.

By application of Theorem 1.3.2, $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$ under Assumptions A.1, A.1, and A.1. We can also provide a result analogous to Corollary 1.3.3 under these assumptions. In this case, if closed-form expressions for $\hat{B}_n$ and $\hat{B}_n^*$ are not available, we can approximate these bootstrap expectations by Monte Carlo simulations and then compute $P^*(T_n^* - \hat{B}_n^* \leq T_n - \hat{B}_n)$ as a valid bootstrap p-value. Note, however, that this approach is computationally as intensive as the prepivoting approach based on $\tilde{p}_n$ since it too requires two layers of resampling.

Remark A.1.1 *In the case of asymptotically Gaussian statistics discussed in this section, the more general Assumptions 1.3.5 and 1.3.5 simplify straightforwardly. In Assumption A.1(i) we assume that $T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v_s^2)$ and in Assumption A.1(i) that $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0, v_s^2)$, in probability, for some $v_s^2 > 0$, while the rest of Assumptions A.1–A.1 are unchanged. The results of this section continue to apply under these more general conditions, replacing $G_\gamma(u) = \Phi(u/v)$ with $J_\gamma(u) = \Phi(u/v_s)$ and consequently defining $m := v_d^2/v_s^2$.*

Remark A.1.2 *Contrary to Beran (1987, 1988), in our context the first level of prepivoting, e.g., by the double bootstrap, is used to obtain an asymptotically valid bootstrap p-value. Therefore, inference based on $\tilde{p}_n$ does not necessarily provide an asymptotic refinement over inference based on an asymptotic approach that does not require the bootstrap. Nevertheless, the Monte Carlo results in Table ?? below seem to suggest an asymptotic refinement for the double bootstrap, at least for the non-parametric bootstrap scheme. In the special case where the bias term B_n is of sufficiently small order, the arguments in Beran (1987, 1988) apply, and an asymptotic refinement can be obtained. We also conjecture that, in the general case, an asymptotic refinement could be obtained by further iterating the bootstrap.*

A.2 Examples with details

A.2.1 Inference after model averaging

In this section we first provide the regularity conditions required in Lemmas 1.4.1 and 1.4.2, and then we give the proofs of the lemmas. We subsequently provide some brief Monte Carlo evidence. Finally, at the end of the section, we provide regularity conditions for the extension to the pairs bootstrap and a proof of the associated Lemma 1.4.3.

Assumptions and notation

We impose the following conditions.

(i) $\varepsilon_t | W \sim$ i.i.d.$(0, \sigma^2)$, where $W := (x, Z)$; (ii) $S_{WW} \to_p \Sigma_{WW}$ with $\text{rank}(\Sigma_{WW}) = q + 1$; (iii) $n^{1/2} S_{W\varepsilon} \to_d N(0, \Omega)$ with $\Omega := \sigma^2 \Sigma_{WW}$.

Remark A.2.1 *We assume that the weights ω are fixed and independent of n. A popular example in forecasting is to use equal weighting. We could allow for stochastic weights as long as these are constant in the limit. This would be the case, for example, when the weights are based on moments that can be consistently estimated.*

To proceed, we introduce the following notation. First, partition Σ_{WW} according to W,

$$\Sigma_{WW} := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xZ} \\ \Sigma_{Zx} & \Sigma_{ZZ} \end{pmatrix}.$$

Let $\Sigma_{xZ_m} := \Sigma_{xZ}R_m$, $\Sigma_{Z_mZ_m} := R_m'\Sigma_{ZZ}R_m$, $\Sigma_{xx.Z_m} := \Sigma_{xx} - \Sigma_{xZ}R_m(R_m'\Sigma_{ZZ}R_m)^{-1}R_m'\Sigma_{Zx}$, and $\Sigma_{xZ.Z_m} := \Sigma_{xZ} - \Sigma_{xZ}R_m(R_m'\Sigma_{ZZ}R_m)^{-1}R_m'\Sigma_{ZZ}$. Also let $A_n := \sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}n^{-1}x'M_{Z_m}$, where $M_{Z_m} := I_n - Z_m(Z_m'Z_m)^{-1}Z_m'$, such that $A_nZ = Q_n$. With this notation,

$$\tilde{\beta}_n = A_ny = A_nx\beta + Q_n\delta + A_n\varepsilon = \beta + Q_n\delta + A_n\varepsilon, \tag{A.2.1}$$

$$\tilde{\beta}_n^* = A_ny^* = \hat{\beta}_n + Q_n\hat{\delta}_n + A_n\varepsilon^*. \tag{A.2.2}$$

Finally, define

$$\bar{d}_{M,n}' := \sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}(1, -S_{xZ_m}S_{Z_mZ_m}^{-1}R_m'),$$

$$\bar{b}_{M,n}' := \sum\nolimits_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}S_{xZ.Z_m}S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q),$$

and let $\bar{d}_M'$ and $\bar{b}_M'$ denote their probability limits, which exist and are well-defined under Assumption A.2.1.

Proofs of lemmas

Proof of Lemma 1.4.1. We first verify Assumption 1.3.1 (or equivalently, Assumption A.1). Using (A.2.1) we can write $T_n = B_n + \xi_{1,n}$ with

$$\xi_{1,n} := n^{1/2}A_n\varepsilon = n^{1/2}\sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}n^{-1}x'M_{Z_m}\varepsilon = n^{1/2}\sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}S_{x\varepsilon.Z_m}.$$

Then

$$\begin{aligned} S_{x\varepsilon.Z_m} &= n^{-1}x'M_{Z_m}\varepsilon = n^{-1}(x'\varepsilon - x'Z_m(Z_m'Z_m)^{-1}R_m'Z'\varepsilon) \\ &= (1, -S_{xZ_m}(S_{Z_mZ_m})^{-1}R_m')S_{W\varepsilon} =: \hat{d}_m'S_{W\varepsilon}, \end{aligned}$$

so that

$$\xi_{1,n} = \sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}\hat{d}_m'n^{1/2}S_{W\varepsilon} = \bar{d}_{M,n}'n^{1/2}S_{W\varepsilon}.$$

Hence, $\xi_{1,n} \to_d N(0, v^2)$ with $v^2 := \bar{d}_M'\Omega\bar{d}_M$.

Next, we verify Assumption 1.3.1 (or Assumption A.1). From (A.2.2) we write $T_n^* = \hat{B}_n + \xi_{1,n}^*$ with $\xi_{1,n}^* := n^{1/2}A_n\varepsilon^* \sim N(0, \hat{\sigma}_n^2A_nA_n')$, conditional on D_n. Part (i) now follows straightforwardly because $\hat{\sigma}_n^2 \to_p \sigma^2$ and $A_n A_n' = \bar{d}_{M,n}' S_{WW} \bar{d}_{M,n} \to_p \bar{d}_M' \Sigma_{WW} \bar{d}_M$. To prove Part (ii), note that

$$n^{1/2}(\hat{\delta}_n - \delta) = S_{ZZ.x}^{-1} S_{Z\varepsilon.x} = S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q) n^{1/2} S_{W\varepsilon},$$

from which it follows that

$$\hat{B}_n - B_n = Q_n n^{1/2}(\hat{\delta}_n - \delta) = Q_n S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q) n^{1/2} S_{W\varepsilon} = \bar{b}_{M,n}' n^{1/2} S_{W\varepsilon}.$$

Hence,

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} = \begin{pmatrix} \bar{d}_{M,n}' \\ \bar{b}_{M,n}' \end{pmatrix} n^{-1/2} W' \varepsilon \xrightarrow{d} N(0, V), \quad V = \begin{pmatrix} \bar{d}_M' \Omega \bar{d}_M & \bar{d}_M' \Omega \bar{b}_M \\ \bar{b}_M' \Omega \bar{d}_M & \bar{b}_M' \Omega \bar{b}_M \end{pmatrix},$$

which completes the proof. □

Proof of Lemma 1.4.2. First note that $\tilde{\beta}_n^{**} = A_n y^{**} = A_n x \hat{\beta}_n^* + A_n Z \hat{\delta}_n^* + A_n \varepsilon^{**}$. It follows that

$$T_n^{**} := n^{1/2}(\tilde{\beta}_n^{**} - \hat{\beta}_n^*) = \hat{B}_n^* + n^{1/2} A_n \varepsilon^{**},$$

where $\hat{B}_n^* := n^{1/2} Q_n \hat{\delta}_n^*$ and $\xi_{1,n}^{**} := n^{1/2} A_n \varepsilon^{**} \sim N(0, \hat{\sigma}_n^{*2} A_n A_n')$, conditional on (D_n, D_n^*). The conditions in Assumption 1.3.2(i) or A.1(i) now follows as in Part (i) of the previous proof because $\hat{\sigma}_n^{*2} \xrightarrow{p^*}_p \sigma^2$. For Assumption 1.3.2(ii) or A.1(ii) we consider the joint convergence of $(T_n^* - \hat{B}_n, \hat{B}_n^* - \hat{B}_n)'$. By noticing that

$$n^{1/2}(\hat{\delta}_n^* - \hat{\delta}_n) = S_{ZZ.x}^{-1} S_{Z\varepsilon^*.x} = S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q) n^{1/2} S_{W\varepsilon^*},$$

it follows that

$$\begin{pmatrix} T_n^* - \hat{B}_n \\ \hat{B}_n^* - \hat{B}_n \end{pmatrix} = \begin{pmatrix} \bar{d}_{M,n}' \\ \bar{b}_{M,n}' \end{pmatrix} n^{1/2} S_{W\varepsilon^*} \sim N(0, \hat{V}_n),$$

conditional on D_n, where

$$\hat{V}_n = \hat{\sigma}_n^2 \begin{pmatrix} \bar{d}_{M,n}' S_{WW} \bar{d}_{M,n} & \bar{d}_{M,n}' S_{WW} \bar{b}_{M,n} \\ \bar{b}_{M,n}' S_{WW} \bar{d}_{M,n} & \bar{b}_{M,n}' S_{WW} \bar{b}_{M,n} \end{pmatrix} \xrightarrow{p} V.$$

The desired result follows. □

A small Monte Carlo experiment

In Table ?? we present the results of a small Monte Carlo simulation experiment to illustrate the above results numerically. We generate the data from the regression model (1.2.1) with sample sizes $n = 10, 20, 40$. The regressors x_t and z_t are both scalar and multivariate normally distributed with unit variances and correlation 0.7, and the errors are either standard normal, t_3, or χ_1^2 distributed. The true values are $\beta = \bar{\beta} + a n^{-1/2}$ with $\bar{\beta} = 1$ and $\delta = 1$ (the results are invariant to $\bar{\beta}$ and δ because we use the unrestricted estimates to construct the bootstrap samples). We test the null hypothesis $\mathsf{H}_0 : \beta = \bar{\beta}$ against a left-sided alternative. Results for right-tailed and two-tailed tests are analogous to those presented here for left-tailed tests. The case $a = 0$ corresponds to rejection frequencies under the null, and $a = -1, -2, -4$ corresponds to rejection frequencies under local alternatives. The estimator puts weight $\omega_1 = 1/2$ on the short model that includes only x (and a constant term) and weight $\omega_2 = 1/2$ on the long model that includes both regressors (and a constant term). We consider two bootstrap schemes. The first is the parametric bootstrap scheme, where $\varepsilon_t^* \sim$ i.i.d.$N(0,1)$, which is denoted as "par." in the table. The second is the non-parametric bootstrap scheme, where ε_t^* is resampled independently from the (centered) residuals from the long regression, which is denoted as "non-par." Results are based on 10,000 Monte Carlo simulations and $B = 999$ bootstrap replications.

First consider the case $a = 0$. The simulation outcomes in Table ?? clearly illustrate our theoretical results. The standard bootstrap p-value, $\hat{p}_n$, is much larger than the nominal level of the test. The plug-in modified p-value, $\tilde{p}_{n,p}$, is close to the nominal level for the parametric bootstrap scheme, but is still over-sized for the non-parametric scheme with the smaller sample sizes. Finally, the double bootstrap modified p-value, $\tilde{p}_{n,d}$, is nearly perfectly sized throughout the table.

Table ?? for $a = -1, -2, -4$ clearly shows nontrivial power, which increases as a increases. The discrepancies in finite-sample power are due to differences in size. For example, consider the standard parametric bootstrap with 5% nominal level and normal errors (top left of the table). It has finite-sample size very close to 10%. Comparing this with our modified bootstrap test with nominal size 10% (towards the right in the same panel of the table), we see that the finite-sample powers are nearly identical.

Extension to the pairs bootstrap

In addition to Assumption A.2.1 we also impose the following conditions.

With $w_t := (x_t, z_t)'$ it holds that (i) $\sup_t E\|w_t\|^4 < \infty$, $E\varepsilon_t^4 < \infty$; (ii) $n^{-1}\sum_{t=1}^n x_t^2\varepsilon_t^2 \to_p \sigma^2\Sigma_{xx}$, $n^{-1}\sum_{t=1}^n x_t^2 w_t w_t' \to_p \Sigma_r > 0$, and $n^{-1}\sum_{t=1}^n x_t^2 w_t\varepsilon_t \to_p 0$.

Proof of Lemma 1.4.3. We first prove that

$$S_{W^*W^*} - S_{WW} \xrightarrow{p^*}_p 0, \tag{A.2.3}$$

$$S_n^* := \begin{pmatrix} n^{1/2}S_{x^*\varepsilon^*} \\ n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} \xrightarrow{d^*}_p N(0, \Sigma_s), \quad \Sigma_s = \begin{pmatrix} \sigma^2\Sigma_{xx} & 0 \\ 0 & \Sigma_r \end{pmatrix}. \tag{A.2.4}$$

Here, (A.2.3) follows by straightforward application of Chebyshev's LLN.

To prove (A.2.4), we first compute the mean and variance of S_n^*. Note that the mean of S_n^* is zero by construction; for example, $E^*(n^{1/2}S_{x^*\varepsilon^*}) = n^{-1/2}\sum_{t=1}^n E^*(x_t^*\varepsilon_t^*) = n^{1/2}S_{x\hat{\varepsilon}} = 0$ by the OLS first-order condition. In addition,

$$\mathrm{Var}^*(n^{1/2}S_{x^*\varepsilon^*}) = n^{-1}\sum_{t=1}^n E^*(x_t^{*2}\varepsilon_t^{*2}) = n^{-1}\sum_{t=1}^n x_t^2\hat{\varepsilon}_t^2 \xrightarrow{p} \sigma^2\Sigma_{xx}$$

Table A.1: Simulated rejection frequencies (%) of bootstrap tests

			5% nominal level						10% nominal level					
			par.			non-par.			par.			non-par.		
dist.	a	n	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$
N	0	10	10.1	5.0	5.0	16.2	11.2	6.3	15.9	10.0	10.0	21.5	16.2	10.5
		20	9.7	5.0	5.1	12.6	7.8	5.4	15.1	9.8	9.8	18.2	12.9	10.4
		40	9.8	5.1	5.2	10.5	5.8	4.9	15.4	10.1	10.2	16.5	11.0	9.8
	-1	10	25.5	15.9	16.0	34.0	26.1	16.2	35.7	25.9	25.8	42.3	34.3	24.8
		20	26.0	16.5	16.7	30.1	21.0	15.9	36.5	26.9	26.6	40.0	30.9	26.1
		40	27.4	17.7	17.9	29.3	19.4	16.9	37.8	28.4	28.4	38.9	30.0	27.6
	-2	10	47.7	35.6	35.7	57.0	47.5	33.2	58.4	48.5	48.1	64.9	57.4	45.7
		20	51.6	38.3	38.3	56.3	44.0	36.2	62.5	52.1	52.3	65.9	56.9	51.4
		40	52.5	39.9	39.8	54.8	43.0	39.1	63.9	53.6	53.8	64.9	55.5	52.8
	-4	10	84.9	75.6	75.6	88.2	82.5	71.6	90.1	84.5	84.3	91.9	87.9	81.2
		20	90.5	82.9	82.7	91.5	85.5	80.2	94.2	90.3	90.0	94.4	91.3	88.8
		40	91.7	85.4	85.3	92.5	87.0	84.7	95.3	92.2	92.0	95.8	92.7	91.7
t_3	0	10	7.3	3.7	3.8	15.6	10.8	5.8	12.0	7.3	7.2	21.5	15.8	10.2
		20	7.5	4.1	4.2	13.2	8.1	5.6	12.7	7.6	7.9	19.0	13.4	10.9
		40	7.5	3.8	3.9	10.5	5.7	4.9	12.8	7.8	7.8	16.6	10.8	9.6
	-1	10	20.9	12.0	11.9	39.4	30.6	19.8	31.7	21.4	21.3	47.7	39.8	29.5
		20	23.3	13.1	13.3	35.2	25.3	19.3	34.2	23.9	24.0	45.0	36.3	31.1
		40	24.6	14.7	14.7	31.8	21.4	19.2	35.6	25.3	25.3	42.3	32.9	30.5
	-2	10	47.5	32.2	32.2	65.2	56.6	42.8	60.3	47.7	47.6	72.7	65.4	55.1
		20	51.4	36.7	37.0	63.7	52.3	45.1	64.4	52.4	52.4	72.5	63.9	59.1
		40	52.8	38.1	38.3	60.8	47.8	44.6	65.6	53.9	53.7	70.9	61.7	58.9
	-4	10	87.7	78.1	77.9	91.3	86.9	78.6	92.1	87.3	87.2	94.1	91.2	85.9
		20	91.8	85.0	84.9	92.6	88.1	84.2	95.1	91.6	91.6	95.3	92.8	91.0
		40	93.2	87.7	87.6	93.2	88.2	86.8	96.1	93.3	93.3	96.0	93.3	92.5
χ^2_1	0	10	8.3	4.7	4.7	16.0	10.7	5.8	12.6	8.0	8.0	21.5	16.2	9.9
		20	8.5	4.9	4.9	12.2	7.0	5.0	13.5	8.6	8.6	18.1	12.4	9.8
		40	9.2	4.9	4.8	10.9	6.1	5.3	14.8	9.7	9.5	17.1	11.2	10.1
	-1	10	21.1	12.6	12.6	41.9	33.2	22.5	30.9	21.7	21.2	50.1	42.0	31.9
		20	23.4	14.3	14.3	35.1	25.1	19.7	33.6	24.0	24.1	45.2	35.9	31.2
		40	25.5	15.8	15.9	31.7	21.2	19.1	36.2	26.6	26.7	42.2	32.7	30.4
	-2	10	46.9	31.3	31.5	65.2	57.2	45.3	60.6	47.6	47.6	72.0	65.4	55.8
		20	51.2	36.3	36.4	62.2	51.4	44.3	64.3	52.4	52.5	71.3	62.9	57.9
		40	53.9	39.2	39.1	59.4	46.9	43.9	65.2	55.1	54.9	69.9	60.4	57.8
	-4	10	87.2	78.5	78.3	88.8	84.3	76.6	91.5	86.6	86.4	91.8	88.6	83.2
		20	91.1	84.7	84.7	90.6	84.6	80.4	94.2	91.0	90.8	93.9	90.5	88.4
		40	92.6	86.8	86.8	91.8	86.7	85.2	95.6	92.7	92.5	94.8	92.0	91.0

Notes: $\hat{p}_n$ denotes the standard bootstrap; $\tilde{p}_{n,p}$ and $\tilde{p}_{n,d}$ denote the modified bootstrap using the plug-in and the double bootstrap methods, respectively. The parametric bootstrap scheme, where $\varepsilon_t^* \sim$ i.i.d.$N(0,1)$, is denoted as "par." and the non-parametric bootstrap scheme, where ε_t^* is re-sampled independently from the long regression (centered) residuals, is denoted as "non-par." The ε_t's are i.i.d. draws from (standardized) N, t_3, and χ^2_1 distributions. The parameter a denotes the drift under the local alternative $\beta_0 = \bar{\beta} + an^{-1/2}$. Results are based on 10,000 Monte Carlo simulations and $B = 999$ bootstrap replications for each level.

under Assumptions A.2.1 and A.2.1. Similarly, letting

$$\begin{pmatrix} n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} = n^{1/2}(S_{x^*W^*} - S_{xW}),$$

we find that

$$\begin{aligned} \mathrm{Var}^*(n^{1/2}\,(S_{x^*W^*} - S_{xW})) &= n^{-1}\sum_{t=1}^{n}(x_t w_t - E^*(x_t^* w_t^*))(x_t w_t - E^*(x_t^* w_t^*))' \\ &= n^{-1}\sum_{t=1}^{n} x_t^2 w_t w_t' - S_{xW}S_{Wx} \xrightarrow{p} \Sigma_r - \Sigma_{xW}\Sigma_{Wx}. \end{aligned}$$

Note also that the covariance between $n^{1/2}S_{x^*\varepsilon^*}$ and $n^{1/2}(S_{x^*W^*} - S_{xW})$ is zero because

$$\begin{aligned} E^*(nS_{x^*\varepsilon^*}S_{x^*W^*}) &= n^{-1}E^*\left(\sum_{t=1}^{n} x_t^*\varepsilon_t^* \sum_{s=1}^{n} x_s^* w_s^*\right) = n^{-1}E^*\left(\sum_{t=1}^{n} x_t^{*2} w_t^* \varepsilon_t^*\right) \\ &= E^*(x_t^{*2} w_t^* \varepsilon_t^*) = n^{-1}\sum_{t=1}^{n} x_t^2 w_t \hat{\varepsilon}_t \xrightarrow{p} 0 \end{aligned}$$

by Assumption A.2.1(ii). Thus, we have shown that $E^*(S_n^*) = 0$ and $E^*(S_n^* S_n^{*\prime}) \to_p \Sigma_s$. The result (A.2.4) now follows because the stated moment conditions imply the Lindeberg condition by standard arguments.

Next we can write

$$T_n^* - \hat{B}_n = n^{1/2} S_{x^*x^*}^{-1} S_{x^*\varepsilon^*} + B_n^* - \hat{B}_n,$$

where

$$B_n^* - \hat{B}_n = (S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz}) n^{1/2}\hat{\delta}_n.$$

Adding and subtracting appropriately, we can write this difference as

$$B_n^* - \hat{B}_n = n^{1/2}(S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz})\delta + (S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz}) n^{1/2}(\hat{\delta}_n - \delta),$$

where $n^{1/2}(\hat{\delta}_n - \delta)$ is $O_p(1)$ by a central limit theorem and $S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz} = o_{p^*}(1)$, in probability, by (A.2.3). The first term in $B_n^* - \hat{B}_n$ can be written as

$$\begin{aligned} &S_{x^*x^*}^{-1} n^{1/2}(S_{x^*z^*} - S_{xz})\delta - S_{x^*x^*}^{-1} S_{xx}^{-1}(S_{x^*x^*} - S_{xx}) n^{1/2} S_{xz}\delta \\ &= \delta(\Sigma_{xx}^{-1}, -\Sigma_{xx}^{-2}\Sigma_{xz}) \begin{pmatrix} n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} + o_{p^*}(1), \end{aligned}$$

in probability, by application of (A.2.3) and Assumption A.2.1(ii). It follows that

$$\begin{aligned} T_n^* - \hat{B}_n &= S_{x^*x^*}^{-1} n^{1/2} S_{x\varepsilon}^* + \delta(\Sigma_{xx}^{-1}, -\Sigma_{xx}^{-2}\Sigma_{xz}) \begin{pmatrix} n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} + o_{p^*}(1) \\ &= (\Sigma_{xx}^{-1}, \Sigma_{xx}^{-1}\delta, -\Sigma_{xx}^{-2}\Sigma_{xz}\delta) S_n^* + o_{p^*}(1), \end{aligned}$$

in probability. The required result now follows from (A.2.4) because

$$(\Sigma_{xx}^{-1}, \Sigma_{xx}^{-1}\delta, -\Sigma_{xx}^{-2}\Sigma_{xz}\delta)\begin{pmatrix}\Sigma_s & 0\\ 0 & \Sigma_r\end{pmatrix}(\Sigma_{xx}^{-1}, \Sigma_{xx}^{-1}\delta, -\Sigma_{xx}^{-2}\Sigma_{xz}\delta)'$$
$$= \Sigma_{xx}^{-1}\Sigma_s\Sigma_{xx}^{-1} + d_r(\delta)'\Sigma_r d_r(\delta) = v^2 + \kappa^2,$$

which completes the proof upon noting that $\Sigma_s = \sigma^2\Sigma_{xx}$ implies $v^2 = \sigma^2\Sigma_{xx}^{-1}$. □

A.2.2 Ridge regression

Assumptions and notation

As in Fu and Knight (2000) we assume the following.

(i) $\varepsilon_t \sim$ i.i.d.$(0,\sigma^2)$; (ii) $\max_{t=1,\dots,n} x_t'x_t = o(n)$; (iii) S_{xx} is nonsingular for any n and converges to a positive definite matrix, Σ_{xx}; (iv) $\theta = \delta n^{-1/2}$; and (v) $n^{-1}c_n \to c_0 \geq 0$.

For the bootstrap we will also need the following.

Assumption A.2.2 holds with (ii) replaced by (ii') $\max_{t=1,\dots,n} x_t'x_t = o(n^{1/2})$ and with the additional condition (vi) $E\varepsilon_t^4 < \infty$.

Finally, we define

$$V = \sigma^2\begin{pmatrix} g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g & -c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g\\ -c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g & c_0^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g\end{pmatrix}$$

where $v^2 := v_{11}$, and it holds that

$$m^2 := \frac{v_{11}+v_{22}-2v_{12}}{v_{11}} = \frac{g'\Sigma_{xx}^{-1}g}{g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g}, \tag{A.2.5}$$

where the last equality is derived in the proof of Lemma 1.4.4.

Proofs of lemmas

Proof of Lemma 1.4.4 and derivation of (A.2.5). The result follows by showing that

$$\begin{pmatrix} T_n - B_n\\ \hat{B}_n - B_n\end{pmatrix} = \begin{pmatrix} g'\tilde{S}_{xx}^{-1}\\ -c_n n^{-1}g'\tilde{S}_{xx}^{-1}S_{xx}^{-1}\end{pmatrix} n^{1/2}S_{x\varepsilon} \xrightarrow{d} \begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \sim N(0,V), \quad V = (v_{ij}), \tag{A.2.6}$$

and that

$$T_n^* - B_n^* = T_n^* - \hat{B}_n + o_{p^*}(1) \xrightarrow{d^*}_p N(0, v^2). \tag{A.2.7}$$

To prove (A.2.6) we first notice that, since $c_n n^{-1} \to c_0$, under Assumption A.2.2 we have that $n^{1/2}S_{x\varepsilon} \to_d N(0, \sigma^2\Sigma_{xx})$ and hence

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} = (I_2 \otimes g'\tilde{S}_{xx}^{-1}) \begin{pmatrix} I_p \\ -n^{-1}c_n S_{xx}^{-1} \end{pmatrix} n^{1/2} S_{x\varepsilon}$$

$$\xrightarrow{d} (I_2 \otimes g'\tilde{\Sigma}_{xx}^{-1}) N \left(0, \sigma^2 \begin{pmatrix} \Sigma_{xx} & -c_0 I_p \\ -c_0 I_p & c_0^2 \Sigma_{xx}^{-1} \end{pmatrix}\right) \sim N(0, V), \quad \text{(A.2.8)}$$

$$V = \sigma^2 \begin{pmatrix} g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g & -c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g \\ -c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g & c_0^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g \end{pmatrix}.$$

This immediately implies that m^2 in (A.2.5) is given by

$$m^2 = \frac{g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g + 2c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g + c_0^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1}g}{g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g}. \quad \text{(A.2.9)}$$

The numerator of m^2 in (A.2.9) can be written as

$$g'\tilde{\Sigma}_{xx}^{-1}(\Sigma_{xx} + 2c_0 I_p + c_0^2\Sigma_{xx}^{-1})\tilde{\Sigma}_{xx}^{-1}g = g'\tilde{\Sigma}_{xx}^{-1}(\tilde{\Sigma}_{xx}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx})\tilde{\Sigma}_{xx}^{-1}g = g'\Sigma_{xx}^{-1}g,$$

and hence (A.2.5) follows.

To prove (A.2.7) we note that $T_n^* - \hat{B}_n = \xi_{1,n}^* + B_n^* - \hat{B}_n$, where $\xi_{1,n}^* := n^{1/2}g'\tilde{S}_{x^*x^*}^{-1}S_{x^*\varepsilon^*}$ and

$$\begin{aligned} B_n^* - \hat{B}_n &= -c_n n^{-1/2} g'\tilde{S}_{x^*x^*}^{-1}\hat{\theta}_n + c_n n^{-1/2} g'\tilde{S}_{xx}^{-1}\hat{\theta}_n \\ &= -c_n n^{-1} g'(\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1}) n^{1/2}(\hat{\theta}_n - \theta) - c_n n^{-1} g'(\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1})\delta, \end{aligned}$$

such that $B_n^* - \hat{B}_n \xrightarrow{p^*}_p 0$ if $\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1} \xrightarrow{p^*}_p 0$. Because $||\tilde{S}_{xx}^{-1}|| = O(1)$ under the stated assumptions, it follows that $||\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1}||$ has the same rate as $||\tilde{S}_{x^*x^*} - \tilde{S}_{xx}||$. Thus, $\tilde{S}_{x^*x^*} - \tilde{S}_{xx} = S_{x^*x^*} - S_{xx} = n^{-1}\sum_{t=1}^n x_t^* x_t^{*\prime} - E^*(x_t^* x_t^{*\prime}) \xrightarrow{p^*}_p 0$ by a straightforward application of Chebyshev's LLN using that $\max_t x_t' x_t = o(n^{1/2})$ by Assumption A.2.2(ii').

The proof is completed by showing that $\xi_{1,n}^*$ satisfies the bootstrap central limit theorem. By the above results it holds that $\xi_{1,n}^* = n^{1/2}g'\tilde{\Sigma}_{xx}^{-1}S_{x^*\varepsilon^*} + o_{p^*}(1)$, so it is only required to analyze the term $n^{1/2}g'\tilde{\Sigma}_{xx}^{-1}S_{x^*\varepsilon^*} = n^{1/2}S_{\tilde{x}^*\varepsilon^*}$, where $\tilde{x}_t^* := g'\tilde{\Sigma}_{xx}^{-1}x_t^*$. First, we have $E^*(n^{1/2}S_{\tilde{x}^*\varepsilon^*}) = g'\tilde{\Sigma}_{xx}^{-1}E^*(n^{1/2}S_{x^*\varepsilon^*}) = n^{1/2}g'\tilde{\Sigma}_{xx}^{-1}S_{x\hat{\varepsilon}} = 0$. Second, with $\tilde{x}_t := g'\tilde{\Sigma}_{xx}^{-1}x_t$,

$$\begin{aligned} \text{Var}^*(n^{1/2}S_{\tilde{x}^*\varepsilon^*}) &= n^{-1}\sum_{t=1}^n \tilde{x}_t^2\hat{\varepsilon}_t^2 = n^{-1}\sum_{t=1}^n \tilde{x}_t^2(\hat{\varepsilon}_t^2 - \sigma^2 + \sigma^2) \\ &= \sigma^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g + n^{-1}\sum_{t=1}^n \tilde{x}_t^2(\varepsilon_t^2 - \sigma^2) + o_p(1). \end{aligned}$$

Because ε_t is i.i.d. and $\tilde{x}_t^2$ is non-stochastic, a sufficient condition for $n^{-1}\sum_{t=1}^n \tilde{x}_t^2(\varepsilon_t^2 - \sigma^2) \to_p 0$ is that $\lambda_{\min}(\sum_{t=1}^n \tilde{x}_t^2) \to \infty$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the argument, and this is implied by $n^{-1}\sum_{t=1}^n \tilde{x}_t^2 \to g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g > 0$.

Third, we check Lindeberg's condition, where we set $s_n^2 := nS_{\tilde{x}\tilde{x}}$. For $\epsilon > 0$ it holds that

$$\frac{1}{s_n^2}\sum_{t=1}^n E^*(\tilde{x}_t^{*2}\varepsilon_t^{*2}\mathbb{I}_{\{|\tilde{x}_t^*\varepsilon_t^*|>\epsilon s_n\}}) = \frac{1}{S_{\tilde{x}\tilde{x}}}E^*(\tilde{x}_t^{*2}\varepsilon_t^{*2}\mathbb{I}_{\{(\tilde{x}_t^*\varepsilon_t^*)^2>\epsilon^2 nS_{\tilde{x}\tilde{x}}\}})$$
$$\leq \frac{1}{\epsilon^2 nS_{\tilde{x}\tilde{x}}^2}E^*(\tilde{x}_t^{*4}\varepsilon_t^{*4})$$
$$= \frac{1}{\epsilon^2 n^2 S_{\tilde{x}\tilde{x}}^2}\sum_{t=1}^n \tilde{x}_t^4\hat{\varepsilon}_t^4 \leq \frac{n^{-1}\max_t \tilde{x}_t^4}{\epsilon^2 S_{\tilde{x}\tilde{x}}^2}\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_t^4 \xrightarrow{p} 0$$

because $n^{-1}\max_t \tilde{x}_t^4 = o(1)$ and ε_t has bounded fourth-order moment. □

Proof of Lemma 1.4.5. The proof follows closely the proofs of Lemma 1.4.4 and is omitted for brevity. □

A.2.3 Nonparametric regression

Assumptions and notation

We impose the following conditions.

(i) $\varepsilon_t \sim$ i.i.d.$(0,\sigma^2)$; (ii) $E|\varepsilon_t|^{2+\delta} < \infty$; (iii) $\beta : [0,1] \to \mathbb{R}$ is three times continuously differentiable with bounded derivatives; (iv) $K : \mathbb{R} \to [0,\infty)$ is symmetric and satisfies $K(u) = 0$ for all $u \notin (-1,1)$, $\int K(u)du = 1$, $\kappa^2 := \int u^2K(u)du \neq 0$, and $R_K := \int K(u)^2du \in (0,\infty)$.

Note that Assumption A.2.3 allows for the most popular choices of symmetric and truncated kernels.

To simplify notation, we define $k_t := K((x_t - x)/h)$ and $k_{tj} := K((x_t - x_j)/h)$. We also define the variance matrix

$$V := \begin{pmatrix} v^2 & \omega_{12} - v^2 \\ \omega_{12} - v^2 & v^2 + \omega_{22} - 2\omega_{12} \end{pmatrix},$$

where $v^2 := \sigma^2 R_K$, $\omega_{12} := \sigma^2\int K(u)\int K(s-u)K(s)dsdu$, and $\omega_{22} := \sigma^2\int(\int K(s-u)K(s)ds)^2du$.

Proofs of (1.2.4) and lemmas

Although it is well known (e.g., Li and Racine, 2007) that (1.2.4) and Assumption 1.3.1 hold in this example, we give short proofs for completeness.

Proof of (1.2.4). Under Assumption A.2.3 we obtain by Taylor expansion the following well-known result,

$$E\hat{\beta}_h(x) = \frac{1}{nh}\sum_{t=1}^n k_t\beta(x_t) = \int K(u)\beta(x+uh)du + o((nh)^{-1})$$
$$= \int K(u)\left(\beta(x) + \beta'(x)uh + \beta''(x)u^2h^2/2 + o(h^2)\right)du + o((nh)^{-1})$$
$$= \beta(x) + h^2\beta''(x)\kappa_2/2 + o(h^2) + o((nh)^{-1}), \tag{A.2.10}$$

where the last equality follows by $\int K(u)du = 1$ and $\int uK(u)du = 0$. Setting the bandwidth as $h = cn^{-1/5}$ thus implies (1.2.4). Note that the limits of integration are $u \in ((n^{-1}-x)/h, (1-x)/h)$, but for n sufficiently large this is the same as $u \in (-1,1)$ because $K(u) = 0$ for all $u \notin (-1,1)$. We use this property throughout the remaining proofs.

Proof of Lemma 1.4.6. First, we verify Assumption 1.3.1 by showing that $\xi_{1,n} := T_n - B_n = (nh)^{-1/2}\sum_{t=1}^{n} k_t\varepsilon_t$ satisfies the central limit theorem. Because $k_t\varepsilon_t, t = 1,\ldots,n$, is a sequence of independent random variables with mean zero and $\mathrm{Var}(k_t\varepsilon_t) = k_t^2\sigma^2$, we have

$$\begin{aligned}\mathrm{Var}(\xi_{1,n}) &= \frac{1}{nh}\sum_{t=1}^{n} k_t^2\sigma^2 = \frac{\sigma^2}{h}\int K\left(\frac{s-x}{h}\right)^2 ds + o((nh)^{-1})\\ &= \frac{\sigma^2}{h}\int K(u)^2 d(x+uh) + o((nh)^{-1}) \to \sigma^2 R_K = v^2.\end{aligned}$$

Moreover, Lyapunov's condition holds because

$$\begin{aligned}(nh)^{-(1+\delta)}\sum_{t=1}^{n} E(k_t^{2+\delta}|\varepsilon_t|^{2+\delta}) &\le c(nh)^{-(1+\delta)}\sum_{t=1}^{n} k_t^{2+\delta}\\ &\le c(nh)^{-(1+\delta)}\sum_{t:|x_t-x|\le h} k_t^{2+\delta} \le c(nh)^{-(1+\delta)}hn \to 0. \quad \text{(A.2.11)}\end{aligned}$$

Next, we verify Assumption 1.3.1(i). Note that $T_n^* - \hat{B}_n = (nh)^{-1/2}\sum_{t=1}^{n} k_t\varepsilon_t^* =: \xi_{1,n}^*$, where, conditional on D_n, $\xi_{1,n}^* \sim N(0, \hat{\sigma}_n^2(nh)^{-1}\sum_{t=1}^{n} k_t^2)$. Hence, the result follows from $\hat{\sigma}_n^2 \to_p \sigma^2$ and $(nh)^{-1}\sum_{t=1}^{n} k_t^2 \to R_K$.

Finally, we verify Assumption 1.3.1(ii). We first show that we can write

$$\hat{B}_n - B_n = \xi_{2,n} + o(1), \quad \xi_{2,n} := \frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\Big(\frac{1}{nh}\sum_{j=1}^{n} k_j k_{tj} - k_t\Big)\varepsilon_t, \quad \text{(A.2.12)}$$

and then we show that

$$\xi_n := (\xi_{1,n}, \xi_{2,n})' \xrightarrow{d} N(0, V). \quad \text{(A.2.13)}$$

To prove (A.2.12) we write

$$\hat{B}_n - B_n = (nh)^{1/2}\left(\frac{1}{nh}\sum_{t=1}^{n} k_t(\hat{\beta}_h(x_t) - \beta(x_t)) - (\hat{\beta}_h(x) - \beta(x))\right),$$

where

$$\begin{aligned}\hat{\beta}_h(x) &= (nh)^{-1}\sum_{t=1}^{n} k_t\beta(x_t) + (nh)^{-1}\sum_{t=1}^{n} k_t\varepsilon_t,\\ k_t\hat{\beta}_h(x_t) &= (nh)^{-1}\sum_{j=1}^{n} k_t k_{tj}\beta(x_j) + (nh)^{-1}\sum_{j=1}^{n} k_t k_{tj}\varepsilon_j.\end{aligned}$$

By reversing the summations and exploiting symmetry of k_{tj}, it immediately follows that

$\hat{B}_n - B_n = \xi_{2,n} + B_{2,n} - B_{1,n}$ with

$$B_{1,n}(x) := B_n = (nh)^{1/2}\left((nh)^{-1}\sum_{t=1}^n k_t\beta(x_t) - \beta(x)\right),$$

$$B_{2,n}(x) := (nh)^{-1}\sum_{t=1}^n k_t B_{1,n}(x_t) = (nh)^{-1/2}\sum_{t=1}^n k_t\left((nh)^{-1}\sum_{j=1}^n k_{tj}\beta(x_j) - \beta(x_t)\right).$$

By (A.2.10) we find

$$\begin{aligned} B_{2,n}(x) - B_{1,n}(x) &= (nh)^{-1/2}\sum_{t=1}^n k_t\left(h^2\beta''(x_t)\kappa_2/2 + o(h^2) + o((nh)^{-1})\right) \\ &\quad - (nh)^{1/2}h^2\beta''(x)\kappa_2/2 + o(h^2) + o((nh)^{-1}) \\ &= (nh)^{1/2}\frac{\kappa_2}{2}h^2\left(\frac{1}{nh}\sum_{t=1}^n k_t\beta''(x_t) - \beta''(x)\right) + o((nh)^{1/2}h^2) + o((nh)^{-1/2}), \end{aligned}$$

where

$$\frac{1}{nh}\sum_{t=1}^n k_t\beta''(x_t) = \int \beta''(x+uh)K(u)du + o((nh)^{-1}) = \beta''(x) + O(h) + o((nh)^{-1})$$

by first-order Taylor expansion, similar to (A.2.10), together with the assumption of continuous and bounded β'''. The result now follows because $h = cn^{-1/5}$.

Finally, to prove (A.2.13) we show that

$$J_n = \frac{1}{\sqrt{nh}}\sum_{t=1}^n \begin{pmatrix} k_t \\ (nh)^{-1}\sum_{j=1}^n k_j k_{tj} \end{pmatrix}\varepsilon_t \xrightarrow{d} N(0,\Omega), \quad \Omega := (\omega_{ij})_{i,j=1,2}, \tag{A.2.14}$$

from which the result follows by noting that $v^2 = \omega_{11}$ and

$$\xi_n = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} J_n.$$

It is clear that J_n has mean zero and independent increments. Approximating summations by integrals, it can be straightforwardly shown that

$$\mathrm{Var}(J_n) = \sigma^2\begin{bmatrix} (nh)^{-1}\sum_{t=1}^n k_t^2 & (nh)^{-2}\sum_{t,j=1}^n k_t k_j k_{tj} \\ (nh)^{-2}\sum_{t,j=1}^n k_t k_j k_{tj} & (nh)^{-1}\sum_{t=1}^n((nh)^{-1}\sum_{j=1}^n k_j k_{tj})^2 \end{bmatrix} \to \Omega.$$

By the same proof as in (A.2.11), we can show that the Lyapunov condition is satisfied, and result (A.2.14) follows. □

PROOF OF LEMMA 1.4.7. We first verify Assumption 1.3.2(i). We notice that $\bar{T}_n^{**} - \bar{B}_n^* = T_n^{**} - \hat{B}_n^* = (nh)^{-1/2}\sum_{t=1}^n k_t\varepsilon_i^{**} =: \xi_{1,n}^{**}$, where, conditional on (D_n, D_n^*), $\xi_{1,n}^{**} \sim N(0, \hat{\sigma}_n^{*2}(nh)^{-1}\sum_{t=1}^n k_t^2)$. Hence, the result follows from $\hat{\sigma}_n^{*2} \xrightarrow{p^*}_p \sigma^2$ and $(nh)^{-1}\sum_{t=1}^n k_t^2 \to R_K$.

Next, we verify Assumption 1.3.2(ii). We first write $T_n^* - \bar{B}_n^* = T_n^* - \hat{B}_n - (\bar{B}_n^* - \hat{B}_n) =$

$\xi^*_{1,n} - (\bar{B}^*_n - \hat{B}_n)$, where $\bar{B}^*_n - \hat{B}_n = \hat{B}^*_n - \hat{B}_{2,n}$. Recall $\hat{B}_{2,n} := (nh)^{-1}\sum_{t=1}^n k_t \hat{B}_n(x_t) = (nh)^{-1/2}\sum_{t=1}^n k_t((nh)^{-1}\sum_{j=1}^n k_{tj}\hat{\beta}_h(x_j) - \hat{\beta}_h(x_t))$ and $\hat{B}^*_n := (nh)^{1/2}((nh)^{-1}\sum_{t=1}^n k_t \hat{\beta}^*_h(x_t) - \hat{\beta}^*_h(x))$, where

$$\hat{\beta}^*_h(x) = (nh)^{-1}\sum_{t=1}^n k_t\hat{\beta}_h(x_t) + (nh)^{-1}\sum_{t=1}^n k_t\varepsilon^*_t,$$

$$k_t\hat{\beta}^*_h(x_t) = (nh)^{-1}\sum_{j=1}^n k_t k_{tj}\hat{\beta}_h(x_j) + (nh)^{-1}\sum_{j=1}^n k_t k_{tj}\varepsilon^*_j,$$

so it follows that

$$\bar{B}^*_n - \hat{B}_n = \hat{B}^*_n - \hat{B}_{2,n} = \xi^*_{2,n}, \quad \xi^*_{2,n} := \frac{1}{\sqrt{nh}}\sum_{t=1}^n (\frac{1}{nh}\sum_{j=1}^n k_j k_{tj} - k_t)\varepsilon^*_t.$$

Thus, the proof is completed by showing that

$$\xi^*_n := (\xi^*_{1,n}, \xi^*_{2,n})' \xrightarrow{d^*}_p N(0, V). \tag{A.2.15}$$

Conditional on D_n, it holds that $\xi^*_n \sim N(0, \hat{V}_n)$, where

$$\hat{V}_n = \hat{\sigma}^2_n \frac{1}{nh}\sum_{t=1}^n \begin{bmatrix} k_t^2 & k_t(\frac{1}{nh}\sum_{j=1}^n k_j k_{tj} - k_t) \\ k_t(\frac{1}{nh}\sum_{j=1}^n k_j k_{tj} - k_t) & (\frac{1}{nh}\sum_{j=1}^n k_j k_{tj} - k_t)^2 \end{bmatrix} \xrightarrow{p} V$$

by approximating the summations by integrals and using $\hat{\sigma}^2_n \to_p \sigma^2$. This proves (A.2.15) and hence completes the proof of Lemma 1.4.7. □

A.2.4 Inference under heavy tails

Setup. We consider a simple location model with heavy-tailed data, thus demonstrating that our analysis applies to a non-Gaussian asymptotic framework. Specifically, consider a sample of n i.i.d. random variables $\{y_t\}$. Interest is in inference on θ in the location model

$$y_t = \theta + \varepsilon_t, \quad E(\varepsilon_t) = 0,$$

when the ε_t's follow a symmetric, stable random variable $S(\alpha)$ with tail index $\alpha \in (1,2)$ and the location parameter is local to zero; i.e., $\theta = n^{1/\alpha - 1}c$.[2] Under these assumptions, $E(|\varepsilon_t|^{\alpha+\delta}) = +\infty$ for any $\delta \geq 0$; in particular, ε_t has infinite variance. Notice that θ is local of order $n^{1/\alpha-1}$ rather than the usual $n^{-1/2}$ because of the slower convergence rate of the OLS-type estimator when the variance of ε_t is infinite. We consider the biased estimator

$$\hat{\theta}_n := \omega\bar{y}_n, \quad \bar{y}_n := n^{-1}\sum_{t=1}^n y_t,$$

[2]The results in this section can easily be generalized to the case where the ε_t's are not necessarily symmetric and/or are in the domain of attraction of a stable law with index $\alpha \in (0,1)$, as in Cornea-Madeira and Davidson (2015). Moreover, the results apply to the case of non-local θ as well; i.e., $\theta \neq 0$ fixed.

where $\omega \in (0,1)$. In the finite variance case, this estimator improves upon $\bar{y}_n$ in terms of MSE when θ is local to zero. It holds that

$$T_n := n^{1-1/\alpha}(\hat{\theta}_n - \theta) = (\omega - 1)c + \omega n^{1-1/\alpha}\bar{\varepsilon}_n \sim B + \omega S(\alpha) \tag{A.2.16}$$

with $B := (\omega - 1)c$; equivalently, $T_n - B \sim \xi_1 := \omega S(\alpha)$. Hence, Assumption 1.3.1 is satisfied with $G_\gamma(u) = P(\omega S(\alpha) \leq u) = \Psi_\alpha(\omega^{-1}u)$, where $\Psi_\alpha(u) := P(S(\alpha) \leq u)$ is continuous. Inference based on quantiles of ξ_1 is invalid because it misses the term B.

Bootstrap. It is well known that the standard bootstrap fails to be valid under infinite variance (Knight, 1989). The 'm out of n' bootstrap (see Politis et al., 1999, and the references therein) is an attractive option, but it fails to mimic the non-centrality parameter B; see Remark A.2.2 below. Instead, we consider the parametric bootstrap of Cornea-Madeira and Davidson (2015), which only requires a consistent estimator $\hat{\alpha}_n$ of the tail index α, assumed to lie in a compact set. The bootstrap sample is generated as

$$y_t^* = \bar{y}_n + \varepsilon_t^*, \quad \varepsilon_t^* \sim \text{ i.i.d.} S(\hat{\alpha}_n),$$

and the bootstrap estimator is $\hat{\theta}_n^* := \omega \bar{y}_n^* = \omega(\bar{y}_n + \bar{\varepsilon}_n^*)$ with $\bar{\varepsilon}_n^* := n^{-1}\sum_{t=1}^n \varepsilon_t^*$. The bootstrap analogue of T_n then satisfies

$$T_n^* := n^{1-1/\alpha}(\hat{\theta}_n^* - \bar{y}_n) = \omega n^{1-1/\alpha}\bar{\varepsilon}_n^* + \hat{B}_n \text{ with } \hat{B}_n := (\omega - 1)n^{1-1/\alpha}\bar{y}_n.$$

Now, $n^{1-1/\alpha}\bar{\varepsilon}_n^* \xrightarrow{d^*}_p S(\alpha)$ by Proposition 1 in Cornea-Madeira and Davidson (2015) and, therefore,

$$T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1 := \omega S(\alpha).$$

This shows that Assumption 1.3.1(i) is satisfied in this example. Notice that the bias term in the bootstrap world satisfies, jointly with (A.2.16),

$$\hat{B}_n - B = (\omega - 1)n^{1-1/\alpha}\bar{\varepsilon}_n \sim (\omega - 1)S(\alpha) =: \xi_2.$$

Specifically, because both T_n and $\hat{B}_n$ depend on the data through $\bar{\varepsilon}_n$ only, we have that $(\xi_1, \xi_2) \sim (\omega, \omega - 1)S(\alpha)$, implying that $\xi_1 - \xi_2 \sim S(\alpha)$. Hence, Assumption 1.3.1(ii) is satisfied with $F_\phi(u) = P(S(\alpha) \leq u) = \Psi_\alpha(u)$. Since the cdf of $\xi_1 \sim \omega S(\alpha)$ can be written as $G_\gamma(u) = \Psi_\alpha(\omega^{-1}u)$, it follows by Theorem 1.3.1 that $\hat{p}_n \to_d G_\gamma(F_\phi^{-1}(U_{[0,1]})) = \Psi_\alpha(\omega^{-1}\Psi_\alpha^{-1}(U_{[0,1]}))$ and, therefore,

$$P(\hat{p}_n \leq u) \to H(u) := P(\Psi_\alpha(\omega^{-1}\Psi_\alpha^{-1}(U_{[0,1]})) \leq u) = \Psi_\alpha(\omega\Psi_\alpha^{-1}(u)),$$

which differs from u unless $\omega = 1$.

Because ω is known and we can estimate α consistently with $\hat{\alpha}_n$, we can estimate $H(u)$ consistently with $\hat{H}_n(u) := \Psi_{\hat{\alpha}_n}(\omega\Psi_{\hat{\alpha}_n}^{-1}(u))$ and obtain a valid plug-in modified p-value,

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Psi_{\hat{\alpha}_n}(\omega\Psi_{\hat{\alpha}_n}^{-1}(\hat{p}_n)),$$

by application of Corollary 1.3.2.

Alternatively, we can estimate $H(u)$ using the double bootstrap estimator $\hat{H}_n(u) := P^*(\hat{p}_n^{**} \leq u)$, where $\hat{p}_n^{**} := P^{**}(T_n^{**} \leq T_n^*)$. Specifically, let the double bootstrap sample $\{y_t^{**}\}$ be generated as

$$y_t^{**} = \bar{y}_n^* + \varepsilon_t^{**}, \quad \varepsilon_t^{**} \sim \text{ i.i.d.} S(\hat{\alpha}_n),$$

and set $\hat{\theta}_n^{**} := \omega \bar{y}_n^{**} = \omega \bar{y}_n^* + \omega \bar{\varepsilon}_n^{**}$, where $\bar{\varepsilon}_n^{**} := n^{-1} \sum_{t=1}^n \varepsilon_t^{**}$. The (second-level) bootstrap analogue of T_n^* then satisfies

$$T_n^{**} := n^{1-1/\alpha}(\hat{\theta}_n^{**} - \bar{y}_n^*) = \omega n^{1-1/\alpha} \bar{\varepsilon}_n^{**} + \hat{B}_n^* \text{ with } \hat{B}_n^* := (\omega - 1) n^{1-1/\alpha} \bar{y}_n^*.$$

Since ε_t^{**} is generated from $S(\hat{\alpha}_n)$, where $\hat{\alpha}_n$ depends only on D_n, the distribution of ε_t^{**}, conditionally on D_n^* and D_n, is the same as the distribution of ε_t^*, conditionally on D_n. This implies that

$$n^{1-1/\alpha} \bar{\varepsilon}_n^{**} \xrightarrow{d^{**}}_{p*} S(\alpha),$$

in probability, by Proposition 1 of Cornea-Madeira and Davidson (2015). Therefore,

$$T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p*} \xi_1 = \omega S(\alpha),$$

in probability, showing that Assumption 1.3.2(i) is satisfied. Since

$$\hat{B}_n^* - \hat{B}_n = (\omega - 1) n^{1-1/\alpha}(\bar{y}_n^* - \bar{y}_n) = (\omega - 1) n^{1-1/\alpha} \bar{\varepsilon}_n^*$$

and $T_n^* - \hat{B}_n = \omega n^{1-1/\alpha} \bar{\varepsilon}_n^*$, Assumption 1.3.2(ii) is also satisfied in this example. Thus, $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \rightarrow_d U_{[0,1]}$ by Theorem 1.3.2.

Remark A.2.2 *Consider the 'm out of n' bootstrap data generating process,*

$$y_t^* = \bar{y}_n + \varepsilon_t^*, \quad t = 1, \ldots, m,$$

where ε_t^* is an i.i.d. sample from the residuals $\hat{\varepsilon}_t = y_t - \bar{y}_n$, $t = 1, \ldots, n$. Then, with $\hat{\theta}_m^* := \omega \bar{y}_m^*$, $\bar{y}_m^* := m^{-1} \sum_{t=1}^m y_t^*$, the 'm out of n' bootstrap statistic is

$$T_m^* := m^{1-1/\alpha}(\hat{\theta}_m^* - \bar{y}_n) = \omega m^{1-1/\alpha} \bar{\varepsilon}_m^* + (\omega - 1) m^{1-1/\alpha} \bar{y}_n,$$

where $m^{1-1/\alpha} \bar{\varepsilon}_m^* \xrightarrow{d^*}_p S(\alpha)$ as $m \rightarrow \infty$; see Arcones and Giné (1989). Moreover, if $m = o(n)$,

$$\hat{B}_m := (\omega - 1) m^{1-1/\alpha} \bar{y}_n = (\omega - 1) m^{1-1/\alpha} n^{1/\alpha - 1}(n^{1-1/\alpha} \bar{y}_n) = O_p((m/n)^{1/\alpha - 1}) = o_p(1),$$

which shows that $T_m^* \xrightarrow{d^*}_p \omega S(\alpha)$. Hence, Assumption 1.3.1(i) is satisfied with $\xi_1 := \omega S(\alpha)$ and $\hat{B}_n = 0$. Since $B := (\omega - 1)c \neq 0$, we have $\xi_2 := -B$ a.s., so that Assumption 1.3.1(ii) does not hold. As in Remark 1.3.1, it then follows that

$$\hat{p}_m := P^*(T_m^* \leq T_n) \xrightarrow{d} G_\gamma(G_\gamma^{-1}(U_{[0,1]}) - B)) = \Psi_\alpha(\Psi_\alpha^{-1}(U_{[0,1]}) - B).$$

This shows that the limiting distribution of $\hat{p}_m$ depends on B. Since B cannot be consistently estimated, the 'm out of n' bootstrap cannot be used to solve the problem.

A.2.5 Nonlinear dynamic panel data models with incidental parameter bias

Another example that fits our framework is inference based on panel data estimators subject to incidental parameter bias. We consider the properties of the cross-sectional pairs bootstrap considered by Kaffo (2014), Dhaene and Jochmans (2015), and Gonçalves and Kaffo (2015) in the context of a general nonlinear panel data model. Although this bootstrap cannot replicate the bias, we show that our prepivoting approach based on a plug-in estimator of the bias is valid. Recently, Higgins and Jochmans (2022) proposed a (double) bootstrap procedure that retains asymptotic validity without an explicit plug-in estimator of the bias, but their procedure relies heavily on the parametric distribution assumption.

Setup. Let z_{it} denote a vector of random variables for a set of n individuals, $i = 1, \ldots, n$, over T time periods, $t = 1, \ldots, T$. Given a model for the density function $f_{it}(\theta, \alpha_i) := f(z_{it}, \theta, \alpha_i)$, the parameter of interest is $\theta \in \Theta$, which is common to all the individuals, while $\alpha_i \in \mathcal{A}$ denote the individual fixed effects. The fixed effects estimator of θ is the maximum likelihood estimator defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \sum_{t=1}^{T} \log f_{it}(\theta, \hat{\alpha}_i(\theta)), \text{ where } \hat{\alpha}_i(\theta) = \arg\max_{\alpha_i \in \mathcal{A}} \sum_{t=1}^{T} \log f_{it}(\theta, \alpha_i). \quad \text{(A.2.17)}$$

Under certain regularity conditions (see, e.g., Hahn and Kuersteiner, 2011), including letting $n, T \to \infty$ jointly such that $n/T \to \rho < \infty$,

$$T_n := \sqrt{nT}(\hat{\theta}_n - \theta) \xrightarrow{d} N(B, v^2), \quad \text{(A.2.18)}$$

where B denotes the incidental parameter bias and v^2 is the asymptotic variance of $\hat{\theta}_n$. Hence, Assumption 1.3.1 is satisfied with $\xi_1 \sim N(0, v^2)$ (equivalently, Assumption A.1 is satisfied).

The exact forms of B and v^2 may be quite involved and depend on the type of heterogeneity and dependence assumptions imposed on z_{it}. A standard assumption is that z_{it} is independent across i while allowing for time series dependence of unknown form; see Hahn and Kuersteiner (2011).

Bootstrap. Given the cross sectional independence assumption, a natural bootstrap method in this context is the cross sectional pairs bootstrap. The idea is to resample $z_i = (z_{i1}, \ldots, z_{iT})'$ in an i.i.d. fashion in the cross sectional dimension. If $z_{it} = (y_{it}, x_{it})'$ and $f(z_{it}, \theta, \alpha_i) = f(y_{it}|x_{it}, \theta, \alpha_i)$ is the conditional density of y_{it} given x_{it}, this is equivalent to a cross sectional pairs bootstrap. As the results of Kaffo (2014, Theorem 3.1) show, this bootstrap fails to capture the bias term B. In particular, letting $\hat{\theta}_n^*$ denote the bootstrap analogue of $\hat{\theta}_n$, we have that

$$T_n^* := \sqrt{nT}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*}_p N(0, v^2),$$

which implies that, as in Remarks 1.3.1 and A.2.2,

$$\hat{p}_n := P^*(T_n^* \leq T_n) = \Phi(v^{-1}T_n) + o_p(1) \xrightarrow{d} \Phi(v^{-1}B + \Phi^{-1}(U_{[0,1]})).$$

Thus,

$$P(\hat{p}_n \leq u) \to H(u) := P(\Phi(\Phi^{-1}(U_{[0,1]}) + v^{-1}B) \leq u) = \Phi(\Phi^{-1}(u) - v^{-1}B),$$

which shows that the bootstrap test based on $\hat{p}_n$ is asymptotically invalid since its limiting distribution is not uniform.

Remark A.2.3 *Note that, in this example, $\hat{L}_n(u) := P^*(T_n^* \leq u) \to_p \Phi(u/v)$, showing that the bootstrap conditional distribution of T_n^* is not random in the limit. The invalidity of $\hat{p}_n$ is due to the fact that the cross sectional pairs bootstrap induces $\hat{B}_n = 0$, whereas $B \neq 0$. This implies that $\hat{B}_n - B = -B := \xi_2$ is not random. The fact that ξ_2 is not zero is the cause of the bootstrap invalidity. See Remark 1.3.1, which contains this example as a special case.*

Contrary to previous examples (e.g., Remark A.2.2), B and v can both be consistently estimated. Hence, in this example we can restore bootstrap validity by modifying the bootstrap p-value using a plug-in approach. More specifically, let $\tilde{B}_n$ and $\hat{v}_n$ denote consistent estimators of B and v, respectively.[3] By Corollary 1.3.2,

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Phi(\Phi^{-1}(\hat{p}_n) - \hat{v}_n^{-1}\tilde{B}_n) \xrightarrow{d} U_{[0,1]}$$

because $\hat{H}_n(u) := \Phi(\Phi^{-1}(u) - \hat{v}_n^{-1}\tilde{B}_n)$ is a consistent estimator of $H(u)$.

Remark A.2.4 *A double bootstrap modified p-value version of $\tilde{p}_n$ is not valid in this setting. The reason is that the double bootstrap mimics the behavior of the first-level bootstrap, i.e.*

$$T_n^{**} := \sqrt{nT}(\hat{\theta}_n^{**} - \hat{\theta}_n^*) \xrightarrow{d^{**}}_p N(0, v^2),$$

so that $\hat{B}_n^*$ in Assumption 1.3.2(i) is zero. Since $\hat{B}_n = 0$, Assumption 1.3.2(ii) holds with $\hat{B}_n^* - \hat{B}_n = 0$, whereas Assumption 1.3.1(ii) has $\hat{B}_n - B_n = -B$ a.s. Then,

$$\hat{p}_n^* = P^{**}(v^{-1}T_n^{**} \leq v^{-1}T_n^*) = \Phi(v^{-1}T_n^*) \xrightarrow{d^*}_p \Phi(\Phi^{-1}(U_{[0,1]})) = U_{[0,1]},$$

[3]Since we reserve the notation $\hat{B}_n$ for the bootstrap-induced bias estimator (which is zero for the cross sectional pairs bootstrap), we use the notation $\tilde{B}_n$ to denote any consistent estimator of B in this setup. For instance, $\tilde{B}_n$ could be the plug-in estimator proposed by Hahn and Kuersteiner (2011), which is based on a closed-form expression of B_1. Another option is the half-split panel jackknife estimator of Dhaene and Jochmans (2015).

whereas

$$\hat{p}_n \xrightarrow{d} \Phi(\Phi^{-1}(U_{[0,1]}) + v^{-1}B).$$

Thus, $\hat{H}_n(u) := P^*(\hat{p}_n^* \leq u)$ *is not a consistent estimator of* $H(u)$*, invalidating* $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$*.*

Remark A.2.5 *A special case of the previous setup is a linear panel dynamic model, where* $z_{it} = (y_{it}, x'_{it})'$ *and* x_{it} *is a vector containing lags of* y_{it} *(Hahn and Kuersteiner, 2002). In this case, the plug-in modified p-value,* $\tilde{p}_n$*, based on the cross sectional pairs bootstrap can be implemented using any consistent estimator of* B*, as described above. However, we can also use a recursive bootstrap that exploits the linearity of the model to obtain an asymptotically valid standard bootstrap p-value,* $\hat{p}_n$*. The validity of* $\hat{p}_n$ *follows from the fact that the recursive bootstrap estimates* B *consistently, contrary to the pairs bootstrap (Gonçalves and Kaffo, 2015). In light of this, prepivoting* $\hat{p}_n$ *by computing a double bootstrap modified p-value* $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$ *is not needed in this example, but it is still a valid alternative.*

Appendix B

Appendix to Chapter 2

B.1 "Modified" prepivoting: high-level conditions

Let us now consider a further generalization of the high-level assumptions in Cavaliere et al. (2024) which allow to avoid the issues discussed in Section 3.1. Let T_n and T_n^* be the asymptotic and bootstrap statistic, with "bias terms" $B_{1,n}$ and $\hat{B}_n$, respectively, defined as general functions of the samples D_n and D_n^*, respectively. We here show that prepivoting can be applied to obtain valid p-values even in cases in which $T_n - B_{1,n}$ is asymptotically centered at zero but $\hat{B}_n - B_{1,n}$ is not. This is done via proper modifications of T_n^* which still do not require estimation of $B_{1,n}$.

Assumption B.1.1 *$T_n - B_n \xrightarrow{d} \xi_1$, where ξ_1 is centered at zero and the cdf $G(u) := \mathbb{P}(\xi_1 \leq u)$ is continuous and strictly increasing over its support.*

Assumption B.1.1 is analogous to Assumption 1 in Cavaliere et al. (2024). The main difference with the setup in Cavaliere et al. (2024) is given by the introduction of a second "bias term" $B_{2,n}$ which is asymptotically different from $B_{1,n}$ and such that $\hat{B}_n - B_{2,n}$ is asymptotically centered at zero.

Assumption B.1.2 *For some D_n-measurable random variable $\hat{B}_n$, it holds that: (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1$, where ζ_1 is centered at zero and the cdf $J(u) := \mathbb{P}(\zeta_1 \leq u)$; (ii)*

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 0 \\ \phi \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

where ξ_1 and ξ_2 are both centered at zero and the cdf $F(u) := \mathbb{P}(\xi_1 - \xi_2 \leq u)$ is continuous.

The setup embedded in Assumption B.1.1 and B.1.2 is a generalization of the setup considered in Cavaliere et al. (2024) – specifically, to the conditions of Theorem 3.4 – which allow $B_{1,n}$ and $B_{2,n}$ not to be the same quantity, not even asymptotically. In case we have that $B_{1,n} - B_{2,n} \to 0$, then the conditions of Theorem 3.4 in Cavaliere et al. (2024)

hold and "standard" prepivoting can be applied. However, if $B_{1,n} - B_{2,n} \nrightarrow 0$ Assumption 5 in Cavaliere et al. (2024) breaks down and the results in Theorem 3.4 are not valid.

We here focus the attention on situations in which $B_{1,n}$ and $B_{2,n}$ are hard or impossible to estimate, but their ratio is measurable or more easily estimable. As we will show in Section 3.3, this is the case in the setup of local polynomial estimation at the boundary of the design space.

Assumption B.1.3 *Suppose $Q := plim_{n\to\infty}\{B_{1,n}/B_{2.n}\}$ with $|Q| \in (0,\infty)$ is D_n-measurable.*

Assumption B.1.3 formalizes measurability of the limit of the ratio between $B_{1,n}$ and $B_{2.n}$. Crucially, the above condition rules out the fact that either $B_{1,n} = 0$ or $B_{2,n} = 0$ (the latter being the case, for instance, for least squares linear regression with omitted variable bias).

Let us consider the modified bootstrap test statistic $\tilde{T}_n^* = QT_n^*$; its "bias term" becomes $\tilde{B}_n := Q\hat{B}_n$. The aim of this modification is to make $\tilde{B}_n - B_{1,n}$ asymptotically centered at zero; in fact,

$$\begin{aligned}\tilde{B}_n - B_{1,n} &= Q\hat{B}_n - B_{1,n} = Q(\hat{B}_n - B_{2,n}) + QB_{2,n} - B_{1,n} \\ &= Q(\hat{B}_n - B_{2,n}) + o_p(1) \xrightarrow{d} Q\xi_2 =: \tilde{\xi}_2\end{aligned}$$

where the last equality is given by the fact that $QB_{2,n} - B_{1,n} = \left(\text{plim}_{n\to\infty}\left\{\frac{B_{1,n}}{B_{2,n}}\right\} \cdot \frac{B_{2,n}}{B_{1,n}} - 1\right) B_{1,n} = o_p(1)$. Moreover, note that $\tilde{\xi}_2$ is centered at zero since ξ_2 is centered at zero.

Theorem B.1.1 *Under Assumptions A.1–A.3, it holds that: (i) $\tilde{T}_n^* - \tilde{B}_n \xrightarrow{d^*}_p Q\zeta_1 =: \tilde{\zeta}_1$ where $\tilde{\zeta}_1$ is centered at zero and the cdf $\tilde{J}(u) := \mathbb{P}(\tilde{\zeta}_1 \leq u)$; (ii)*

$$\begin{pmatrix} T_n - B_{1,n} \\ \tilde{B}_n - B_{1,n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \tilde{\xi}_2 \end{pmatrix}$$

where ξ_1 and $\tilde{\xi}_2$ are both centered at zero and the cdf $\tilde{F}(u) := \mathbb{P}(\xi_1 - \tilde{\xi}_2 \leq u)$ is continuous.

Theorem B.1.1 shows that considering the modified bootstrap test statistic allows the application of prepivoting since $\tilde{\zeta}_1$, ξ_1 and $\tilde{\xi}_2$ are all centered at zero and do not depend on the "bias terms". Specifically, Theorem B.1.1 states that the conditions of Theorem 3.4 in CGNZ hold under Assumptions A-C.

Suppose now that Q_n is not observable by the researcher but a consistent estimator of Q exists. Then a result analogous to Theorem 3 can be derived when Assumption B.1.3 is replaced by Assumption B.1.4 below.

Assumption B.1.4 *(i) For a sequence r_n such that $r_n \to 0$, suppose there exists an estimator $\hat{Q}_n$ such that $\hat{Q}_n - Q = O_p(r_n)$, where $|Q| \in (0,\infty)$; (ii) $B_{2,n} = O_p\left(r_n^{\delta-1}\right)$ for some $\delta > 0$.*

Let us now define the modified bootstrap test statistic $\check{T}_n^* := \hat{Q}_n T_n^*$, with "bias term" given by $\check{B}_n^* := \hat{Q}_n \hat{B}_n$; then, we have that

$$\check{B}_n - B_{1,n} = \hat{Q}_n \hat{B}_n - B_{1,n} = Q\hat{B}_n - B_{1,n} + (\hat{Q}_n - Q)\hat{B}_n,$$

where $Q\hat{B}_n - B_{1,n} \xrightarrow{d} \tilde{\xi}_2$ by the same arguments in Theorem 3, and

$$(\hat{Q}_n - Q)\hat{B}_n = (\hat{Q}_n - Q)(\hat{B}_n - B_{2,n}) + (\hat{Q}_n - Q)B_{2,n} = o_p(1).$$

where the last equality is given by the fact that $(\hat{Q}_n - Q)(\hat{B}_n - B_{2,n}) = o_p(1)O_p(1) = o_p(1)$ and $(\hat{Q}_n - Q)B_{2,n} = O_p(r_n^\delta) = o_p(1)$.

THEOREM B.1.2 *Under Assumptions A.1–A.3 it holds that: (i)* $\check{T}_n^* - \check{B}_n \xrightarrow{d^*}_p Q\zeta_1 =: \tilde{\zeta}_1$ *where* $\tilde{\zeta}_1$ *is centered at zero and the cdf* $\tilde{J}(u) := \mathbb{P}(\tilde{\zeta}_1 \leq u)$*; (ii)*

$$\begin{pmatrix} T_n - B_{1,n} \\ \check{B}_n - B_{1,n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \tilde{\xi}_2 \end{pmatrix}$$

where ξ_1 *and* $\tilde{\xi}_2$ *are both centered at zero and the cdf* $\tilde{F}(u) := \mathbb{P}(\xi_1 - \tilde{\xi}_2 \leq u)$ *is continuous.*

Theorem B.1.2 formalizes the validity of "modified" prepivoting which - analogously than for Theorem B.1.1 - implies that the conditions of Theorem 3.4 in Cavaliere et al. (2024) are satisfied.

B.2 ASYMPTOTIC VALIDITY OF THE FL BOOTSTRAP-BASED CIS

PROPOSITION B.2.1 *Let Assumptions 1-3 hold, then,*

$$v_{1n}^{-1}\xi_{1n}^* := v_{1n}^{-1}(T_n^* - \tilde{B}_n) \xrightarrow{d^*}_p N(0,1).$$

PROPOSITION B.2.2 *Let the conditions of Proposition 2.3.1 hold, then: (i)*

$$V_{FL,n}^{-1/2}\tilde{\xi}_n \xrightarrow{d} N(0, I_2); \tag{B.2.1}$$

(ii) moreover, if x *is an interior point,*

$$V_{FL,n} \xrightarrow{p} V_{FL}; \tag{B.2.2}$$

whereas if x *is a boundary point,*

$$V_{FL,n} \xrightarrow{p} \ddot{V}_{FL} \tag{B.2.3}$$

where $V_{FL} := (v_{ij,FL}^2)$ *and* $\ddot{V}_{FL} := (\ddot{v}_{ij,FL}^2)$*, with* $v_{2,FL}, \ddot{v}_{2,FL} > 0$ *and* V_{FL} *and* $\ddot{V}_{FL}$ *are defined in Appendix B.*

PROPOSITION B.2.3 *Let Assumptions 1-3 hold, then: (i) if* x *is an interior point,*

$$\hat{p}_n \xrightarrow{d} \Phi\left(m_{FL}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{B.2.4}$$

where $m_{FL} = \sqrt{v_{1,FL}^2 + v_{2,FL}^2 - 2v_{12,FL}}/v_{1,FL}$*; and (ii) if* x *is a boundary point,*

$$\hat{p}_n \xrightarrow{d} \Phi\left(\ddot{m}_{FL}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{B.2.5}$$

where $\ddot{m}_{FL} := \ddot{v}_{d,FL}/\ddot{v}_{1,FL} := \sqrt{\ddot{v}_{1,FL}^2 + \ddot{v}_{2,FL}^2 - 2\ddot{v}_{12,FL}}/\ddot{v}_{1,FL}$.

B.3 Proof of the Main Results

B.3.1 Proof of Proposition 3.1

The proof of Proposition 3.1 follows analogous steps as the proof of Proposition 3.2 and is thus omitted for brevity.

B.3.2 Proof of Proposition 3.2

We let $\varepsilon^* = (\varepsilon_1^*, ..., \varepsilon_n^*)'$ and note that:

$$T_n^* - \hat{B}_n = \sqrt{nh}e_1'(Z_x'W_xZ_x)^{-1}Z_x'W_x\varepsilon^*$$

Let us first focus on $Z_x'W_xZ_x$ and notice that:

$$\Gamma_{1n} := \frac{Z_x'W_xZ_x}{n} = \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right) & \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right) \\ \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right) & \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^2 \end{pmatrix}$$

So that, by Lemma B.4.1,

$$\Gamma_{1n} = \Gamma_1 + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is interior}$$

$$\Gamma_{1n} = \ddot{\Gamma}_1 + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is boundary}$$

Let us now consider the term $\sqrt{nh}Z_x'W_x\varepsilon^*/n$. We can notice that:

$$\begin{aligned} \sqrt{nh}Z_x'W_x\varepsilon^*/n &= \frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\begin{pmatrix} 1 \\ \frac{x_i-x}{h} \end{pmatrix}\varepsilon_i^* \\ &= \frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\begin{pmatrix} 1 \\ \frac{x_i-x}{h} \end{pmatrix}\varepsilon_i e_i^* + o_p(1) \end{aligned}$$

So that

$$\mathbb{E}^*\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\begin{pmatrix} 1 \\ \frac{x_i-x}{h} \end{pmatrix}\varepsilon_i e_i^*\right] = 0$$

and

$$\mathbb{E}^*\left[\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\begin{pmatrix} 1 \\ \frac{x_i-x}{h} \end{pmatrix}\varepsilon_i e_i^*\right)\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\begin{pmatrix} 1 \\ \frac{x_i-x}{h} \end{pmatrix}\varepsilon_i e_i^*\right)'\right] = hZ_{1x}'W_x\Sigma W_xZ_{1x}/$$

Moreover,

$$hZ'_{1x}W_x\Sigma W_x Z_{1x}/n = \Psi_n = \begin{pmatrix} \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i-x}{h}\right) & \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right) \\ \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right) & \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^2 \end{pmatrix}$$

where, by Lemma B.4.2

$$\Psi_{11n} = \Psi_{11} + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is interior}$$
$$\Psi_{11n} = \ddot{\Psi}_{11} + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is boundary}$$

such that

$$\Psi_{22} := \begin{pmatrix} \psi_0 & \psi_1 & \psi_2 \\ \psi_1 & \psi_2 & \psi_3 \\ \psi_2 & \psi_3 & \psi_4 \end{pmatrix} \qquad \ddot{\Psi}_{22} := \begin{pmatrix} \ddot{\psi}_0 & \ddot{\psi}_1 & \ddot{\psi}_2 \\ \ddot{\psi}_1 & \ddot{\psi}_2 & \ddot{\psi}_3 \\ \ddot{\psi}_2 & \ddot{\psi}_3 & \ddot{\psi}_4 \end{pmatrix}$$
$$\Gamma_1 := \begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_2 \end{pmatrix} \qquad \ddot{\Gamma}_1 := \begin{pmatrix} \ddot{\gamma}_0 & \ddot{\gamma}_1 \\ \ddot{\gamma}_1 & \ddot{\gamma}_2 \end{pmatrix}$$

where the elements of the above matrices are defined in Lemmas B.4.1 and B.4.2. Moreover,

$$v_{1n} = \mathbb{V}\left[\sqrt{nh}e_1'(Z_x'W_xZ_x)^{-1}Z_x'W_x\varepsilon|\mathcal{X}_n\right] = e_1'\Gamma_{1n}^{-1}\Psi_{11n}\Gamma_{1n}^{-1}e_1 \xrightarrow{p} \begin{cases} e_1'\Gamma_1^{-1}\Psi_{11}\Gamma_1^{-1}e_1 & \text{if } x \text{ is interior} \\ e_1'\ddot{\Gamma}_1^{-1}\ddot{\Psi}_{11}\ddot{\Gamma}_1^{-1}e_1 & \text{if } x \text{ is boundary} \end{cases}$$

We are now left with proving asymptotic normality. To do so, we observe that:

$$v_{1,LP,n}^{-1}(T_n^* - \hat{B}_n) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \omega_i(x)\varepsilon_i e_i^* + o_p(1)$$

where $\omega_i(x) = e_1'(\text{plim}_{n\to\infty}S_n)^{-1}Z_{ix}K((x_i-x)/h)/\text{plim}_{n\to\infty}v_{1n}$. Then, asymptotic normality follows from a bootstrap version of Lyapunov's CLT, noting that $\mathbb{E}^*[(nh)^{-1/2}\sum_{i=1}^n \omega_i(x)\varepsilon_i e_i^*] = 0$ and $\mathbb{E}^*[(nh)^{-1}\sum_{i=1}^n \omega_i^2(x)\varepsilon_i^2 e_i^{*2}] \xrightarrow{p} 1$ since, for $\delta > 1$,

$$\frac{1}{(nh)^\delta}\sum_{i=1}^n \mathbb{E}^*\left(w_i(x)\varepsilon_i e_i^*\right)^{2\delta} = \frac{1}{(nh)^\delta}\sum_{i=1}^n \omega_i^{2\delta}(x)\mathbb{E}^*\left(\varepsilon_i e_i^*\right)^{2\delta}$$
$$\leq C_1\frac{1}{(nh)^\delta}\sum_{i=1}^n \omega_i^{2\delta}(x)\varepsilon_i^{2\delta} = O_p\left(\frac{1}{(nh)^{\delta-1}}\right)$$

where the last result is given by Markov's inequality given the fact that $E[\varepsilon_i^4|x_i] \leq \infty$ by Assumption 1.

B.3.3 Proof of Proposition 3.3

Let $w_i(x) = e_1'\Gamma_{1n}^{-1} Z_{ix} K((x_i - x)/h)$, our aim is to derive a CLT for:

$$\begin{pmatrix} \xi_{1n} \\ \xi_{2n} \end{pmatrix} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \begin{pmatrix} w_i(x) \\ \tilde{w}_i(x) \end{pmatrix} \varepsilon_i$$

where $\tilde{w}_i(x) = (nh)^{-1} \sum_{j=1}^{n} w_j(x) w_i(x_j) - w_i(x)$. We will do so by considering the two cases of interior and boundary point separately. Let us consider the case in which x is an interior point first. First of all, by noting that that $\Gamma_{1n} = \Gamma_1 + o_p(1)$ it immediately follows that:

$$\xi_{1n} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \bar{w}_i(x) \varepsilon_i + o_p(1)$$

where

$$\bar{w}_i(x) := e_1'\Gamma_1^{-1} \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} K\left(\frac{x_i - x}{h}\right)$$

We now consider how to apply the same idea to ξ_2, and we let $b_i(x) := (nh)^{-1} \sum_{j=1}^{n} w_j(x) w_i(x_j)$. By the same reasoning than above, we have that:

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} b_i(x) \varepsilon_i = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \bar{b}_i(x) \varepsilon_i + o_p(1)$$

where

$$\bar{b}_i(x) := e_1'\Gamma_1^{-1} \begin{pmatrix} 1 \\ \frac{x_j - x}{h} \end{pmatrix} K\left(\frac{x_j - x}{h}\right) w_i(x_j)$$

Let us expand the term $w_i(x_j)$. Following its definition,

$$w_i(x_j) := e_1'\Gamma_{1n,j}^{-1} Z_{1x_j i} K\left(\frac{x_i - x_j}{h}\right);$$

$$\Gamma_{1n,j} := \begin{pmatrix} \frac{1}{nh}\sum_{l=1}^{n} K\left(\frac{x_l - x_j}{h}\right) & \frac{1}{nh}\sum_{l=1}^{n} K\left(\frac{x_l - x_j}{h}\right)\left(\frac{x_l - x_j}{h}\right) \\ \frac{1}{nh}\sum_{l=1}^{n} K\left(\frac{x_l - x_j}{h}\right)\left(\frac{x_l - x_j}{h}\right) & \frac{1}{nh}\sum_{l=1}^{n} K\left(\frac{x_l - x_j}{h}\right)\left(\frac{x_l - x_j}{h}\right)^2 \end{pmatrix}; \qquad Z_{1x_j i} := \begin{pmatrix} 1 \\ \left(\frac{x_i - x_j}{h}\right) \end{pmatrix}$$

We can note that, differently than before, we cannot take $\Gamma_{1n,j}^{-1}$ out of the summations. What we can do is to remove the randomness of $\Gamma_{1n,j}^{-1}$ coming from the summations over the $l = 1, ..., n$, and to replace it with *deterministic* functions of the random quantity x_j. If we have that, for $r = 0, 1, 2$:

$$\sup_{\tilde{x} \in int\mathbb{S}_x} \left| \frac{1}{nh} \sum_{l=1}^{n} K\left(\frac{x_l - \tilde{x}}{h}\right)\left(\frac{x_l - \tilde{x}}{h}\right)^r - \mathbb{E}\left[\frac{1}{nh} \sum_{l=1}^{n} K\left(\frac{x_l - \tilde{x}}{h}\right)\left(\frac{x_l - \tilde{x}}{h}\right)^r\right] \right| = o_p(1)$$

and

$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{x_l-\tilde{x}}{h}\right)\right]-f(\tilde{x})\right|=o_p(1)$$
$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{x_l-\tilde{x}}{h}\right)\left(\frac{x_l-\tilde{x}}{h}\right)\right]-f(\tilde{x})\mu_1\right|=o_p(1)$$
$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{x_l-\tilde{x}}{h}\right)\left(\frac{x_l-\tilde{x}}{h}\right)^2\right]-f(\tilde{x})\mu_2\right|=o_p(1)$$

Conditions which are satisfied by (B.61) and the discussion above in Hall and Horowitz's supplement, then we can write

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}b_i(x)\varepsilon_i=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\bar{\bar{b}}_i(x)\varepsilon_i+o_p(1)$$

where

$$\bar{\bar{b}}_i(x)=e_1'\Gamma_1^{-1}\begin{pmatrix}1\\ \frac{x_j-x}{h}\end{pmatrix}K\left(\frac{x_j-x}{h}\right)e_1'\Gamma_{1,j}^{-1}\begin{pmatrix}1\\ \frac{x_i-x_j}{h}\end{pmatrix}K\left(\frac{x_i-x_j}{h}\right)$$

and

$$\Gamma_{1,j}:=\begin{pmatrix}f(x_j) & f(x_j)\mu_1\\ f(x_j)\mu_1 & f(x_j)\mu_2\end{pmatrix}$$

If x is a boundary point, by the same steps, we have that:

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}b_i(x)\varepsilon_i=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\bar{\bar{b}}_{i,\texttt{bnd}}(x)\varepsilon_i+o_p(1)$$

where

$$\bar{\bar{b}}_{i,\texttt{bnd}}(x)=e_1'\Gamma_1^{-1}\begin{pmatrix}1\\ \frac{x_j-x}{h}\end{pmatrix}K\left(\frac{x_j-x}{h}\right)e_1'\Gamma_{1,j}^{-1}\begin{pmatrix}1\\ \frac{x_i-x_j}{h}\end{pmatrix}K\left(\frac{x_i-x_j}{h}\right)$$

and

$$\ddot{\Gamma}_{1,j}:=\begin{pmatrix}f(x_j)\ddot{\mu}_{0,x_j/h} & f(x_j)\ddot{\mu}_{1,x_j/h}\\ f(x_j)\ddot{\mu}_{1,x_j/h} & f(x_j)\ddot{\mu}_{2,x_j/h}\end{pmatrix}$$

such that $\ddot{\mu}_{l,x_j/h}:=\int_{-x_j/h}^{1}u^lK(u)du$.

From now on, we focus on the interior point case only, although the same steps apply also to the boundary case. A remark below will specify the parallelism with the following approximation and an analogous approximation for the boundary point case. By noting that $\mu_1 = 0$ if x is an interior point, we can simplify our arguments to get:

$$\begin{aligned}\frac{1}{\sqrt{nh}}\sum_{i=1}^n b_i\varepsilon_i &= \frac{1}{f(x)}\frac{1}{(nh)^{3/2}}\sum_{i=1}^n\sum_{j=1}^n \frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\varepsilon_i + o_p(1)\\ &= \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{K}\left(\frac{x_i-x}{h}\right)\varepsilon_i + o_p(1)\end{aligned}$$

where

$$\tilde{K}\left(\frac{x_i-x}{h}\right) = \frac{1}{nh}\sum_{j=1}^n \frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)$$

We now want to approximate $\tilde{K}$ with a function not involving a convolution sum, replacing the summation over the j index with an integral. This can be seen as an asymptotic approximation of that summation. Specifically, we want to show that:

$$\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{K}\left(\frac{x_i-x}{h}\right)\varepsilon_i = \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i + o_p(1) \qquad \text{(B.3.1)}$$

with

$$\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right) := \frac{1}{h}\int_{a-x}^{b-x} K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du$$

where $a = 1$ and $b = 0$ since we assumed without loss of generality that $\mathbb{S}_x = [0, 1]$.

Remark B.3.1 *In the boundary case, analogous steps as for the following proof of (B.3.1) yield to*

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^n b_i(x)\varepsilon_i = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{b}_{i,\mathsf{bnd}}(x)\varepsilon_i + o_p(1)$$

where

$$\tilde{b}_{i,\mathsf{bnd}}(x) = e_1'\ddot{\Gamma}_1^{-1}\int_0^1 \begin{pmatrix}1\\ \frac{u}{h}\end{pmatrix} K\left(\frac{u}{h}\right) e_1'\ddot{\Gamma}_{1,u}^{-1}\begin{pmatrix}1\\ \frac{x_i-x-u}{h}\end{pmatrix} K\left(\frac{x_i-x-u}{h}\right)du$$

Proof of (B.3.1). We have that

$$\begin{aligned}\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{K}\left(\frac{x_i-x}{h}\right)\varepsilon_i &= \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i +\\ &\quad + \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n\left(\tilde{K}\left(\frac{x_i-x}{h}\right) - \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)\varepsilon_i\\ &=: \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i + R_n\end{aligned}$$

we aim to show that $R_n = o_p(1)$. To do so, we observe that, for $\eta > 0$:

$$\mathbb{P}\left(|R_n| \geq \eta\right) \leq \eta^{-1}\mathbb{E}\left|\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i - x}{h}\right) - \tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\right)\varepsilon_i\right|$$

$$\leq \eta^{-1}\sqrt{\mathbb{E}\left|\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i - x}{h}\right) - \tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\right)\varepsilon_i\right|^2}$$

$$= (\eta f(x))^{-1}\sqrt{\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i - x}{h}\right) - \tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\right)^2\sigma^2(x_i)\right]}$$

where the last equality follows from $\mathbb{E}[\varepsilon_i|x_1, ..., x_n] = 0$. Now we expand the squared difference inside the summation to get that:

$$\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i - x}{h}\right) - \tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\right)^2\sigma^2(x_i)\right] =$$

$$= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j - x}{h}\right)K\left(\frac{x_i - x_j}{h}\right) - \frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i - x - u}{h}\right)du\right)^2\sigma^2(x_i)\right.$$

$$= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j - x}{h}\right)K\left(\frac{x_i - x_j}{h}\right)\right)^2\sigma^2(x_i)\right] +$$

$$+ \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i - x - u}{h}\right)du\right)^2\sigma^2(x_i)\right]$$

$$- 2\cdot\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j - x}{h}\right)K\left(\frac{x_i - x_j}{h}\right)\right)\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i - x - u}{h}\right)du\right)\sigma^2\right.$$

$$=: R_{1n} + R_{2n} - 2R_{12n}$$

We will now prove the result by showing that R_{1n}, R_{2n} and R_{12n} have the same limit.

We start with deriving the limit of R_{1n} as follows.

$$R_{1n} := \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\right)^2\sigma^2(x_i)\right] =$$

$$= \mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)K\left(\frac{x_{j'}-x}{h}\right)K\left(\frac{x_i-x_{j'}}{h}\right)\sigma^2(x_i)\right] =$$

$$= \mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{j'=1,j'\neq i,j'\neq j}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)K\left(\frac{x_{j'}-x}{h}\right)K\left(\frac{x_i-x_{j'}}{h}\right)\sigma\right.$$

$$+ 2K(0)\,\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f(x_i)}\frac{1}{f(x_j)}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right] +$$

$$+ \mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f^2(x_j)}K^2\left(\frac{x_j-x}{h}\right)K^2\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right] +$$

$$+ \mathbb{E}\left[\frac{K^2(0)}{(nh)^3}\sum_{i=1}^{n}\frac{1}{f^2(x_i)}K^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right] =:$$

$$=: R_{1n,1} + R_{1n,2} + R_{1n,3} + R_{1n,4}$$

Note that, referring to the indexes in the second line above, $R_{1n,1}$ refers to the contributions to the triple sum such that $i \neq j \neq j'$, $R_{1n,2}$ to the contributions such that $i = j \neq j'$ and $i = j' \neq j$, $R_{1n,3}$ to the contributions such that $j = j' \neq i$ and $R_{1n,4}$ to the contributions such that $i = j = j'$. We will show that $R_{1n,1}$ is the dominant term of R_{1n} by proving that $R_{1n,2} = o(1)$, $R_{1n,3} = o(1)$ and $R_{1n,4} = o(1)$. To do so, note that:

$$\begin{aligned}
R_{1n,2} &= 2K(0)\,\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right] \\
&= 2K(0)\,\frac{n(n-1)}{(nh)^3}\mathbb{E}\left[\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right] \\
&= 2K(0)\,\frac{n(n-1)}{(nh)^3}\int\int K\left(\frac{x_1-x}{h}\right)K\left(\frac{x_2-x}{h}\right)K\left(\frac{x_1-x_2}{h}\right)\sigma^2(x_1)dx_1dx_2 \\
&= 2K(0)\,\frac{n(n-1)}{n^3h}\int\int K(s+u)\,K(s)\,K(u)\,\sigma^2(x+(s+u)h)duds \\
&= (1+o(1))2\sigma^2(x)K(0)\,\frac{n(n-1)}{n^3h}\int\int K(s+u)\,K(s)\,K(u)\,duds = O(n^{-1}h^{-1}) = o(1)
\end{aligned}$$

where the third equality above follows from the change of variables $x_2 = x + sh; x_1 =$

$x_2+uh = x+(u+s)h$. One can show that $R_{1n,3} = O(n^{-1}h^{-1})$ by analogous steps. Finally,

$$\begin{aligned} R_{1n,4} &= \frac{K^2(0)}{n^2h^3}\mathbb{E}\left[\frac{1}{f^2(x_1)}K^2\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)\right] \\ &= \frac{K^2(0)}{n^2h^3}\int\frac{1}{f(x_1)}K^2\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)dx_1 \\ &= \frac{K^2(0)}{n^2h^2}\int\frac{1}{f(x+uh)}K^2(u)\,\sigma^2(x+uh)du \\ &= (1+o(1))\frac{K^2(0)}{n^2h^2}\frac{\sigma^2(x)}{f(x)}\int K^2(u)\,du = O(n^{-2}h^{-2}) = o(1) \end{aligned}$$

We now focus on the leading term, $R_{1n,1}$, and derive its limit. We have that

$$\begin{aligned} R_{1n,1} := \mathbb{E}\Bigg[&\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{j'=1,j'\neq i,j'\neq j}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)K\left(\frac{x_{j'}-x}{h}\right)K\left(\frac{x_i-x}{h}\right. \\ &\cdot\sigma^2(x_i)\Bigg] = \\ &= (h^{-3}+o(1))\int\int\int K\left(\frac{x_2-x}{h}\right)K\left(\frac{x_1-x_2}{h}\right)K\left(\frac{x_3-x}{h}\right)K\left(\frac{x_1-x_3}{h}\right)\sigma^2(x_1)f(x_1)dx_1dx_2 \\ &= (h^{-3}+o(1))h^3\int\int\int K(r)\,K(s+u)\,K(s+r)\,K(u)\,\cdot \\ &\quad\cdot\sigma^2(x+(u+s+r)h)f(x+(u+s+r)h)dudrds = \\ &= (1+o(h^3))(1+o(1))\sigma^2(x)f(x)\int\int\int K(r)\,K(s+u)\,K(s+r)\,K(u)\,dudrds \\ &= \sigma^2(x)f(x)\int\left[\int K(r)\,K(s+r)\,ds\right]^2 dr + o(1) =: R_1 + o(1) \end{aligned}$$

where the third equality follows from the change of variables:

$$\begin{cases} x_1 = x_3+uh \\ x_3 = x_2+sh \\ x_2 = x+rh \end{cases} \leftrightarrow \begin{cases} x_1 = x+(r+s+u)h \\ x_3 = x+(r+s)h \\ x_2 = x+rh \end{cases}$$

We now move to the derivation of the limit of R_{2n}. This limit can be achieved again by a similar procedure than for the leading term of $R_{1n,1}$. To see this, note that:

$$\begin{aligned}
R_{2n} &:= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)^2\sigma^2(x_i)\right]\\
&= \frac{1}{h^3}\int\left(\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)du\right)^2\sigma^2(x_1)f(x_1)dx_1\\
&= \frac{1}{h^3}\int\int_{a-x}^{b-x}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)K\left(\frac{u'}{h}\right)K\left(\frac{x_1-x-u'}{h}\right)\sigma^2(x_1)f(x_1)dudu'dx_1\\
&= \int\int_{(a-x)/h}^{(b-x)/h}\int_{(a-x)/h-v}^{(b-x)/h-v}K(v)K(r+v)K(r+s)K(s)f(x+(v+r+s)h)\sigma^2(x+(v+r+s)h)ds\\
&= (1+o(1))\sigma^2(x)f(x)\int\int\int K(v)K(r+v)K(r+s)K(s)\,dsdvdr\\
&= \sigma^2(x)f(x)\int\left[\int K(r)K(s+r)\,ds\right]^2dr+o(1) =: R_1+o(1)
\end{aligned}$$

where the fourth equality follows from the change of variables:

$$\begin{cases} x_1 = x+u'+sh\\ u' = u+rh\\ u = vh\end{cases} \leftrightarrow \begin{cases} x_1 = x+(v+r+s)h\\ u' = (v+r)h\\ u = vh\end{cases}$$

We are left with deriving the limit of R_{12n}. We have that:

$$\begin{aligned}
R_{12n} &:= \mathbb{E}\Bigg[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\right)\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)\sigma\\
&= \frac{1}{n^2h^3}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\sigma^2(x_i)\right]+o(1)\\
&= \frac{1}{h^3}\int\int\int_{a-x}^{b-x}K\left(\frac{x_2-x}{h}\right)K\left(\frac{x_1-x_2}{h}\right)K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)f(x_1)\sigma^2(x_1)dudx_1dx_2+o(1)\\
&= \int\int\int_{(a-x)/h-v}^{(b-x)/h-v}K(v)K(r+s)K(r+v)K(s)f(x+(r+s+v)h)\sigma^2(x+(r+s+v)h)drdsdv+o(\\
&= (1+o(1))f(x)\sigma^2(x)\int\int\int K(v)K(r+s)K(r+v)K(s)\,drdsdv =: R_1+o(1)
\end{aligned}$$

where the fourth equality follows from the change of variables:

$$\begin{cases} x_1 = x+u+sh\\ u = x_2-x+rh\\ x_2 = x+vh\end{cases} \leftrightarrow \begin{cases} x_1 = x+(v+r+s)h\\ u = (v+r)h\\ x_2 = x+vh\end{cases}$$

This concludes the proof of (1.1).

Following (1.1), we can write

$$\begin{pmatrix}\xi_{1n}\\ \xi_{2n}\end{pmatrix} = \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix} K\left(\frac{x_i-x}{h}\right) \\ \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right) - K\left(\frac{x_i-x}{h}\right)\end{pmatrix}\varepsilon_i + o_p(1)$$
$$= \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\underbrace{\begin{pmatrix} K\left(\frac{x_i-x}{h}\right) \\ \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\end{pmatrix}\varepsilon_i}_{=:\tilde{s}_i} + o_p(1)$$

We are now interested in deriving a CLT for

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{s}_i := \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix} K\left(\frac{x_i-x}{h}\right) \\ \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\end{pmatrix}\varepsilon_i = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix} K\left(\frac{x_i-x}{h}\right) \\ \frac{1}{h}\int_{a-x}^{b-x} K\left(\frac{u}{h}\right) K\left(\frac{x_i-x-u}{h}\right)du\end{pmatrix}\varepsilon_i$$

We start by proving asymptotic normality. To do so, we note that $\tilde{s}_i$ is a sequence of independent random variables. Hence, we can check if Lyapunov's condition holds together with the Cramer-Wold device. Specifically, for $(\alpha,\beta)\in\mathbb{R}^2$, we want to verify asymptotic normality of $\sum_{i=1}^{n}(nh)^{-1/2}(\alpha K((x_i-x)/h+\beta\tilde{\tilde{K}}((x_i-x)/h)\varepsilon_i =: \sum_{i=1}^{n}\eta_i$. To do so, we first note that:

$$\sum_{i=1}^{n}\mathbb{V}\left[(nh)^{-1/2}\left(\alpha K\left(\frac{x_i-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)\varepsilon_i\right] =$$
$$= (nh)^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(\alpha K\left(\frac{x_i-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)^2\sigma^2(x_i)\right]$$
$$= (nh)^{-1}\alpha^2\sum_{i=1}^{n}\mathbb{E}\left[K^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right] + (nh)^{-1}\beta^2\sum_{i=1}^{n}\mathbb{E}\left[\tilde{\tilde{K}}^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]$$
$$+ (nh)^{-1}2\alpha\beta\mathbb{E}\left[K\left(\frac{x_i-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]$$

where

$$(nh)^{-1}\alpha^2\sum_{i=1}^{n}\mathbb{E}\left[K^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right] = h^{-1}\alpha^2\int K^2\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)f(x_1)dx_1 \quad \text{(B.3.2)}$$
$$= \alpha^2\int K^2(u)\,\sigma^2(x+uh)f(x+uh)du = \quad \text{(B.3.3)}$$
$$= (1+o(1))\sigma^2(x)f(x)\alpha^2\int K^2(u)\,du = O(1) \quad \text{(B.3.4)}$$

and

$$(nh)^{-1}2\alpha\beta\mathbb{E}\left[K\left(\frac{x_i-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]=h^{-1}2\alpha\beta\int K\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)f(x_1)dx_1 \tag{B.3.5}$$

$$=h^{-2}2\alpha\beta\int\int_{a-x}^{b-x}K\left(\frac{x_1-x}{h}\right)K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)f(x_1)\sigma^2(x_1)dudx_1 \tag{B.3.6}$$

$$=2\alpha\beta\int\int_{(a-x)/h}^{(b-x)/h}K(r)K(v)K(r-v)f(x+rh)\sigma^2(x+rh)drdv \tag{B.3.7}$$

$$=2\alpha\beta f(x)\sigma^2(x)\int\int K(r)K(v)K(r-v)drdv+o(1)=O(1) \tag{B.3.8}$$

Moreover, note that

$$\beta^2(nh)^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\tilde{\tilde{K}}^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]=\beta^2R_{2n}=\beta^2R_1+o(1)=O(1) \tag{B.3.9}$$

We can now show that Lyapunov's condition holds. Specifically, we have to prove that, for some $\delta>0$,

$$\frac{\sum_{i=1}^n\mathbb{E}|\eta_i|^{2+\delta}}{\left(\sum_{i=1}^n\mathbb{V}(\eta_i)\right)^{2+\delta}}=o(1)$$

Since we already proved that $\left(\sum_{i=1}^n\mathbb{V}(\omega_i)\right)^{2+\delta}=O(1)$ for all $\delta>0$, it suffices to show that $\sum_{i=1}^n\mathbb{E}|\eta_i|^{2+\delta}=o(1)$; for simplicity, we take $\delta=2$ and note that:

$$\begin{aligned}\sum_{i=1}^n\mathbb{E}|\eta_i|^4&=\frac{1}{n^2h^2}\sum_{i=1}^n\mathbb{E}\left[\left(\alpha K\left(\frac{x_i-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)^4\varepsilon_i^4\right]\leq\\
&\leq\frac{c_1}{nh^2}\mathbb{E}\left[\left(\alpha K\left(\frac{x_1-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_1-x}{h}\right)\right)^4\right]=\\
&=\frac{c_1\alpha^4}{nh^2}\mathbb{E}\left[K^4\left(\frac{x_1-x}{h}\right)\right]+\beta\frac{c_1\alpha^4}{nh^2}\mathbb{E}\left[\tilde{\tilde{K}}^4\left(\frac{x_1-x}{h}\right)\right]+\\
&\quad+\frac{4c_1\alpha^3\beta}{nh^2}\mathbb{E}\left[K^3\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_1-x}{h}\right)\right]+\frac{4c_1\alpha^2\beta^2}{nh^2}\mathbb{E}\left[K^2\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^2\left(\frac{x_1-x}{h}\right)\right]\\
&\quad+\frac{4c_1\alpha\beta^3}{nh^2}\mathbb{E}\left[K\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^3\left(\frac{x_1-x}{h}\right)\right]\end{aligned}$$

We are going to conclude this proof by showing that each term above is $o(1)$. First of all, we can see that

$$\begin{aligned}\mathbb{E}\left[K^4\left(\frac{x_1-x}{h}\right)\right]&=\int K^4\left(\frac{x_1-x}{h}\right)f(x_1)dx_1\\
&=(1+o(1))hf(x)\int K^4(u)\,du\end{aligned}$$

which proves that the first term is $O(n^{-1}h^{-1})$. To handle the remaining four terms, we will show that for $\gamma = 0, 1, 2, 3$:

$$\mathbb{E}\left[K^{\gamma}\left(\frac{x_1 - x}{h}\right)\tilde{\tilde{K}}^{4-\gamma}\left(\frac{x_1 - x}{h}\right)\right] = O(h)$$

To see why, we write:

$$\mathbb{E}\left[K^{\gamma}\left(\frac{x_1 - x}{h}\right)\tilde{\tilde{K}}^{4-\gamma}\left(\frac{x_1 - x}{h}\right)\right] := \int K^{\gamma}\left(\frac{x_1 - x}{h}\right)\tilde{\tilde{K}}^{4-\gamma}\left(\frac{x_1 - x}{h}\right) f(x_1)dx_1$$

$$= \int K^{\gamma}\left(\frac{x_1 - x}{h}\right)\left(\frac{1}{h}\int_{a-x}^{b-x} K\left(\frac{u}{h}\right) K\left(\frac{x_1 - x - u}{h}\right) du\right)^{4-\gamma} f(x_1)dx_1$$

$$= h^{\gamma-4}\int K^{\gamma}\left(\frac{x_1 - x}{h}\right)\left(\int_{a-x}^{b-x} K\left(\frac{u}{h}\right) K\left(\frac{x_1 - x - u}{h}\right) du\right)^{4-\gamma} f(x_1)dx_1$$

$$= h^{\gamma-4}\int \underbrace{\int_{a-x}^{b-x} ... \int_{a-x}^{b-x}}_{4-\gamma} K^{\gamma}\left(\frac{x_1 - x}{h}\right)\prod_{j=1}^{4-\gamma}\left\{K\left(\frac{u_j}{h}\right) K\left(\frac{x_1 - x - u_j}{h}\right)\right\} du_1...du_{4-\gamma} f(x_1)dx_1$$

$$= h\int \underbrace{\int_{(a-x)/h}^{(b-x)/h} ... \int_{(a-x)/h}^{(b-x)/h}}_{4-\gamma} K^{\gamma}(v)\prod_{j=1}^{4-\gamma}\left\{K(s_j) K(v - s_j)\right\} f(x + vh)ds_1...ds_{4-\gamma}dv$$

$$= (1 + o(1))hf(x)\int \underbrace{\int ... \int}_{4-\gamma} K^{\gamma}(v)\prod_{j=1}^{4-\gamma}\left\{K(s_j) K(v - s_j)\right\} ds_1...ds_{4-\gamma}dv = O(h)$$

where the fourth equality follows from the change of variables

$$\begin{cases} x_1 = x + vh \\ u_1 = s_1 h \\ ... \\ u_{4-\gamma} = s_{4-\gamma}h \end{cases}$$

with the corresponding Jacobian matrix:

$$J := \begin{bmatrix} \partial x_1/\partial v & \partial x_1/\partial s_1 & ... & \partial x_1/\partial s_{4-\gamma} \\ \partial u_1/\partial v & \partial u_1/\partial s_1 & ... & \partial u_1/\partial s_{4-\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ \partial u_{4-\gamma}/\partial v & \partial u_{4-\gamma}/\partial s_1 & ... & \partial u_{4-\gamma}/\partial s_{4-\gamma} \end{bmatrix} = \text{diag}(h, h, ..., h)$$

so that $\det(J) = h^{5-\gamma}$. This concludes the proof of asymptotic normality of $\sum_{i=1}^{n} \eta_i$. We are now only left with deriving the asymptotic variance of $(nh)^{-1/2}\sum_{i=1}^{n} \tilde{s}_i$. From (1.4), we have that

$$\mathbb{E}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K^2\left(\frac{x_i-x}{h}\right)\varepsilon_i\right]=\sigma^2(x)f(x)\int K^2(u)du+o(1)=:\omega_1+o(1)$$

From (1.8), we have that

$$\mathbb{E}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K\left(\frac{x_i-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i\right]=f(x)\sigma^2(x)\int\int K(s)K(v)K(s+v)\,dsdv+o(1)$$
$$=:\omega_{12}+o(1)$$

Finally, from (1.9), we obtain

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{\tilde{K}}^2\left(\frac{x_i-x}{h}\right)\varepsilon_i\right]&=R_1+o(1)\\&=\sigma^2(x)f(x)\int\left[\int K(r)K(s+r)\,ds\right]^2dr+o(1)\\&=:\omega_2+o(1)\end{aligned}$$

Hence, defining

$$\Omega=\begin{bmatrix}\omega_1 & \omega_{12}\\ \omega_{12} & \omega_2\end{bmatrix}$$

we have that

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{s}_i\xrightarrow{d}N(0,\Omega)\tag{B.3.10}$$

and

$$\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}\xrightarrow{d}N(0,V_{LP})\tag{B.3.11}$$

with

$$V_{LP}:=\frac{1}{f^2(x)}\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\Omega\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}\tag{B.3.12}$$

Finally, the proof is completed by noting that

$$\mathbb{V}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{s}_i|\mathcal{X}_n\right]\xrightarrow{p}V_{LP}$$

and from the following remark.

Remark B.3.2 *In the boundary case, by exploiting the approximation in Remark B.3.1, we obtain that*

$$\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}\xrightarrow{d}N(0,\ddot{V}_{LP})\tag{B.3.13}$$

and

$$V_n\xrightarrow{p}\ddot{V}_{LP}\tag{B.3.14}$$

where

$$\ddot{V}_{LP} := \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \ddot{\omega}_1 & \ddot{\omega}_{12} \\ \ddot{\omega}_{12} & \ddot{\omega}_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

with

$$\ddot{\omega}_1 = \sigma^2(0) e_1' \ddot{\Gamma}_1^{-1} \ddot{\Psi}_{11} \ddot{\Gamma}_1^{-1} e_1$$

$$\ddot{\omega}_{12} := \sigma^2(0) e_1' \ddot{\Gamma}_1^{-1} \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \ddot{\Gamma}_1^{-1} e_1$$

$$\ddot{\omega}_2^2 := \sigma^2(0) e_1' \ddot{\Gamma}_1^{-1} \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \ddot{\Gamma}_1^{-1} e_1$$

$$\gamma_{ll'} := f(0) \int_0^1 \int_{-u}^1 (s)^{l-1} (u+s)^{l'-1} K(s) K(u+s) e_1' (\ddot{\Gamma}^{\dagger}_{1,s})^{-1} \begin{pmatrix} 1 \\ u \end{pmatrix} K(u) \, ds du$$

$$\lambda_{ll'} := f(0) \int_0^1 \int_{-r}^1 \int_{-s-r}^1 (r)^{l-1} (s+r)^{l'-1} K(r) K(s+r) e_1' (\ddot{\Gamma}^{\dagger}_{1,r})^{-1} \begin{pmatrix} 1 \\ u+s \end{pmatrix}$$

$$\cdot \, e_1' (\ddot{\Gamma}^{\dagger}_{1,(s+r)})^{-1} \begin{pmatrix} 1 \\ u \end{pmatrix} K(u+s) K(u) \, du ds dr$$

and

$$\ddot{\Gamma}^{\dagger}_{1,u} = \begin{pmatrix} \ddot{\mu}_{0,u} & \ddot{\mu}_{1,u} \\ \ddot{\mu}_{1,u} & \ddot{\mu}_{2,u} \end{pmatrix}$$

B.3.4 Proof of Proposition 3.4

We start by considering the expansion on $\hat{B}_n$, which holds for both interior and boundary points. Let us write

$$\begin{aligned} \hat{B}_n &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i(x) \left(\hat{m}(x_i) - \hat{m}(x) \right) \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i(x) \left(\frac{1}{nh} \sum_{j=1}^n w_j(x_i) g(x_j) - g(x_i) \right) \\ &\quad + \frac{1}{nh} \sum_{i=1}^n w_i(x) \left(\frac{1}{nh} \sum_{j=1}^n w_j(x_i) \varepsilon_j - \varepsilon_i \right) \\ &=: \hat{B}_{1,n} + \xi_{2,n} \end{aligned}$$

Note that, by the mean value theorem:

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)g(x_j) =$$

$$= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)\Bigg[g(x_i) + hg'(x_i)\left(\frac{x_j - x_i}{h}\right) + \frac{h^2 g''(x_i)}{2}\left(\frac{x_j - x_i}{h}\right)^2 + \frac{h^3 g^{(3)}(\tilde{x}_{ij})}{6}\left(\frac{x_j - x_i}{h}\right)^3\Bigg]$$

$$= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)\Bigg[\left(1\ \left(\tfrac{x_j - x_i}{h}\right)\right)\begin{pmatrix} g(x_i) \\ hm'(x_i)\end{pmatrix} + \frac{h^2 g''(x_i)}{2}\left(\frac{x_j - x_i}{h}\right)^2 + \frac{h^3 g^{(3)}(\tilde{x}_{ij})}{6}\left(\frac{x_j - x_i}{h}\right)^3\Bigg]$$

$$= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)g(x_i) + \frac{h^2}{2}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)g''(x_i)\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)\left(\frac{x_j - x_i}{h}\right)^2 + O_p(h)$$

with $\tilde{x}_{ij}$ some value between x_i and x_j. The above implies that:

$$B_{2n} = \frac{h^2}{2}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)g''(x_i)\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)\left(\frac{x_j - x_i}{h}\right)^2 + O_p(h)$$

and

$$B_{2n} - B_n = \frac{h^2}{2}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} w_i(x)g''(x_i)\left[\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)\left(\frac{x_j - x_i}{h}\right)^2 - \left(\frac{x_i - x}{h}\right)^2\right] + O_p(h)$$

$$= \kappa^{1/2}\frac{g''(x)}{2}\frac{1}{nh}\sum_{i=1}^{n} w_i(x)\left[\frac{1}{nh}\sum_{j=1}^{n} w_j(x_i)\left(\frac{x_j - x_i}{h}\right)^2 - \left(\frac{x_i - x}{h}\right)^2\right] + O_p(h) + O_p(h) =$$

$$= \kappa^{1/2}\frac{g''(x)}{2}\left[C_{2n} - C_n\right] + O_p(h)$$

For part (i), and by the same reasoning as in the Proof of Proposition 3.3, we have that

$$C_{2n} = \frac{1}{f(x)}\frac{1}{nh}\sum_{i=1}^{n}\tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^2 + o_p(1)$$

so that by mean squared error convergence we can prove that

$$C_{2n} = \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^2\right] + o_p(1)$$

where

$$\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^2\right] = \mu_2 \qquad \text{(B.3.15)}$$

and

$$B_{2n} = \kappa^{1/2}\frac{g''(x)\mu_2}{2} + o_p(1)$$

By combining this result with the probability limit of B_n, it follows that $B_{2n} - B_n = o_p(1)$.

For part (ii), from the results in Proposition 3.3:

$$C_{2n} = \frac{1}{nh}\sum_{i=1}^{n} \tilde{b}_{i,\texttt{bnd}}(x)\left(\frac{x_i - x}{h}\right)^2 + o_p(1)$$

so that by mean squared error convergence we can prove that

$$C_{2n} = \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} \tilde{b}_{i,\texttt{bnd}}(x)\left(\frac{x_i - x}{h}\right)^2\right] + o_p(1)$$

where

$$\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} \tilde{b}_{i,\texttt{bnd}}(x)\left(\frac{x_i - x}{h}\right)^2\right] = e_1'\ddot{\Gamma}_1^{-1}\int_0^1 K(s)\begin{pmatrix}1\\ s\end{pmatrix} e_1'\ddot{\Gamma}_{1,sh}^{-1}\left[\int_{-s}^{1} K(u)\begin{pmatrix}u^2\\ u^3\end{pmatrix} du\right] ds =: C_2$$

and

$$B_{2n} = \kappa^{1/2}\frac{g''(x)C_2}{2} + o_p(1)$$

By combining this result with the probability limit of B_n, it follows that $B_{2n} - B_n = A + o_p(1)$.

B.3.5 Proof of Proposition 3.5

We note that

$$\begin{aligned}
\mathbb{P}^*\left(T_n^* \leq T_n\right) &= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,LP,n}} \leq \frac{T_n - \hat{B}_n}{v_{1,LP,n}}\right) \\
&= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,LP,n}} \leq \frac{v_{d,LP,n}}{v_{1,LP,n}}\frac{\xi_{1n} - \xi_{2n}}{v_{d,LP,n}} + \frac{B_{2n} - B_n}{v_{1,LP,n}}\right) \\
&\xrightarrow{d} \Phi\left(\text{plim}\{(B_{2n} - B_n)/v_{1,LP,n}\} + \text{plim}\{v_{d,LP,n}/v_{1,n}\}\Phi^{-1}\left(U_{[0,1]}\right)\right)
\end{aligned}$$

where the last convergence result is given by Propositions 2.3.2 and 2.3.3. The result then applies to the case of interior and boundary points by considering the different specifications of the probability limits included in Propositions 2.3.2, 2.3.3 and 2.3.4.

B.3.6 Proof of Theorem 3.1

Note that Proposition 2.3.1 ensures that Assumption 1 in Cavaliere et al. (2024) is satisfied. Moreover, Propositions 2.3.2 and 2.3.3 ensure that Assumption 2 in Cavaliere et al. (2024) is satisfied. Then, the conditions of Corollary 3.2 in Cavaliere et al. (2024) are satisfied because H is continuous in m_{LP} and $\hat{m}_{LP,n} = m_{LP} + o_p(1)$. Hence, we have that:

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LP}\right) = \mathbb{P}\left(\alpha/2 \leq \hat{H}_{LP,n}\left(\hat{p}_n\right) \leq 1 - \alpha/2\right) \to 1 - \alpha$$

B.3.7 Proof of Proposition 3.6

For part (i), just note that

$$(v_{1n}Q_n)^{-1}(T^*_{n,mod} - \hat{B}_{n,mod}) = v_{1n}^{-1}(T^*_n - \hat{B}_n) = v_{1n}^{-1}\xi^*_{1n}$$

so that the result follows from Proposition 3.2 directly. For part (ii), just note that

$$\hat{B}_{mod,n} - B_n = Q_n B_{LP,n} - B_{AT,n} + Q_n\xi_{2,n} + o_p(1) =: \xi_{2,mod,n} + o_p(1)$$

where the first equality is given by (2.3.4) and (2.3.5) and by Proposition 2.3.4. Finally, note that part (iii) follows directly from Proposition 2.3.3 and from the fact that $\xi_{mod,n} := \text{diag}(1, Q_n)(\xi_{1n}, \xi_{2n})'$, which ensures that $V_{LP,mod,n} = \text{diag}(1, Q_n)V_{LP,n}\text{diag}(1, Q_n)$.

B.3.8 Proof of Proposition 3.7

By the usual expansion

$$\begin{aligned}
\mathbb{P}^*\left(T^*_{mod,n} \leq T_n\right) &= \mathbb{P}^*\left(\frac{T^*_{mod,n} - \hat{B}_{mod,n}}{Q_n v_{1,LP,n}} \leq \frac{T_n - \hat{B}_{mod,n}}{Q_n v_{1,LP,n}}\right) \\
&= \mathbb{P}^*\left(\frac{T^*_{mod,n} - \hat{B}_{mod,n}}{Q_n v_{1n}} \leq \frac{v_{d,LP,mod,n}}{Q_n v_{1,LP,n}}\frac{\xi_{1n} - \xi_{2,mod,n}}{v_{d,LP,mod,n}} + \frac{Q_n B_{2n} - B_n}{Q_n v_{1,LP,n}}\right) \\
&\xrightarrow{d} \Phi\left(\text{plim}\{v_{d,LP,mod,n}/Q_n v_{1,n}\}\Phi^{-1}\left(U_{[0,1]}\right)\right)
\end{aligned}$$

where the last convergence result is given by Propositions 2.3.6. The result then applies to the case of interior and boundary points by considering the different specifications of the probability limits included in Propositions 2.3.2, 2.3.3 and 2.3.4.

B.3.9 Proof of Theorem 3.2

For interior points $Q_n = 1 + o_p(1)$, so that the result follows directly from Theorem 3.1.

For boundary points, note that Proposition 2.3.1 ensures that Assumption A in Appendix A is satisfied. Moreover, Propositions 2.3.2, 2.3.3 and 2.3.4 ensure that Assumption B and C in Appendix A are satisfied. Then, the conditions of Theorem B.1.2 in Appendix A are satisfied because H is continuous in $\ddot{m}_{LP,mod}$ and $\hat{m}_{LP,mod,n} = \ddot{m}_{LP,mod} + o_p(1)$. Hence, we have that:

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LP,mod}\right) = \mathbb{P}\left(\alpha/2 \leq \hat{H}_{LP,mod,n}\left(\hat{p}_{mod,n}\right) \leq 1 - \alpha/2\right) \to 1 - \alpha$$

B.3.10 Proof of Proposition 3.8

The proof of Proposition 3.8 is analogous to that of proposition 3.2 and it thus omitted for the seek of brevity.

B.3.11 Proof of Proposition 3.9

Let $w_i(x) = e_1'\Gamma_{1n}^{-1}Z_{ix}K((x_i - x)/h)$, our aim is to derive a CLT for:

$$\begin{pmatrix} \xi_{1n} \\ \tilde{\xi}_{2n} \end{pmatrix} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \begin{pmatrix} w_i(x) \\ C_n \tilde{l}_i(x) \end{pmatrix} \varepsilon_i$$

First of all, we note that

$$\begin{pmatrix} \xi_{1n} \\ \tilde{\xi}_{2n} \end{pmatrix} = \mathrm{diag}(1, C) \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \begin{pmatrix} \bar{w}_i(x) \\ \bar{l}_i(x) \end{pmatrix} \varepsilon_i + o_p(1)$$

where

$$\begin{aligned} \bar{w}_i(x) &:= e_1'\Gamma_1^{-1} \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} K\left(\frac{x_i - x}{h}\right) \\ \bar{l}_i(x) &:= e_3'\Gamma_2^{-1} \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \\ \left(\frac{x_i - x}{h}\right)^2 \end{pmatrix} K\left(\frac{x_i - x}{h}\right) \end{aligned}$$

Then the result follows immediately as a bivariate estension (i.e., by exploiting the Cramer-Wold device) of the central limit theorem proposed in Lemma A2 in Calonico et al. (2014), where

$$\begin{aligned} v_{1,FL}^2 &:= v_{1,LP}^2 \\ v_{12,FL}^2 &:= \sigma^2(x) e_1'\Gamma_1^{-1}\Psi_{12}\Gamma_2^{-1}e_3 \\ v_{22,FL}^2 &:= \sigma^2(x) e_3'\Gamma_2^{-1}\Psi_{22}\Gamma_2^{-1}e_3 \\ \ddot{v}_{1,FL}^2 &:= \ddot{v}_{1,LP}^2 \\ \ddot{v}_{12,FL}^2 &:= \sigma^2(0) e_1'\ddot{\Gamma}_1^{-1}\ddot{\Psi}_{12}\ddot{\Gamma}_2^{-1}e_3 \\ \ddot{v}_{22,FL}^2 &:= \sigma^2(0) e_3'\ddot{\Gamma}_2^{-1}\ddot{\Psi}_{22}\ddot{\Gamma}_2^{-1}e_3 \end{aligned}$$

such that:

$$\Psi_{12} := \begin{pmatrix} \psi_0 & \psi_1 & \psi_2 \\ \psi_1 & \psi_2 & \psi_3 \end{pmatrix} \qquad \ddot{\Psi}_{12} := \begin{pmatrix} \ddot{\psi}_0 & \ddot{\psi}_1 & \ddot{\psi}_2 \\ \ddot{\psi}_1 & \ddot{\psi}_2 & \ddot{\psi}_3 \end{pmatrix}$$

$$\Gamma_2 := \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_2 & \gamma_3 & \gamma_4 \end{pmatrix} \qquad \ddot{\Gamma}_2 := \begin{pmatrix} \ddot{\gamma}_0 & \ddot{\gamma}_1 & \ddot{\gamma}_2 \\ \ddot{\gamma}_1 & \ddot{\gamma}_2 & \ddot{\gamma}_3 \\ \ddot{\gamma}_2 & \ddot{\gamma}_3 & \ddot{\gamma}_4 \end{pmatrix}$$

where the elements of the above matrices are defined in Lemmas B.4.1 and B.4.2.

B.3.12 Proof of Proposition 3.10

We note that

$$\begin{aligned}\mathbb{P}^*\left(T_n^* \leq T_n\right) &= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,FL,n}} \leq \frac{T_n - \hat{B}_n}{v_{1,FL,n}}\right) \\ &= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,FL,n}} \leq \frac{v_{d,FL,n}}{v_{1,FL,n}}\frac{\xi_{1n} - \xi_{2n}}{v_{d,FL,n}}\right) \\ &\xrightarrow{d} \Phi\left(\text{plim}\{v_{d,FL,n}/v_{1,FL,n}\}\Phi^{-1}\left(U_{[0,1]}\right)\right)\end{aligned}$$

where the last convergence result is given by Propositions B.2.1 and B.2.2. The result then applies to the case of interior and boundary points by considering the different specifications of the probability limits included in Propositions B.2.1 and B.2.2.

B.3.13 Proof of Theorem 3.3

Note that Proposition 2.3.1 ensures that Assumption 1 in Cavaliere et al. (2024) is satisfied. Moreover, Propositions B.2.1 and B.2.2 ensure that Assumption 2 in Cavaliere et al. (2024) is satisfied. Then, the conditions of Corollary 3.2 in Cavaliere et al. (2024) are satisfied because H is continuous in m_{FL} and $\hat{m}_{FL,n} = m_{FL} + o_p(1)$. Hence, we have that:

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{FL}\right) = \mathbb{P}\left(\alpha/2 \leq \hat{H}_{FL,n}\left(\hat{p}_n\right) \leq 1 - \alpha/2\right) \to 1 - \alpha$$

B.3.14 Proof of Proposition 3.11

The results follows immediately from the proofs of Propositions 3.2, 3.3, 3.8 and 3.9, where we define the elements in the covariance matrices in the according central limit theorems, noting that the quantities:

$$\begin{aligned}\mathcal{K}_{v_d,LP} &:= \frac{f(x)}{\sigma^2(x)}v_{d,LP}^2 \\ \mathcal{K}_{v_d,FL} &:= \frac{f(x)}{\sigma^2(x)}v_{d,FL}^2 \\ \ddot{\mathcal{K}}_{v_d,LP} &:= \frac{f(0)}{\sigma^2(0)}\ddot{v}_{d,LP}^2 \\ \ddot{\mathcal{K}}_{v_d,FL} &:= \frac{f(0)}{\sigma^2(0)}\ddot{v}_{d,FL}^2\end{aligned}$$

are measurable functions of the kernel K only.

B.4 Auxiliary results

Lemma B.4.1 *Let Assumptions 1-3 hold, then: (i) if x is an interior point*

$$\gamma_{j,n} = f(x)\mu_j + O_p\left(\frac{1}{\sqrt{nh}}\right) = \gamma_j + O_p\left(\frac{1}{\sqrt{nh}}\right), \quad j = 0, 1, 2, ... \tag{B.4.1}$$

(ii) whereas if x is a boundary point

$$\gamma_{j,n} = f(0)\ddot{\mu}_j + O_p\left(\frac{1}{\sqrt{nh}}\right) = \ddot{\gamma}_j + O_p\left(\frac{1}{\sqrt{nh}}\right) \quad j = 0, 1, 2, ... \tag{B.4.2}$$

Proof of Lemma B.4.1. For part (i), note that

$$\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^j\right] = \frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\left[K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^j\right] =$$
$$= \frac{1}{h}\mathbb{E}\left[K\left(\frac{x_1 - x}{h}\right)\left(\frac{x_1 - x}{h}\right)^j\right] = \frac{1}{h}\int_0^1 K\left(\frac{x_1 - x}{h}\right)\left(\frac{x_1 - x}{h}\right)^j f(x_1)dx_1$$
$$= \int_{-x/h}^{(1-x)/h} K(u)\, u^j f(x + uh)du \to f(x)\int_{-\infty}^{+\infty} K(u)\, u^j du = f(x)\int_{-1}^{1} K(u)\, u^j du =: f(s)\mu_j$$

and

$$\frac{1}{(nh)^2}\sum_{i=1}^{n}\mathbb{E}\left[K^2\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^{2j}\right] +$$
$$+ \frac{1}{(nh)^2}\sum_{i\neq i'}\mathbb{E}\left[K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^j\right]\mathbb{E}\left[K\left(\frac{x_{i'} - x}{h}\right)\left(\frac{x_{i'} - x}{h}\right)^j\right] =$$
$$= O\left(\frac{1}{nh}\right) + \frac{1}{h^2}\left[\mathbb{E}\left[K\left(\frac{x_1 - x}{h}\right)\left(\frac{x_1 - x}{h}\right)^j\right]\right]^2$$
$$= O\left(\frac{1}{nh}\right) + f^2(x)\mu_j^2$$

For part (ii), just note that, if x is a boundary point in the sense of Remark 2.3.1,

$$\int_{-x/h}^{(1-x)/h} K(u)\, u^j f(x + uh)du = \int_0^{(1-x)/h} K(u)\, u^j f(uh)\sigma^2(uh)du$$
$$\to \sigma^2(0)f(0)\int_0^{+\infty} K(u)\, u^j du = \sigma^2(0)f(0)\int_0^1 K(u)\, u^j du =: \sigma^2(0)f(0)\ddot{\mu}_j$$

which concludes the proof.

Lemma B.4.2 *Let Assumptions 1-3 hold, then: (i) if x is an interior point*

$$\psi_{j,n} = \sigma^2(x)f(x)\nu_j + O_p\left(\frac{1}{\sqrt{nh}}\right) = \psi_j + O_p\left(\frac{1}{\sqrt{nh}}\right), \qquad j = 0, 1, 2, ... \tag{B.4.3}$$

(ii) whereas if x is a boundary point

$$\psi_{j,n} = \sigma^2(0)f(0)\ddot{\nu}_j + O_p\left(\frac{1}{\sqrt{nh}}\right) = \ddot{\psi}_j + O_p\left(\frac{1}{\sqrt{nh}}\right), \qquad j = 0, 1, 2, ... \tag{B.4.4}$$

Proof of Lemma B.4.2. For part (i), note that

$$\begin{aligned}
\mathbb{E}\left[\psi_{j,n}\right] &= \frac{1}{h}\mathbb{E}\left[K^2\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j\sigma^2(x_1)\right] \\
&= \frac{1}{h}\int_0^1 K^2\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j f(x_1)\sigma^2(x_1)dx_1 \\
&= \int_{-x/h}^{(1-x)/h} K^2(u)\,u^j f(x+uh)du \to \sigma^2(x)f(x)\int_{-1}^1 K^2(u)\,u^j du =: \sigma^2(x)f(x)\nu_j
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\psi_{j,n}^2(x)\right] &= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j\sigma^2(x_i)\right]^2 \\
&= \frac{1}{(nh)^2}\sum_{i=1}^n \mathbb{E}\left[K^4\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^{2j}\sigma^4(x_i)\right] + \\
&\quad + \frac{1}{(nh)^2}\sum_{i\neq i'}\mathbb{E}\left[K^2\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j\sigma^2(x_i)\right]\mathbb{E}\left[K^2\left(\frac{x_{i'}-x}{h}\right)\left(\frac{x_{i'}-x}{h}\right)^j\sigma^2(x_{i'})\right] \\
&= O_p\left(\frac{1}{nh}\right) + f^2(x)\nu_j^2
\end{aligned}$$

For part (ii), note that

$$\begin{aligned}
\int_{-x/h}^{(1-x)/h} K^2(u)\,u^j f(x+uh)\sigma^2(x+uh)du &= \int_0^{(1-x)/h} K^2(u)\,u^j f(uh)\sigma^2(uh)du \\
&\to \sigma^2(0)f(0)\int_0^1 K^2(u)\,u^j du =: f(0)\ddot{\nu}_j
\end{aligned}$$

Appendix C

Appendix to Chapter 3

C.1 Mathematical Appendix

C.1.1 Proof of Theorem 3.4.1

Introduce $\tilde{x}_t := (1, x_{n,t-1})'$. Let $\mu_n := n^{1/2}(\hat{\theta} - \theta_0)$, $M_n := n^{-1}\sum_{t=1}^n \tilde{x}_t\tilde{x}_t'$ and $N_n^* := n^{-1/2}\sum_{t=1}^n \varepsilon_t^*\tilde{x}_t$. Moreover, let the normalized bootstrap estimator be denoted by $\mu_n^* := n^{1/2}(\hat{\theta}^* - \hat{\theta})$; similarly, $\tilde{\mu}_n^* := n^{1/2}(\tilde{\theta}^* - \hat{\theta})$, where $\tilde{\theta}^*$ is the unrestricted (OLS) bootstrap estimator. On the event $\{\det(M_n) > 0\}$ with $P(\det(M_n) > 0) \to 1$, the estimator $\tilde{\theta}^*$ is well-defined and unique. As we are interested in distributional convergence results, without loss of generality we proceed as if $P(\det(M_n) > 0) = 1$.

By arguments similar to the proof of Theorem 4.1 in Cavaliere and Georgiev (2020), it can be concluded that $(\mu_n, M_n, N_n^*) \xrightarrow{w^*}_w (\ell(\theta_0), M, M^{1/2}\xi^*)|(M, \ell(\theta_0))$ in $\mathbb{R}^{2\times 4}$, where M is of full rank with probability one, $\xi^*|(M, \ell(\theta_0)) \sim N(0, \sigma_e^2 I_2)$ and σ_e^2 denotes the variance of ε_t corrected for $\Delta x_{n,t}$. To derive the result (3.4.1), we analyze the properties of μ_n^* on a special probability space where (μ_n, M_n, N_n^*) given the data converge weakly a.s. rather than weakly in distribution. Specifically, by Lemma A.2(a) in Cavaliere and Georgiev (2020) we can consider a probability space (where $\ell(\theta_0), M$ and, for every $n \in N$, also the original data and the bootstrap sample can be redefined, maintaining their distribution), such that

$$\mu_n \overset{a.s.}{\to} \ell(\theta_0),\ M_n \overset{a.s.}{\to} M,\ N_n^* \xrightarrow{w^*}_{a.s.} M^{1/2}\xi^*|(M, \ell(\theta_0)) = M^{1/2}\xi^*|M, \tag{C.1.1}$$

the last equality being an a.s. equality of conditional distributions.

Let $q_n^*(\theta) := n^{-1}\sum_{t=1}^n (y_t^* - \theta'\tilde{x}_t)^2$ with $\tilde{\theta}^* := \arg\min_{\theta\in\mathbb{R}^2} q_n^*(\theta)$ being well-defined and unique for outcomes in the event $\{\det(M_n) > 0\}$. On the special probability space, the asymptotic distribution of $\tilde{\mu}_n^* = n^{1/2}(\tilde{\theta}^* - \hat{\theta}) = M_n^{-1}N_n^*$ follows from (C.1.1) and a CMT (Theorem 10 of Sweeting, 1989):

$$\tilde{\mu}_n^* \xrightarrow{w^*}_{a.s.} \tilde{\ell}^*|(M, \ell(\theta_0)) = \tilde{\ell}^*|M,\ \tilde{\ell}^* := \sigma_e^2 M^{-1/2}\xi^*. \tag{C.1.2}$$

Let us turn now to the bootstrap estimator $\hat{\theta}^*$. If $g(\theta_0) > g^*(\theta_0)$, then the consistency facts $\hat{\theta} \overset{a.s.}{\to} \theta_0$ (from (C.1.1)) and $\tilde{\theta}^* \overset{w^*}{\to}_{a.s.} \theta_0$ (from (C.1.2)), jointly with the continuity of g, g^* at θ_0, imply that $P^*(g(\tilde{\theta}^*) \geq g^*(\hat{\theta})) \overset{a.s.}{\to} 1$. Hence, $\tilde{\theta}^*$ uniquely minimizes q_n^* on Θ^* with P^*-probability approaching one a.s. This establishes the existence of $\hat{\theta}^*$ with P^*-probability approaching one a.s., as well as the facts $P^*(\hat{\theta}^* = \tilde{\theta}^*) \overset{a.s.}{\to} 1$ and $P^*(\mu_n^* = \tilde{\mu}_n^*) \overset{a.s.}{\to} 1$. Using also (C.1.2), it follows that $\mu_n^* \overset{w}{\to}_{a.s.} \tilde{\ell}^*|M$ on the special probability space, and since $\mu_n \overset{a.s.}{\to} \ell(\theta_0)$ on this space, it follows further that $(\mu_n, (\mu_n^*|D_n)) \overset{w}{\to}_w (\ell(\theta_0), (\tilde{\ell}^*|M))$ on a general probability space, as asserted in (3.4.1).

In the case where $g^*(\theta_0) = g(\theta_0)$, it still holds that $\tilde{\theta}^*$ uniquely minimizes q_n^* on Θ^* whenever $g(\tilde{\theta}^*) \geq g^*(\hat{\theta})$, such that $\hat{\theta}^*$ exists and equals $\tilde{\theta}^*$ on the event $\{g(\tilde{\theta}^*) \geq g^*(\hat{\theta})\}$. However, the probability of this event no longer tends to one. Whenever $g(\tilde{\theta}^*) < g^*(\hat{\theta})$, a minimizer of q_n^* on Θ^* exists if and only if a minimizer, say $\check{\theta}^*$, of q_n^* on $\partial\Theta^*$ exists and minimizes q_n^* over the entire Θ^* (this claim is due to the fact that, for outcomes in the event $\{\det(M_n) > 0\}$, the function $q_n^*(\theta)$ is locally minimized uniquely at $\tilde{\theta}^*$). Let $\mathbb{I}_n^* := \mathbb{I}_{\{b(\tilde{\theta}^*)\geq 0\}}$ with $b(\theta) := g(\theta) - g^*(\hat{\theta})$. We show in Section C.1.2 below that $\check{\theta}^*(1-\mathbb{I}_n^*)$, with a measurable $\check{\theta}^*$, is well-defined with P^*-probability approaching one a.s. and $(q_n^*(\check{\theta}^*) - q_n^*(\theta))(1-\mathbb{I}_n^*) \leq 0$ for all $\theta \in \Theta^*$, with P^*-probability approaching one a.s. This establishes the possibility to define the bootstrap estimator $\hat{\theta}^*$ as

$$\hat{\theta}^* = \tilde{\theta}^*\mathbb{I}_n^* + \check{\theta}^*(1-\mathbb{I}_n^*) \tag{C.1.3}$$

and, therefore, the existence of $\hat{\theta}^*$ with P^*-probability approaching one a.s. The existence result carries over to a general probability space with P^*-probability approaching one in probability.

In Section C.1.2 we also show that $\|\check{\theta}^* - \hat{\theta}\|(1-\mathbb{I}_n^*) = O_{p^*}(n^{-1/2})$ a.s., and as a result, $\|\hat{\theta}^* - \hat{\theta}\| = O_{p^*}(n^{-1/2})$ a.s., using also (C.1.2). We do not discuss the uniqueness of $\check{\theta}^*$ but instead we argue next that the measurable minimizers of q_n^* over the bootstrap boundary are asymptotically equivalent, as they give rise to the same asymptotic distribution of $\hat{\theta}^*$.

To accomplish this, we use the result of Section C.1.2 that $\check{\theta}^*$ satisfies a first-order condition [foc] with P^*-probability approaching one a.s. Let dots over function names denote differentiation w.r.t. θ (e.g., $\dot{q}_n^*(\theta) := (\partial q_n^*/\partial\theta')(\theta)$, a column vector). Then the foc takes the form

$$\{\dot{q}_n^*(\check{\theta}^*) + \check{\delta}_n\dot{b}(\check{\theta}^*)\}(1-\mathbb{I}_n^*) = \{\dot{q}_n^*(\check{\theta}^*) + \check{\delta}_n\dot{g}(\check{\theta}^*)\}(1-\mathbb{I}_n^*) = 0,\ b(\check{\theta}^*)(1-\mathbb{I}_n^*) = 0,$$

where $\check{\delta}_n \in \mathbb{R}$ is a Lagrange multiplier. The foc implies, by means of a standard argument, the existence of a measurable $\bar{\theta}^*$ between $\check{\theta}^*$ and $\hat{\theta}$ such that

$$\{n^{1/2}(\check{\theta}^* - \hat{\theta}) - (I_2 - A_n^*\dot{g}(\bar{\theta}^*)')\tilde{\mu}_n^* + A_n^*n^{1/2}b(\hat{\theta})\}(1-\mathbb{I}_n^*) = 0,$$

where $A_n^* := M_n^{-1}\dot{g}(\check{\theta}^*)[\dot{g}(\bar{\theta}^*)'M_n^{-1}\dot{g}(\check{\theta}^*)]^{-1}$ is well-defined with P^*-probability approaching one a.s. As further $\|\breve{\theta}^* - \hat{\theta}\|(1-\mathbb{I}_n^*) = O_{p^*}(n^{-1/2})$ a.s., $\|\bar{\theta}^* - \hat{\theta}\|(1-\mathbb{I}_n^*) = O_{p^*}(n^{-1/2})$ a.s. and $\hat{\theta} - \theta_0 = O(n^{-1/2})$ a.s., using the continuity of $\dot{g}(\theta)$ at θ_0 it follows that

$$\{n^{1/2}(\breve{\theta}^* - \hat{\theta}) - [(I_2 - A^*\dot{g}')\tilde{\mu}_n^* - A^*(\dot{g} - \dot{g}^*)'n^{1/2}(\hat{\theta} - \theta_0)]\}(1-\mathbb{I}_n^*) = o_{p^*}(1) \text{ a.s.,}$$

where $A^* := M^{-1}\dot{g}[\dot{g}'M^{-1}\dot{g}]^{-1}$ and $P^*(|o_{p^*}(1)| > \eta) \overset{a.s.}{\to} 1$ for all $\eta > 0$.

Returning to (C.1.3), we conclude that

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) = \tilde{\mu}_n^*\mathbb{I}_n^* + \{(I_2 - A^*\dot{g}')\tilde{\mu}_n^* - A^*(\dot{g} - \dot{g}^*)'n^{1/2}(\hat{\theta} - \theta_0)\}(1-\mathbb{I}_n^*) + o_{p^*}(1) \text{ a.s.} \quad \text{(C.1.4)}$$

Consider the event indicated by $\mathbb{I}_n^*$. As $\|\hat{\theta}^* - \hat{\theta}\| = O_{p^*}(n^{-1/2})$ a.s. and $\hat{\theta} - \theta_0 = O(n^{-1/2})$ a.s., by the mean value theorem and the continuous differentiability of g, g^* it holds that

$$n^{1/2}b(\tilde{\theta}^*) = \dot{g}'\tilde{\mu}_n^* + (\dot{g} - \dot{g}^*)'\mu_n + o_{p^*}(1) \text{ a.s.}$$

Then $\mathbb{I}_n^* \overset{w^*}{\to}_{a.s.} \mathbb{I}_\infty|(M, \ell(\theta_0))$ with $\mathbb{I}_\infty := \mathbb{I}_{\{\dot{g}'\tilde{\ell}^* \geq (\dot{g}^* - \dot{g})'\ell(\theta_0)\}}$, by (C.1.1)-(C.1.2) and the CMT for weak a.s. convergence (Theorem 10 of Sweeting, 1989), as the probability of the limiting discontinuities is 0: $P(\dot{g}'\tilde{\ell}^* = (\dot{g}^* - \dot{g})'\ell(\theta_0)|(M, \ell(\theta_0))) = 0$ a.s. By exactly the same facts, passage to the limit directly in (C.1.4) yields

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) \overset{w^*}{\to}_{a.s.} \{\tilde{\ell}^*\mathbb{I}_\infty + \breve{\ell}^*(1-\mathbb{I}_\infty)\}|(M, \ell(\theta_0)),\ \breve{\ell}^* := (I_2 - A^*\dot{g}')\tilde{\ell}^* - A^*(\dot{g} - \dot{g}^*)'\ell$$

on the special probability space, where also $\mu_n \overset{a.s.}{\to} \ell(\theta_0)$ by (C.1.1). Therefore, on a general probability space it holds that

$$(\mu_n, (n^{1/2}(\hat{\theta}^* - \hat{\theta})|D_n)) \overset{w^*}{\to}_w (\ell(\theta_0), [\{\tilde{\ell}^*\mathbb{I}_\infty + \breve{\ell}^*(1-\mathbb{I}_\infty)\}|(M, \ell(\theta_0))]).$$

As $I_2 - A^*\dot{g}' = \dot{g}_\perp(\dot{g}_\perp'M\dot{g}_\perp)^{-1}\dot{g}_\perp'M$ and $\tilde{\ell}^* = M^{-1/2}\xi^*$, it follows that

$$\begin{aligned}\tilde{\ell}^*\mathbb{I}_\infty + \breve{\ell}^*(1-\mathbb{I}_\infty) &= \dot{g}_\perp(\dot{g}_\perp'M\dot{g}_\perp)^{-1}\dot{g}_\perp'M^{1/2}\xi^* \\ &\quad + M^{-1}\dot{g}(\dot{g}'M^{-1}\dot{g})^{-1}\max\{(\dot{g}^* - \dot{g})'\ell, \dot{g}'M^{-1/2}\xi^*\},\end{aligned}$$

which is $\arg\min_{\{\dot{g}'\lambda \geq (\dot{g}^* - \dot{g})'\ell\}} ||\lambda - M^{-1/2}\xi^*||_M$ a.s. as asserted in (3.4.2). □

For use in the proof of Corollary 3.4.1, we notice here a useful consequence of the previous argument. Return to the special probability space where

$$(\mu_n, (n^{1/2}(\hat{\theta}^* - \hat{\theta})|D_n)) \overset{w^*}{\to}_{a.s.} (\ell(\theta_0), [\{\tilde{\ell}^*\mathbb{I}_\infty + \breve{\ell}^*(1-\mathbb{I}_\infty)\}|(M, \ell(\theta_0))]).$$

Let $\tau_n := \phi(\mu_n)$, $\tau_n^* := \phi(n^{1/2}(\hat{\theta}^* - \hat{\theta}))$, $\tau := \phi(\ell(\theta_0))$ and $\tau^* := \phi(\tilde{\ell}^*\mathbb{I}_\infty + \breve{\ell}^*(1-\mathbb{I}_\infty))$ for a continuous $\phi : \mathbb{R} \to \mathbb{R}$. Then

$$(\tau_n, (\tau_n^*|D_n)) \overset{w^*}{\to}_{a.s.} (\tau, \tau^*|(M, \ell(\theta_0)))$$

by the CMT of Sweeting (1989). Furthermore, the regular conditional distributions $\tau_n^*|D_n$ converge weakly to the regular conditional distribution $\tau^*|(M, \ell(\theta_0))$ for almost all outcomes; see Theorem 2.2 of Berti, Pratelli and Rigo, (2006). For any fixed outcome such that the previous convergence holds, also $F_n^{*-1}(q_i) \to F_{M,\ell}^{-1}(q_i)$, $i = 1, 2$, hold for the sample paths of the respective conditional quantile functions, provided that q_1, q_2 are continuity points of the sample path of $F_{M,\ell}^{-1}$. If q_1, q_2 are continuity points of almost all sample paths of $F_{M,\ell}^{-1}$, it follows that $F_n^{*-1}(q_i) \to_{a.s.} F_{M,\ell}^{-1}(q_i)$, $i = 1, 2$. Therefore, on a general probability space,

$$(\tau_n, F_n^{*-1}(q_1), F_n^{*-1}(q_2), (\tau_n^*|D_n)) \xrightarrow{w^*}_w (\tau, F_{M,\ell}^{-1}(q_1), F_{M,\ell}^{-1}(q_2), \tau^*|(M, \ell(\theta_0))) \qquad \text{(C.1.5)}$$

provided that $F_{M,\ell}^{-1}$ is a.s. continuous at q_1, q_2.

C.1.2 Details of the proof of Theorem 3.4.1

Let $g^*(\theta_0) = g(\theta_0)$ throughout this subsection. For outcomes such that $\tilde{\theta}^* \notin \Theta^*$ and $\lambda_{\min}(M_n) > 0$, the quadratic function q_n^* is not minimized over Θ^* at any interior point of Θ^* (for otherwise this point would have to be the stationary point $\tilde{\theta}^* \notin \Theta^*$ of q_n^*, a contradiction). For such outcomes, if q_n^* is at all minimized over Θ^*, then this has to occur at a boundary point of Θ^*. Since $\partial\Theta^* \subseteq \{\theta \in \mathbb{R}^2 : g(\theta) = g^*(\hat{\theta})\} =: \tilde{\partial}\Theta^*$, we proceed by constructing a minimizer of q_n^* over the latter set and by showing that this minimizer is in fact a global one over Θ^*. This (and some added measurability considerations) establishes the well-definition of $\check{\theta}^*$ in (C.1.3). Then we establish the $n^{-1/2}$ consistency rate of $\check{\theta}^*$ in the sense that $\|\check{\theta}^* - \hat{\theta}\|(1 - \mathbb{I}_n^*) = O_{p^*}(n^{-1/2})$ a.s.

Step 1. Existence of a minimizer of q_n^* over a portion of $\tilde{\partial}\Theta^*$ close to θ_0. The point $(\theta', c)' = (\theta_0', g(\theta_0))' \in \mathbb{R}^3$ trivially satisfies the equation $g(\theta) = c$. Since g is continuously differentiable in a neighborhood of θ_0 and $\dot{g} = (\dot{g}_1(\theta_0), \dot{g}_2(\theta_0))' \neq 0$ (say that $\dot{g}_1(\theta_0) \neq 0$, with the subscript denoting partial differentiation), by the implicit function theorem there exist an $r > 0$ and a unique function $\gamma : [\theta_{2,0} - r, \theta_{2,0} + r] \times [g(\theta_0) - r, g(\theta_0) + r] \to [\theta_{1,0} - r, \theta_{1,0} + r]$ such that $\gamma(\theta_{2,0}, g(\theta_0)) = \theta_{1,0}$, $g(\gamma(\theta_2, c), \theta_2) = c$; moreover, γ is continuously differentiable. For outcomes such that $|g^*(\hat{\theta}) - g(\theta_0)| \leq r$, the (non-empty) portion of the curve $\tilde{\partial}\Theta^* = \{\theta \in \mathbb{R}^2 : g(\theta) = g^*(\hat{\theta})\}$ contained in the square $\Pi := [\theta_{1,0} - r, \theta_{1,0} + r] \times [\theta_{2,0} - r, \theta_{2,0} + r]$ can be parameterized as $\theta_1 = \gamma(\theta_2, g^*(\hat{\theta}))$, $\theta_2 \in [\theta_{2,0} - r, \theta_{2,0} + r]$. Define $\check{\theta}^* := (\gamma(\check{\theta}_2^*, g^*(\hat{\theta}^r)), \check{\theta}_2^*)'$, where $\check{\theta}_2^*$ is a measurable minimizer of the continuous function $q_n^*(\gamma(\theta_2, g^*(\hat{\theta}^r)), \theta_2)$ over $\theta_2 \in [\theta_{2,0} - r, \theta_{2,0} + r]$, with $\hat{\theta}^r := \hat{\theta}\mathbb{I}_{\{|g^*(\hat{\theta}) - g(\theta_0)| \leq r\}} + \theta_0 \mathbb{I}_{\{|g^*(\hat{\theta}) - g(\theta_0)| > r\}}$. Since $\mathbb{I}_{\{|g^*(\hat{\theta}) - g(\theta_0)| \leq r\}} \xrightarrow{a.s.} 1$ under $g^*(\theta_0) = g(\theta_0)$, it follows that $\check{\theta}^*$ minimizes q_n^* over $\tilde{\partial}\Theta^* \cap \Pi$ with P^*-probability approaching one a.s.

Step 2. Minimization of q_n^* over the entire bootstrap parameter space. For outcomes in

$$\mathcal{A}_n := \{|g^*(\hat{\theta}) - g(\theta_0)| \leq r\} \cap \{g(\tilde{\theta}^*) < g^*(\hat{\theta})\} \cap \{\|\hat{\theta} - \theta_0\| + \|\tilde{\theta}^* - \hat{\theta}\| \leq \tfrac{r}{2}\},$$

the minimum of q_n^* over the entire bootstrap parameter space Θ^* exists and is attained only in Π (e.g., at $\check{\theta}^*$ defined in Step 1), provided that

$$\alpha_n := \lambda_{\min}(M_n)\tfrac{r^2}{4} - \lambda_{\max}(M_n)\|\tilde{\theta}^* - \hat{\theta}\|^2 > 0.$$

To see this, consider $\theta^c := c\hat{\theta} + (1-c)\tilde{\theta}^*$ where $c := \inf\{a \in [0,1] : b(a\hat{\theta} + (1-a)\tilde{\theta}^*) = 0\}$; θ^c is well-defined whenever $g(\tilde{\theta}^*) < g^*(\hat{\theta})$ because $g(\hat{\theta}) \geq g^*(\hat{\theta})$ and b is continuous. Moreover, $\theta^c \in \Pi$ for outcomes in $\mathcal{A}_n$ because $\|\theta^c - \theta_0\| \leq \|\hat{\theta} - \theta_0\| + \|\tilde{\theta}^* - \hat{\theta}\| \leq \frac{r}{2}$ and, hence, $q_n^*(\theta^c) \geq q_n^*(\check{\theta}^*)$ for outcomes in $\mathcal{A}_n$, by the minimizing property of $\check{\theta}^*$ on $\tilde{\partial}\Theta^* \cap \Pi$ and the fact that $b(\theta^c) = 0$. For any $\theta \notin \Pi$ and outcomes in $\mathcal{A}_n$, we therefore find that

$$\begin{aligned}
q_n^*(\theta) - q_n^*(\check{\theta}^*) &\geq q_n^*(\theta) - q_n^*(\theta^c) = q_n^*(\theta) - q_n^*(\tilde{\theta}^*) + q_n^*(\tilde{\theta}^*) - q_n^*(\theta^c) \\
&\geq \lambda_{\min}(M_n)\|\theta - \tilde{\theta}^*\|^2 - \lambda_{\max}(M_n)\|\tilde{\theta}^* - \theta^c\|^2 \\
&\geq \lambda_{\min}(M_n)\{\|\theta - \theta_0\| - \|\tilde{\theta}^* - \theta_0\|\}^2 - \lambda_{\max}(M_n)\|\tilde{\theta}^* - \hat{\theta}\|^2 \\
&\geq \lambda_{\min}(M_n)\{r - \|\tilde{\theta}^* - \hat{\theta}\| - \|\hat{\theta} - \theta_0\|\}^2 - \lambda_{\max}(M_n)\|\tilde{\theta}^* - \hat{\theta}\|^2 \\
&\geq \lambda_{\min}(M_n)\tfrac{r^2}{4} - \lambda_{\max}(M_n)\|\tilde{\theta}^* - \hat{\theta}\|^2 = \alpha_n.
\end{aligned}$$

Thus, for outcomes in $\mathcal{A}_n \cap \{\alpha_n > 0\}$, q_n^* out of Π is larger than $\min_{\theta \in \tilde{\partial}\Theta^* \cap \Pi} q_n^*(\theta)$. As $\tilde{\partial}\Theta^* \subseteq \Theta^*$, it follows that $\min_{\theta \in \Theta^* \cap \Pi} q_n^*(\theta)$ (which exists) for such outcomes is actually $\min_{\theta \in \Theta^*} q_n^*(\theta)$. Moreover,

$$\min_{\theta \in \Theta^*} q_n^*(\theta) = \min_{\theta \in \Theta^* \cap \Pi} q_n^*(\theta) = \min_{\theta \in \tilde{\partial}\Theta^* \cap \Pi} q_n^*(\theta) = \min_{\theta \in \partial\Theta^* \cap \Pi} q_n^*(\theta),$$

for if $\min_{\theta \in \Theta^* \cap \Pi} q_n^*(\theta) < \min_{\theta \in \partial\Theta^* \cap \Pi} q_n^*(\theta)$, then $\min_{\theta \in \Theta^* \cap \Pi} q_n^*(\theta)$ (and thus, $\min_{\theta \in \Theta^*} q_n^*(\theta)$) is achieved at an interior point of Θ^*, which can only be $\tilde{\theta}^*$, a contradiction with $\tilde{\theta}^* \notin \Theta^*$ (i.e., with $g(\tilde{\theta}^*) < g^*(\hat{\theta})$). To summarize, for outcomes in $\mathcal{A}_n \cap \{\alpha_n > 0\}$, $\check{\theta}^*$ minimizes q_n^* over Θ^* and is at the boundary of Θ^*.

We find the associated probability

$$\begin{aligned}
&P^*\left((1 - \mathbb{I}_n^*)q_n^*(\check{\theta}^*) < (1 - \mathbb{I}_n^*)q_n(\theta) \quad \forall \theta \in \Theta^* \setminus \Pi\right) \\
&\quad \geq P^*\left(|g^*(\hat{\theta}) - g(\theta_0)| \leq r, \|\hat{\theta} - \theta_0\| \leq \tfrac{r}{4}, \|\tilde{\theta}^* - \hat{\theta}\| \leq \frac{r}{4}, \alpha_n > 0\right) \\
&\quad = \mathbb{I}_{\{|g^*(\hat{\theta}) - g(\theta_0)| \leq r\} \cap \{\|\hat{\theta} - \theta_0\| \leq r/4\}} P^*\left(\|\tilde{\theta}^* - \hat{\theta}\| \leq \tfrac{r}{4}, \alpha_n > 0\right) \overset{a.s.}{\to} 1
\end{aligned}$$

because $g(\hat{\theta}) \overset{a.s.}{\to} g(\theta_0)$, $\lambda_{\min}(M_n) \to \lambda_{\min}(M) > 0$ a.s., $\lambda_{\max}(M_n) \to \lambda_{\max}(M) < \infty$ a.s. and $\|\tilde{\theta}^* - \hat{\theta}\| \overset{w^*}{\to}_{a.s.} 0$. This establishes the fact that $\hat{\theta}^*$ of (C.1.3), with $\check{\theta}^*$ as defined in Step 1, minimizes q_n^* over the bootstrap parameter space Θ^* with P^*-probability approaching one a.s.

Step 3. Consistency rate of $\check{\theta}^*$. Similarly to Step 2, for outcomes in $\mathcal{A}_n$,

$$0 \geq q_n^*(\check{\theta}^*) - q_n^*(\theta^c) \geq \lambda_{\min}(M_n)\|\check{\theta}^* - \tilde{\theta}^*\|^2 - \lambda_{\max}(M_n)\|\tilde{\theta}^* - \hat{\theta}\|^2,$$

the first inequality by the minimizing property of $\check{\theta}^*$ over $\tilde{\partial}\Theta^* \cap \Pi$. Therefore,

$$\begin{aligned} & P^*\left((1-\mathbb{I}_n^*)\|\check{\theta}^*-\tilde{\theta}^*\|^2 \leq (1-\mathbb{I}_n^*)\frac{\lambda_{\max}(M_n)}{\lambda_{\min}(M_n)}\|\tilde{\theta}^*-\hat{\theta}\|^2\right) \\ \geq \;& P^*\left(|g^*(\hat{\theta})-g(\theta_0)| \leq r, \|\hat{\theta}-\theta_0\| \leq \tfrac{r}{4}, \|\tilde{\theta}^*-\hat{\theta}\| \leq \tfrac{r}{4}\right) \\ = \;& \mathbb{I}_{\{|g^*(\hat{\theta})-g(\theta_0)|\leq r\}\cap\{\|\hat{\theta}-\theta_0\|\leq r/4\}} P^*\left(\|\tilde{\theta}^*-\hat{\theta}\| \leq \tfrac{r}{4}\right) \overset{a.s.}{\to} 1. \end{aligned}$$

As $\lambda_{\max}(M_n)/\lambda_{\min}(M_n) \overset{a.s.}{\to} \lambda_{\max}(M)/\lambda_{\min}(M)$ and $\|\tilde{\theta}^*-\hat{\theta}\| = O_{p^*}(n^{-1/2})$ P-a.s. (the latter, by (C.1.2)), it follows that $(1-\mathbb{I}_n^*)\|\check{\theta}^*-\tilde{\theta}^*\| = O_{p^*}(n^{-1/2})$ P-a.s. and $\|\hat{\theta}^*-\tilde{\theta}^*\| = O_{p^*}(n^{-1/2})$ P-a.s. for $\hat{\theta}^*$ of (C.1.3). Thus, $\hat{\theta}^*$ has the same consistency rate as $\tilde{\theta}^*$. This argument applies to any $\check{\theta}^*$ which is measurable and minimizes q_n^* over $\tilde{\partial}\Theta^* \cap \Pi$ for outcomes in $\mathcal{A}_n$. This completes Step 3.

Finally, consider the first-order condition [foc] for minimization of q_n^* on $\tilde{\partial}\Theta^*$. As $\|\check{\theta}^*-\theta_0\|(1-\mathbb{I}_n^*) \leq \{\|\check{\theta}^*-\tilde{\theta}^*\|+\|\check{\theta}^*-\theta_0\|\}(1-\mathbb{I}_n^*) \overset{w^*}{\to}_{a.s.} 0$, it follows that $\mathbb{I}_{\{\check{\theta}^*\in \text{int}(\Pi)\}}(1-\mathbb{I}_n^*)+\mathbb{I}_n^* \overset{w}{\to}_{a.s.} 1$. As additionally $\dot{g}(\theta_0) \neq 0$, by continuity of $\dot{g}(\theta) := (\partial g/\partial\theta')(\theta)$, the foc takes the form

$$P^*\left(\{\dot{q}_n(\check{\theta}^*)+\check{\delta}_n\dot{g}(\check{\theta}^*)\}(1-\mathbb{I}_n^*)=0\right) \overset{a.s.}{\to} 1,$$

where $\check{\delta}_n \in \mathbb{R}$ are measurable Lagrange multipliers that can be determined, for outcomes in the event $\mathbb{I}_n^* = 1$, by involving also the constraint $b(\check{\theta}^*)(1-\mathbb{I}_n^*) = 0$. □

C.1.3 Proof of Corollary 3.4.1

We only discuss the bootstrap validity part of the corollary, as the convergence part (3.4.3) was explained in the main text.

Let $\tau_n := \phi(n^{1/2}(\hat{\theta}-\theta_0))$, $\tau_n^* := \phi(n^{1/2}(\hat{\theta}^*-\hat{\theta}))$ and $\tau := \phi(\ell(\theta_0))$. Convergence (3.4.3) and the continuity of ϕ imply that

$$(\tau_n, (\tau_n^*|D_n)) \overset{w}{\to}_w (\tau, (\tau|M)).$$

If the (random) cdf of $\tau|M$ is sample-path continuous, bootstrap validity follows from Theorem 3.1 and Lemma A.2(b) of Cavaliere and Georgiev (2020). We reduce the general case to the globally continuous case by a local argument for the cdf's $F(\cdot) := P(\tau \leq \cdot)$ and $F_M(\cdot) := P(\tau \leq \cdot|M)$. For concreteness, we focus on the technically more involved possibility $g(\theta_0) = 0$, such that $\theta_0 \in \partial\Theta$ given the assumption $\dot{g} \neq 0$. With

$$l(B) := \dot{g}_\perp(\dot{g}_\perp' B\dot{g}_\perp)^{-1}\dot{g}_\perp' B^{1/2}\xi + B^{-1}\dot{g}(\dot{g}'B^{-1}\dot{g})^{-1}\max\{0, \dot{g}'B^{-1/2}\xi\}$$

for positive definite $B \in \mathbb{R}^{2\times 2}$ and with $\ell = l(M)$, notice the following. If B is a fixed positive definite matrix such that

$$P\left(\phi\left(l(B)\right) = a\right) > 0 \qquad \text{(C.1.6)}$$

for some $a \in \mathbb{R}$, then by equivalence (i.e., mutual absolute continuity) considerations for non-singular Gaussian distributions, also

$$P\left(\phi\left(l(D)\right)=a\right)>0$$

for any positive definite $D\in\mathbb{R}^{2\times 2}$. In fact, let $\psi:\mathbb{R}\to\mathbb{R}$ be defined as $\psi(\cdot):=\phi(\dot{g}_\perp(\cdot))$ and let $\phi^{\leftarrow}(\cdot),\psi^{\leftarrow}(\cdot)$ denote inverse images. Then the probability in (C.1.6) equals

$$\begin{aligned}
P\left(l(B)\in\phi^{\leftarrow}(\{a\})\cap\partial\Lambda\right)+P(l(B)\in\phi^{\leftarrow}(\{a\})\cap\mathrm{int}\Lambda)&\\
&=P(\{\dot{g}'B^{-1/2}\xi\le 0\}\cap\{(\dot{g}_\perp'B\dot{g}_\perp)^{-1}\dot{g}_\perp'B^{1/2}\xi\in\psi^{\leftarrow}(\{a\})\})\\
&+P(\{\dot{g}'B^{-1/2}\xi>0\}\cap\{B^{-1/2}\xi\in\phi^{\leftarrow}(\{a\})\})\\
&=P(\dot{g}'B^{-1/2}\xi\le 0)P((\dot{g}_\perp'B\dot{g}_\perp)^{-1}\dot{g}_\perp'B^{1/2}\xi\in\psi^{\leftarrow}(\{a\})\})\\
&+P(B^{-1/2}\xi\in\phi^{\leftarrow}(\{a\})\cap\mathrm{int}\Lambda),
\end{aligned}$$

the equality because $Cov(\dot{g}'B^{-1/2}\xi,(\dot{g}_\perp'B\dot{g}_\perp)^{-1}\dot{g}_\perp'B^{1/2}\xi)=0$ and ξ is Gaussian. In the previous display, $P(\dot{g}'B^{-1/2}\xi\le 0)=P(N(0,\dot{g}'B^{-1}\dot{g})\le 0)>0$ for all positive definite B,

$$P((\dot{g}_\perp'B\dot{g}_\perp)^{-1}\dot{g}_\perp'B^{1/2}\xi\in\psi^{\leftarrow}(\{a\})\})=P(N(0,(\dot{g}_\perp'B\dot{g}_\perp)^{-1})\in\psi^{\leftarrow}(\{a\}))$$

is either 0 for all positive definite B or positive for all positive definite B, and the same applies to $P(B^{-1/2}\xi\in\phi^{\leftarrow}(\{a\})\cap int\Lambda)$. Therefore, the sign of the probability in (C.1.6) is the same (zero or positive) for all positive definite B.

The cdf F_M is a measurable transformation of M determined a.s. uniquely by the distribution of (M,ξ); it can be identified (up to a set of measure zero) as

$$F_M(\cdot)=P\left(\phi\left(l(B)\right)\le\cdot\right)|_{B=M}$$

by virtue of the independence of M and ξ. Since M is positive definite a.s., from the argument in the previous paragraph we can conclude that every point on the line is either a discontinuity point of almost all sample paths of F_M, or a continuity point of almost all sample paths of F_M. By averaging, a point on the line is a discontinuity point of F if and only if it is a discontinuity point of almost all sample paths of F_M.

Let now q_0 be an interior point of the set

$$C=\{q\in(0,1):\lim_{n\to\infty}P(F(\tau_n)\le q)\to q|\mathsf{H}_0\}$$

such that the asymptotic test is correctly sized for $q\in(q_0-2\epsilon,q_0+2\epsilon)\subset(0,1)$ for some $\epsilon>0$. As $\tau_n\xrightarrow{w}\tau\sim F$, this implies that F and (by the discussion in previous paragraph) F_M skip no values from the interval $(q_0-2\epsilon,q_0+2\epsilon)$ (for F_M, a.s.). In particular, almost all sample paths of F_M are continuous on the (random) open superset $\left(F_M^{-1}(q_0-\frac{3}{2}\epsilon),F_M^{-1}(q_0+\frac{3}{2}\epsilon)\right)$ of $I_\epsilon:=[F_M^{-1}(q_0-\epsilon),F_M^{-1}(q_0+\epsilon)]$, with

$$F_M^{-1}(q_0-\frac{3}{2}\epsilon)<F_M^{-1}(q_0-\epsilon)<F_M^{-1}(q_0+\epsilon)<F_M^{-1}(q_0+\frac{3}{2}\epsilon)\quad\text{a.s.}\qquad\text{(C.1.7)}$$

Without loss of generality, ϵ can be considered such that $q_0 \pm \epsilon$ are continuity points of F_M^{-1} a.s. (because F_M^{-1} is chosen to be càdlàg and its discontinuity points on, say $[\frac{q_0}{2}, \frac{q_0+1}{2}]$ are countably many). Let $\Psi^-(a, x)$ and $\Psi^+(a, x)$ be generalized inverses of the cdf's of a standard Gaussian variable conditioned to take values respectively in $(-\infty, a]$ and $[a, \infty)$. On extensions of the probability spaces where the data and (τ, M) are defined, consider a $U_{[0,1]}$ variable υ. Define $F_n^*(\cdot) := P^*(\tau_n^* \leq \cdot)$, $I_{n,\epsilon} := [F_n^{*-1}(q_0 - \epsilon), F_n^{*-1}(q_0 + \epsilon)]$ and

$$
\begin{aligned}
\tilde{\tau}_n &= \tau_n \mathbb{I}_{\{\tau_n \in I_{n,\epsilon}\}} + \Psi^-(F_n^{*-1}(q_0 - \epsilon), \upsilon)\mathbb{I}_{\{\tau_n < F_n^{*-1}(q_0-\epsilon)\}} \\
&\quad + \Psi^+(F_n^{*-1}(q_0 + \epsilon), \upsilon)\mathbb{I}_{\{\tau_n > F_n^{*-1}(q_0+\epsilon)\}}, \\
\tilde{\tau}_n^* &= \tau_n^* \mathbb{I}_{\{\tau_n^* \in I_{n,\epsilon}\}} + \Psi^-(F_n^{*-1}(q_0 - \epsilon), \upsilon)\mathbb{I}_{\{\tau_n^* < F_n^{*-1}(q_0-\epsilon)\}} \\
&\quad + \Psi^+(F_n^{*-1}(q_0 + \epsilon), \upsilon)\mathbb{I}_{\{\tau_n^* > F_n^{*-1}(q_0+\epsilon)\}}, \\
\tilde{\tau} &= \tau \mathbb{I}_{\{\tau \in I_\epsilon\}} + \Psi^-(F_M^{-1}(q_0 - \epsilon), \upsilon)\mathbb{I}_{\{\tau < F_M^{-1}(q_0-\epsilon)\}} \\
&\quad + \Psi^+(F_M^{-1}(q_0 + \epsilon), \upsilon)\mathbb{I}_{\{\tau > F_M^{-1}(q_0+\epsilon)\}}.
\end{aligned}
$$

Then

$$(\tilde{\tau}_n, (\tilde{\tau}_n^*|D_n)) \xrightarrow{w}_w (\tilde{\tau}, (\tilde{\tau}|M))$$

because

$$(f_1(\tilde{\tau}_n), E\{f_2(\tilde{\tau}_n^*)|D_n\}) \xrightarrow{w} (f_1(\tilde{\tau}_n), E\{f_2(\tilde{\tau})|M\})$$

for any continuous and bounded real functions f_1, f_2, as a result of (C.1.5) with $\tau^*|(M, \ell(\theta_0) = \tau|M$ in the sense of a.s. equality of conditional distributions and the fact that $P(\tau = F_M^{-1}(q_0 \pm \epsilon)|M) = 0$ a.s. by sample-path continuity of F_M an open superset of I_ϵ. As the cdf of $\tilde{\tau}|M$ is a.s. sample-path continuous by construction, it follows that $P^*(\tilde{\tau}_n^* \leq \tilde{\tau}_n) \xrightarrow{w} U_{[0,1]}$, by Theorem 3.1 and Lemma A.2(b) of Cavaliere and Georgiev (2020).

Let $\tilde{F}_n^*(\cdot) := P^*(\tilde{\tau}_n^* \leq \cdot)$. We now return to the original variables. By considerations of equalities of events, it holds that

$$P(F_n^*(\tau_n) \leq q_0) = P(F_n^*(\tilde{\tau}_n) \leq q_0) = P(\tilde{F}_n^*(\tilde{\tau}_n) \leq q_0) = P(P^*(\tilde{\tau}_n^* \leq \tilde{\tau}_n) \leq q_0) = q_0$$

using the fact that $P^*(\tilde{\tau}_n^* \leq \tilde{\tau}_n) \xrightarrow{w} U_{[0,1]}$. This completes the proof.

Table 1: Empirical rejection probabilities (ERPs) of bootstrap tests under the null.

Nominal level: 0.05

		$\theta_0 = (0,0)'$					$\theta_0 = (-0.75, 0.75)'$					$\theta_0 = (-1.50, 1.50)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0
ξ_1	100	4.2	4.7	5.0	5.3	5.4	6.9	7.0	7.2	7.3	7.5	6.3	6.3	6.3	6.4	6.5
	400	3.9	4.8	5.1	5.3	5.3	5.5	5.6	5.8	6.2	6.7	5.3	5.3	5.3	5.3	5.3
	800	3.7	4.8	5.1	5.2	5.2	5.2	5.3	5.4	5.6	6.2	5.2	5.2	5.2	5.2	5.2
ξ_2	100	4.2	4.7	5.0	5.3	5.5	7.1	7.3	7.4	7.6	7.8	6.2	6.3	6.3	6.4	6.5
	400	3.8	4.7	5.0	5.1	5.2	5.7	5.9	6.1	6.4	6.9	5.3	5.3	5.3	5.3	5.3
	800	3.6	4.6	4.8	4.9	4.9	5.1	5.2	5.3	5.5	6.0	5.1	5.1	5.1	5.1	5.1
ξ_3	100	4.3	4.7	5.0	5.3	5.5	7.1	7.2	7.3	7.5	7.7	6.4	6.4	6.4	6.5	6.6
	400	3.7	4.6	4.9	5.1	5.1	5.5	5.7	5.9	6.2	6.7	5.2	5.2	5.2	5.2	5.2
	800	3.7	4.8	5.0	5.1	5.2	5.1	5.2	5.3	5.5	6.0	5.1	5.1	5.1	5.1	5.1

Nominal level: 0.10

		$\theta_0 = (0,0)'$					$\theta_0 = (-0.75, 0.75)'$					$\theta_0 = (-1.50, 1.50)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0
ξ_1	100	8.0	9.0	9.7	10.3	10.6	13.0	13.3	13.6	14.1	14.6	11.5	11.6	11.6	11.7	11.8
	400	7.7	9.5	10.1	10.4	10.5	10.4	10.5	10.8	11.3	12.4	10.3	10.3	10.3	10.3	10.3
	800	7.4	9.4	9.9	10.1	10.1	10.4	10.4	10.5	10.7	11.5	10.1	10.1	10.1	10.1	10.1
ξ_2	100	8.1	9.0	9.7	10.3	10.5	13.2	13.5	13.8	14.3	14.7	11.3	11.3	11.4	11.5	11.6
	400	7.5	9.2	9.9	10.2	10.3	10.7	10.9	11.1	11.6	12.5	10.2	10.2	10.2	10.2	10.3
	800	7.2	9.2	9.8	10.0	10.0	10.2	10.3	10.3	10.5	11.3	10.3	10.3	10.3	10.3	10.3
ξ_3	100	8.3	9.2	9.9	10.5	10.8	13.3	13.7	14.0	14.5	15.0	11.7	11.7	11.8	11.9	12.0
	400	7.6	9.4	10.0	10.3	10.3	10.4	10.5	10.8	11.3	12.4	10.2	10.2	10.2	10.2	10.2
	800	7.4	9.3	9.9	10.1	10.1	10.1	10.1	10.2	10.4	11.2	10.0	10.0	10.0	10.0	10.0

Note: bootstrap tests are based on a standard fixed-regressor wild bootstrap (b_1) and on the proposed corrected wild bootstrap method (b_2) of Section 4, using $g^* = g - |g|^{1+\kappa}$. ERPs are estimated using 50,000 Monte Carlo replications and 999 bootstrap repetitions. The column "dist." shows the distributions of ε_t: $\xi_1 \sim iidN(0,1)$, $\xi_2 \sim ARCH(1)$ and $\xi_3 = \sqrt{0.5}v_t + \sqrt{0.5}\varepsilon_{x,t}$, where $v_t \sim iidN(0,1)$ and $\varepsilon_{x,t}$ is the error term of the predictive variable $x_{n,t}$.

Table 2: Empirical rejection probabilities (ERPs) of bootstrap tests under local alternatives.

Nominal level: 0.05

		$a_0 = (-3, 0)'$					$a_0 = (3, 0)'$					$a_0 = (5, 0)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0
ξ_1	100	21.0	21.0	21.1	21.2	21.3	40.6	40.9	41.0	41.0	41.0	68.0	68.0	68.0	68.0	68.0
	400	18.9	19.1	19.3	19.4	19.5	38.5	38.8	38.8	38.8	38.8	64.9	64.9	64.9	64.9	64.9
	800	18.6	18.8	19.0	19.1	19.1	37.6	37.9	37.9	37.9	38.0	64.0	64.0	64.0	64.0	64.0
ξ_2	100	21.7	21.8	21.9	22.0	22.1	41.9	42.1	42.2	42.2	42.2	68.5	68.5	68.5	68.5	68.5
	400	19.2	19.4	19.5	19.7	19.8	38.3	38.7	38.7	38.7	38.7	64.7	64.8	64.8	64.8	64.8
	800	18.6	18.8	19.0	19.1	19.1	37.8	38.1	38.1	38.1	38.1	64.2	64.2	64.2	64.2	64.2
ξ_3	100	20.6	20.7	20.8	20.8	21.0	40.8	41.0	41.1	41.1	41.1	67.3	67.3	67.3	67.3	67.3
	400	19.0	19.1	19.3	19.4	19.4	38.1	38.4	38.5	38.5	38.5	65.0	65.0	65.0	65.0	65.0
	800	18.3	18.5	18.8	18.9	18.9	37.7	38.0	38.1	38.1	38.1	63.5	63.5	63.5	63.5	63.5

Nominal level: 0.10

		$a_0 = (-3, 0)'$					$a_0 = (3, 0)'$					$a_0 = (5, 0)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0
ξ_1	100	29.6	29.8	29.9	30.1	30.3	54.7	55.0	55.1	55.2	55.2	81.7	81.7	81.8	81.8	81.8
	400	27.0	27.3	27.7	28.1	28.2	52.2	52.6	52.7	52.7	52.7	79.6	79.6	79.6	79.6	79.6
	800	26.4	26.9	27.3	27.6	27.6	51.7	52.1	52.2	52.2	52.2	78.7	78.7	78.7	78.7	78.7
ξ_2	100	30.2	30.4	30.6	30.8	31.0	55.7	55.9	56.0	56.0	56.0	82.0	82.0	82.0	82.0	82.0
	400	27.1	27.4	27.9	28.2	28.3	51.8	52.1	52.2	52.2	52.2	79.3	79.3	79.3	79.3	79.3
	800	26.6	27.0	27.5	27.7	27.7	51.5	51.9	51.9	51.9	51.9	78.6	78.6	78.6	78.6	78.6
ξ_3	100	29.1	29.3	29.4	29.7	29.9	54.2	54.5	54.6	54.6	54.6	80.9	80.9	80.9	80.9	80.9
	400	26.7	27.0	27.4	27.7	27.8	51.7	52.1	52.2	52.2	52.2	79.4	79.4	79.4	79.4	79.4
	800	26.2	26.6	27.1	27.3	27.3	51.3	51.7	51.7	51.7	51.7	78.5	78.5	78.5	78.5	78.5

Note: bootstrap tests are based on a standard fixed-regressor wild bootstrap (b_1) and on the proposed corrected wild bootstrap method (b_2) of Section 4, using $g^* = g - |g|^{1+\kappa}$. ERPs are estimated using 50,000 Monte Carlo replications and 999 bootstrap repetitions. The column "dist." shows the distributions of ε_t: $\xi_1 \sim iidN(0,1)$, $\xi_2 \sim ARCH(1)$ and $\xi_3 = \sqrt{0.5}v_t + \sqrt{0.5}\varepsilon_{x,t}$, where $v_t \sim iidN(0,1)$ and $\varepsilon_{x,t}$ is the error term of the predictive variable $x_{n,t}$.

C.2 Alternative data generating processes

The asymptotic theory in the paper is presented under the assumption that $x_{n,t}$ is a unit-root non-stationary process. Here we show that the choice of a bootstrap parameter space is fundamental for bootstrap validity also under alternative stochastic specifications for $x_{n,t}$, e.g., a near-unit root and a stationary specification. More importantly, a common definition of the bootstrap parameter space could be appropriate for all the considered specifications of $x_{n,t}$. Still, the functional forms of the limit distributions are not identical across the specifications of $x_{n,t}$ and, in the stationary case, we perform OLS estimation under the additional constraint $\hat{\delta} = 0$ in (3.3.2). The implications for bootstrap inference are discussed below.

C.2.1 Near-unit root regressor

Consider a modification of Assumption 1 where in part (c) the limit process becomes

$$(X,Z)' = \left(\int^{\cdot} e^{c(s-\cdot)}dW(s), Z\right)', \quad c>0,$$

for a Brownian motion $(W,Z)' \sim BM(0,\Omega)$. Thus, X is an Ornstein-Uhlenbeck process originating from a near-UR posited predicting variable $x_{n,t}$. The asymptotic distribution of $\hat{\theta}$ has a more complex structure than in the unit root case. Now $n^{1/2}(\hat{\theta}-\theta_0) \xrightarrow{w} M^{-1/2}\xi + v_c$ with $v_c := (0, c\omega_{xz}\omega_{xx}^{-1})'$ if $\theta_0 \in int\Theta$. On the other hand,

$$n^{1/2}(\hat{\theta}-\theta_0) \xrightarrow{w} \arg\min_{\lambda\in\Lambda} ||\lambda - M^{-1/2}\xi - v_c||_M, \quad \Lambda := \{\lambda \in \mathbb{R}^2 : \dot{g}'\lambda \geq 0\} \tag{C.2.1}$$

if $g(\theta_0)=0$. The limiting shift by v_c is due to the fact that $n^{1/2}\Delta x_{n,t}$ in the near-unit root case is not a sufficiently good proxy for the innovations driving $x_{n,t}$. Eqs. (3.3.5)–(3.3.6) for the standard bootstrap hold in the near-unit root case if X in the definition of M is understood as an Ornstein-Uhlenbeck process. Therefore, the possibility that $\theta_0 \in \partial\Theta$ induces the same kind of limiting bootstrap randomness as in the exact unit-root case. Additionally, the bootstrap limit distribution does not replicate the shift in the limit distribution of $n^{1/2}(\hat{\theta}-\theta_0)$ induced by the vector v_c, as a consequence of the conditional independence of the bootstrap innovations and the regressor $x_{n,t-1}$. This fact is not related to the position of θ_0 relative to Θ and requires separate treatment. Consider now the bootstrap estimator of Corollary 3.4.1 with the choice $g^* = g - |g|^{1+\kappa}$ for $\kappa > 0$. In the case where $x_{n,t}$ is near-unit root non-stationary, instead of (3.4.3) it holds that

$$(n^{1/2}(\hat{\theta}-\theta_0), (n^{1/2}(\hat{\theta}^*-\hat{\theta})|D_n)) \xrightarrow{w}_w \left(M^{-1/2}\xi + v_c, (M^{-1/2}\xi|M)\right)$$

if $g(\theta_0) > 0$, and

$$(n^{1/2}(\hat{\theta}-\theta_0), (n^{1/2}(\hat{\theta}^*-\hat{\theta})|D_n)) \xrightarrow{w}_w \Bigg(\arg\min_{\lambda\in\Lambda} ||\lambda - M^{-1/2}\xi - v_c||_M, \left(\arg\min_{\lambda\in\Lambda} ||\lambda - M^{-1/2}\xi||_M \Big| M\right)\Bigg)$$

if $g(\theta_0) = 0$, where X in the definition of M should again be read as an Ornstein-Uhlenbeck process. This means that g^* still does the job it is designed for (remove the random shift from the half-plane in the limiting bootstrap distribution). Nevertheless, bootstrap invalidity due to the limiting shift by v_c, not related to the position of θ_0 in Θ, remains to be tackled.

C.2.2 Stationary regressor

If $x_{n,t} = x_t$ is stationary, then the inclusion of $\Delta x_{n,t} = \Delta x_t$ among the regressors of (3.3.2) will, in general, compromise the consistency of $\hat{\theta}$ for the true value θ_0 in the predictive regression (3.3.1). Assume, however, that $n^{-1}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t' \xrightarrow{p} M$ for $\tilde{x}_t := (1, x_{n,t-1})'$ and a non-random positive definite matrix M, and that the unconstrained OLS estimator of θ from the predictive regression (3.3.1) is consistent at the $n^{-1/2}$ rate and has asymptotic $N(0, \omega_{zz}M^{-1})$ distribution. Then the constrained OLS estimator $\hat{\theta}$ of (3.3.1) subject to $g(\hat{\theta}) \geq 0$ (equivalently, the constrained OLS estimator of (3.3.2) subject to $g(\hat{\theta}) \geq 0$, $\hat{\delta} = 0$) satisfies $n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{w} \ell_{st}(\theta_0) = \tilde{\ell}_{st} := M^{-1/2}\zeta$ with $\zeta \sim N(0, \omega_{zz}I_2)$ in the case where $\theta_0 \in int\Theta$, and

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{w} \ell_{st}(\theta_0) = \ell_{st} := \arg\min_{\lambda\in\Lambda} ||\lambda - M^{-1/2}\zeta||_M, \quad \Lambda := \{\lambda \in \mathbb{R}^2 : \dot{g}'\lambda \geq 0\}$$

in the case where $g(\theta_0) = 0$. In the stationary case with a non-random limiting M, the limiting behavior of the standard bootstrap is entirely analogous to the introductory location model example, as the possibility that $\theta_0 \in \partial\Theta$ is the only source of bootstrap randomness in the limit. For $\hat{\theta}$ defined in the previous paragraph, it holds that $n^{1/2}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_p M^{-1/2}\zeta^*$ with $\zeta^* \sim N(0, \omega_{zz}I_2)$ in the case where $\theta_0 \in int\Theta$, such that the limit bootstrap distribution is non-random in this case, and

$$n^{1/2}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_w \left(\arg\min_{\lambda\in\Lambda_\ell^*} ||\lambda - M^{-1/2}\zeta^*||_M\right)\Big|\ell, \quad \Lambda_\ell^* := \{\lambda \in \mathbb{R}^2 : \dot{g}'\lambda \geq -\dot{g}'\ell\},$$

with $\zeta^*|\ell \sim N(0, \omega_{zz}I_2)$ in the case where $g(\theta_0) = 0$. We conclude that the same discrepancy between Λ and Λ_ℓ^* emerges in the case $g(\theta_0) = 0$ irrespective of the stochastic properties of the regressor. Consider now the bootstrap estimator of Corollary 3.4.1 with the choice $g^* = g - |g|^{1+\kappa}$ for $\kappa > 0$. For a stationary $x_{n,t}$ and a non-random M, the original and the bootstrap estimators satisfy

$$(n^{1/2}(\hat{\theta} - \theta_0), (n^{1/2}(\hat{\theta}^* - \hat{\theta})|D_n)) \xrightarrow{w}_p (\ell_{st}(\theta_0), \ell_{st}(\theta_0))$$

and bootstrap validity is restored as in Corollary 3.4.1, in particular because the random shift from the half-plane in the limiting bootstrap distribution is again removed.

C.2.3 Concluding remarks

An inferential framework that would be asymptotically valid in the unit root, near-unit root, and stationary cases, allowing the researcher to remain agnostic to the stochastic properties of the regressor, could be based on two main ingredients. First, the definition of the bootstrap parameter space in a way such that it approximates sufficiently well the geometry of the original parameter space; e.g., by setting $g^* = g - |g|^{1+\kappa}$ in the definition of Θ^* for some $\kappa > 0$, see above. This definition is independent of the stochastic properties of the regressor. Second, the use of an estimator (different from our choice of OLS) that gives rise to limit distributions that (a) in the near-unit root case depend on c only through the process X (and thus, the matrix M), but are free from shifts in the direction of v_c, and (b) allow for a common treatment of the contemporaneous correlation between the innovations of the predictive regression and the shocks driving $x_{n,t}$ (vs. the inclusion or omission of $\Delta x_{n.t}$ in the estimated eq. (3.3.2)). We conjecture that constrained versions of both the IVX (extended instrumental variables) estimator and the associated bootstrap schemes as discussed in Demetrescu et al. (2023) would give rise to asymptotically valid bootstrap inference. A detailed discussion is beyond the scope of this appendix due to our focus on issues attributable to the boundary of the parameter space.

C.3 Additional Monte Carlo simulations

In this section, we present additional numerical results in support of the theoretical arguments provided in CGZ. In particular, Tables S.1 and S.2 refer to the same testing procedure considered in Tables 1 and 2 in CGZ, respectively, but focus on the case $g^* = g_2^* := g - n^{-\kappa}|g|$. Furthermore, in Tables S.3 and S.4 we present the simulated ERPs of bootstrap tests under local alternatives such that $\theta_0 \in int(\Theta)$, using $g^* = g_1^*$ and $g^* = g_2^*$, respectively.

Table S1: Empirical rejection probabilities (ERPs) of bootstrap tests under the null.

Nominal level: 0.05																
		$\theta_0 = (0,0)'$					$\theta_0 = (-0.75, 0.75)'$					$\theta_0 = (-1.50, 1.50)'$				
		b_1	b_2 κ				b_1	b_2 κ				b_1	b_2 κ			
dist.	n		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40
ξ_1	100	4.2	4.9	5.3	5.5	5.6	6.9	7.0	7.3	8.3	9.6	6.3	6.4	6.6	7.3	9.6
	400	3.9	4.8	5.1	5.3	5.3	5.5	5.7	6.0	7.1	9.2	5.3	5.3	5.3	5.7	8.6
	800	3.7	4.7	5.0	5.2	5.2	5.2	5.3	5.6	6.7	9.4	5.2	5.2	5.2	5.3	8.4
ξ_2	100	4.2	4.9	5.3	5.6	5.7	7.1	7.3	7.5	8.4	9.9	6.2	6.4	6.6	7.2	9.5
	400	3.8	4.6	5.0	5.1	5.2	5.7	6.0	6.3	7.3	9.4	5.3	5.3	5.3	5.7	8.7
	800	3.6	4.5	4.8	4.9	4.9	5.1	5.2	5.5	6.7	9.3	5.1	5.1	5.1	5.3	8.6
ξ_3	100	4.3	4.9	5.3	5.6	5.7	7.1	7.2	7.4	8.5	9.9	6.4	6.5	6.7	7.4	9.8
	400	3.7	4.6	4.9	5.1	5.1	5.5	5.8	6.1	7.2	9.3	5.2	5.2	5.2	5.6	8.6
	800	3.7	4.6	5.0	5.1	5.2	5.1	5.2	5.4	6.5	9.1	5.1	5.1	5.1	5.3	8.4
Nominal level: 0.10																
		$\theta_0 = (0,0)'$					$\theta_0 = (-0.75, 0.75)'$					$\theta_0 = (-1.50, 1.50)'$				
		b_1	b_2 κ				b_1	b_2 κ				b_1	b_2 κ			
dist.	n		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40
ξ_1	100	8.0	9.1	9.9	10.5	10.7	13.0	13.3	13.7	15.4	18.6	11.5	11.7	12.0	12.9	17.2
	400	7.7	9.2	9.9	10.3	10.5	10.4	10.6	11.1	12.9	17.6	10.3	10.3	10.3	10.7	15.9
	800	7.4	9.0	9.7	10.0	10.1	10.4	10.4	10.7	12.2	18.1	10.1	10.1	10.1	10.2	15.5
ξ_2	100	8.1	9.2	9.9	10.5	10.7	13.2	13.5	13.9	15.6	18.7	11.3	11.5	11.8	12.7	16.9
	400	7.5	9.0	9.7	10.2	10.3	10.7	11.0	11.4	13.2	18.0	10.2	10.3	10.3	10.7	15.9
	800	7.2	8.9	9.5	9.9	10.0	10.2	10.3	10.5	12.0	17.7	10.3	10.3	10.3	10.4	15.7
ξ_3	100	8.3	9.4	10.2	10.8	11.0	13.3	13.7	14.1	15.8	19.0	11.7	11.9	12.2	13.2	17.5
	400	7.6	9.1	9.8	10.2	10.3	10.4	10.6	11.1	13.1	17.7	10.2	10.2	10.2	10.6	15.9
	800	7.4	9.0	9.6	10.0	10.1	10.1	10.1	10.4	11.9	17.6	10.0	10.0	10.0	10.1	15.5

Note: bootstrap tests are based on a standard fixed-regressor wild bootstrap (b_1) and on the proposed corrected wild bootstrap method (b_2) of Section 4, using $g^* = g - n^{-\kappa}|g|$. ERPs are estimated using 50,000 Monte Carlo replications and 999 bootstrap repetitions. The column "dist." shows the distributions of ε_t: $\xi_1 \sim iidN(0,1)$, $\xi_2 \sim ARCH(1)$ and $\xi_3 = \sqrt{0.5}v_t + \sqrt{0.5}\varepsilon_{x,t}$, where $v_t \sim iidN(0,1)$ and $\varepsilon_{x,t}$ is the error term of the predictive variable $x_{n,t}$.

Table S2: Empirical rejection probabilities (ERPs) of bootstrap tests under local alternatives.

Nominal level: 0.05																
		$a_0=(-3,0)'$					$a_0=(3,0)'$					$a_0=(5,0)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40
ξ_1	100	21.0	21.1	21.3	21.5	21.5	40.6	40.8	40.9	41.0	41.0	68.0	68.0	68.0	68.0	68.0
	400	18.9	19.1	19.3	19.5	19.5	38.5	38.7	38.8	38.8	38.8	64.9	64.9	64.9	64.9	64.9
	800	18.6	18.8	19.0	19.1	19.1	37.6	37.8	37.9	37.9	37.9	64.0	64.0	64.0	64.0	64.0
ξ_2	100	21.7	21.9	22.0	22.2	22.3	41.9	42.1	42.2	42.2	42.3	68.5	68.5	68.5	68.5	68.5
	400	19.2	19.4	19.6	19.7	19.8	38.3	38.6	38.7	38.7	38.7	64.7	64.8	64.8	64.8	64.8
	800	18.6	18.8	19.0	19.1	19.1	37.8	38.0	38.1	38.1	38.1	64.2	64.2	64.2	64.2	64.2
ξ_3	100	20.6	20.7	20.9	21.2	21.3	40.8	41.0	41.1	41.1	41.1	67.3	67.3	67.3	67.3	67.3
	400	19.0	19.1	19.3	19.4	19.4	38.1	38.3	38.4	38.5	38.5	65.0	65.0	65.0	65.0	65.0
	800	18.3	18.5	18.7	18.8	18.9	37.7	38.0	38.0	38.1	38.1	63.5	63.5	63.5	63.5	63.5

Nominal level: 0.10																
		$a_0=(-3,0)'$					$a_0=(3,0)'$					$a_0=(5,0)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40
ξ_1	100	29.6	29.8	30.1	30.5	30.7	54.7	55.0	55.1	55.2	55.2	81.7	81.7	81.7	81.8	81.8
	400	27.0	27.3	27.8	28.1	28.2	52.2	52.5	52.6	52.7	52.7	79.6	79.6	79.6	79.6	79.6
	800	26.4	26.8	27.2	27.5	27.6	51.7	52.1	52.1	52.2	52.2	78.7	78.7	78.7	78.7	78.7
ξ_2	100	30.2	30.4	30.7	31.2	31.4	55.7	55.9	55.9	56.0	56.1	82.0	82.0	82.0	82.0	82.0
	400	27.1	27.4	27.9	28.2	28.3	51.8	52.0	52.1	52.2	52.2	79.3	79.3	79.3	79.3	79.3
	800	26.6	26.9	27.4	27.7	27.7	51.5	51.8	51.9	51.9	51.9	78.6	78.6	78.6	78.6	78.6
ξ_3	100	29.1	29.3	29.6	30.1	30.3	54.2	54.4	54.5	54.6	54.6	80.9	80.9	80.9	80.9	80.9
	400	26.7	27.0	27.4	27.8	27.8	51.7	52.0	52.1	52.2	52.2	79.4	79.4	79.4	79.4	79.4
	800	26.2	26.5	27.0	27.3	27.3	51.3	51.6	51.7	51.7	51.8	78.5	78.5	78.5	78.5	78.5

Note: bootstrap tests are based on a standard fixed-regressor wild bootstrap (b_1) and on the proposed corrected wild bootstrap method (b_2) of Section 4, using $g^* = g - n^{-\kappa}|g|$. ERPs are estimated using 50,000 Monte Carlo replications and 999 bootstrap repetitions. The column "dist." shows the distributions of ε_t: $\xi_1 \sim iidN(0,1)$, $\xi_2 \sim ARCH(1)$ and $\xi_3 = \sqrt{0.5}v_t + \sqrt{0.5}\varepsilon_{x,t}$, where $v_t \sim iidN(0,1)$ and $\varepsilon_{x,t}$ is the error term of the predictive variable $x_{n,t}$.

Table S3: Empirical rejection probabilities (ERPs) of bootstrap tests under local alternatives.

Nominal level: 0.05																
		$a_0 = (-3,1)'$					$a_0 = (2,2)'$					$a_0 = (3,4)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0
ξ_1	100	12.8	12.9	13.0	13.2	13.4	48.4	49.6	50.1	50.3	50.4	73.0	73.9	74.4	74.7	74.7
	400	11.4	11.6	11.9	12.2	12.3	45.4	47.2	47.5	47.6	47.6	70.0	71.6	72.0	72.0	72.0
	800	10.9	11.2	11.6	11.7	11.8	44.8	46.9	47.1	47.1	47.2	69.3	71.1	71.4	71.4	71.4
ξ_2	100	13.1	13.2	13.3	13.5	13.6	49.6	50.8	51.3	51.6	51.6	73.2	74.1	74.7	75.0	75.0
	400	11.4	11.6	11.8	12.1	12.2	46.1	48.0	48.3	48.3	48.3	70.2	71.8	72.2	72.3	72.3
	800	11.0	11.3	11.7	11.9	11.9	45.2	47.2	47.4	47.4	47.4	69.6	71.5	71.7	71.7	71.7
ξ_3	100	12.3	12.4	12.5	12.7	12.9	48.1	49.3	49.9	50.1	50.1	72.4	73.2	73.8	74.1	74.1
	400	11.4	11.6	11.9	12.2	12.3	46.0	47.8	48.2	48.2	48.3	69.9	71.5	72.0	72.0	72.0
	800	11.1	11.4	11.8	12.0	12.1	45.0	46.9	47.1	47.1	47.1	69.4	71.3	71.6	71.6	71.6

Nominal level: 0.10																
		$a_0 = (-3,1)'$					$a_0 = (2,2)'$					$a_0 = (3,4)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0		0.25	0.50	1.0	2.0
ξ_1	100	21.2	21.5	21.6	22.0	22.4	58.8	60.4	61.1	61.5	61.5	80.7	81.6	82.2	82.5	82.5
	400	19.2	19.6	20.2	21.0	21.2	56.0	58.2	58.6	58.7	58.7	78.2	79.9	80.3	80.4	80.4
	800	18.3	18.9	19.7	20.2	20.2	55.8	58.1	58.5	58.5	58.5	77.8	79.8	80.1	80.1	80.2
ξ_2	100	21.8	22.0	22.1	22.5	23.0	59.6	61.1	61.8	62.1	62.2	81.0	81.9	82.5	82.9	82.9
	400	19.1	19.5	20.1	20.7	21.0	56.8	59.0	59.5	59.6	59.6	78.6	80.4	80.8	80.8	80.9
	800	18.9	19.5	20.2	20.7	20.8	56.0	58.4	58.7	58.8	58.8	78.0	79.9	80.2	80.3	80.3
ξ_3	100	20.6	20.8	20.9	21.3	21.8	58.5	60.1	60.8	61.1	61.2	80.2	81.2	81.7	82.0	82.1
	400	19.1	19.5	20.1	20.8	21.0	56.6	58.7	59.2	59.3	59.3	78.3	80.1	80.5	80.6	80.6
	800	18.7	19.2	20.0	20.5	20.6	55.7	58.2	58.5	58.6	58.6	77.8	79.5	79.9	79.9	79.9

Note: bootstrap tests are based on a standard fixed-regressor wild bootstrap (b_1) and on the proposed corrected wild bootstrap method (b_2) of Section 4, using $g^* = g - |g|^{1+\kappa}$. ERPs are estimated using 50,000 Monte Carlo replications and 999 bootstrap repetitions. The column "dist." shows the distributions of ε_t: $\xi_1 \sim iidN(0,1)$, $\xi_2 \sim ARCH(1)$ and $\xi_3 = \sqrt{0.5}v_t + \sqrt{0.5}\varepsilon_{x,t}$, where $v_t \sim iidN(0,1)$ and $\varepsilon_{x,t}$ is the error term of the predictive variable $x_{n,t}$.

Table S4: Empirical rejection probabilities (ERPs) of bootstrap tests under local alternatives.

Nominal level: 0.05

		$a_0 = (-3,1)'$					$a_0 = (2,2)'$					$a_0 = (3,4)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40
ξ_1	100	12.8	13.0	13.2	13.6	13.7	48.4	49.6	50.1	50.4	50.4	73.0	74.0	74.5	74.7	74.7
	400	11.4	11.6	12.0	12.2	12.3	45.4	47.0	47.4	47.6	47.6	70.0	71.4	71.9	72.0	72.0
	800	10.9	11.1	11.5	11.7	11.8	44.8	46.5	47.0	47.1	47.2	69.3	70.8	71.3	71.4	71.4
ξ_2	100	13.1	13.3	13.5	13.9	14.0	49.6	50.8	51.3	51.6	51.6	73.2	74.2	74.7	75.0	75.0
	400	11.4	11.6	11.9	12.1	12.2	46.1	47.7	48.1	48.3	48.3	70.2	71.6	72.1	72.3	72.3
	800	11.0	11.3	11.7	11.8	11.9	45.2	46.9	47.3	47.4	47.4	69.6	71.2	71.6	71.7	71.7
ξ_3	100	12.3	12.4	12.8	13.2	13.3	48.1	49.3	49.9	50.1	50.2	72.4	73.4	73.9	74.1	74.2
	400	11.4	11.7	12.0	12.2	12.3	46.0	47.6	48.0	48.2	48.3	69.9	71.4	71.8	72.0	72.0
	800	11.1	11.4	11.8	12.0	12.1	45.0	46.5	47.0	47.1	47.2	69.4	71.0	71.5	71.6	71.6

Nominal level: 0.10

		$a_0 = (-3,1)'$					$a_0 = (2,2)'$					$a_0 = (3,4)'$				
		b_1	b_2				b_1	b_2				b_1	b_2			
			κ					κ					κ			
dist.	n		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40		0.05	0.10	0.20	0.40
ξ_1	100	21.2	21.5	21.9	22.7	23.0	58.8	60.2	60.9	61.4	61.5	80.7	81.6	82.1	82.4	82.5
	400	19.2	19.6	20.3	21.0	21.2	56.0	57.7	58.3	58.6	58.7	78.2	79.6	80.1	80.4	80.4
	800	18.3	18.8	19.6	20.1	20.2	55.8	57.7	58.2	58.5	58.5	77.8	79.4	79.9	80.1	80.1
ξ_2	100	21.8	22.0	22.5	23.3	23.7	59.6	61.0	61.6	62.1	62.2	81.0	81.9	82.5	82.9	82.9
	400	19.1	19.5	20.1	20.8	21.0	56.8	58.5	59.2	59.5	59.6	78.6	80.0	80.6	80.8	80.8
	800	18.9	19.4	20.1	20.7	20.8	56.0	57.9	58.5	58.7	58.8	78.0	79.5	80.1	80.3	80.3
ξ_3	100	20.6	20.8	21.3	22.2	22.6	58.5	59.9	60.6	61.1	61.2	80.2	81.1	81.7	82.0	82.1
	400	19.1	19.5	20.2	20.8	21.0	56.6	58.3	58.9	59.2	59.3	78.3	79.7	80.3	80.5	80.6
	800	18.7	19.1	19.9	20.5	20.6	55.7	57.7	58.3	58.6	58.6	77.8	79.2	79.7	79.9	79.9

Note: bootstrap tests are based on a standard fixed-regressor wild bootstrap (b_1) and on the proposed corrected wild bootstrap method (b_2) of Section 4, using $g^* = g - n^{-\kappa}|g|$. ERPs are estimated using 50,000 Monte Carlo replications and 999 bootstrap repetitions. The column "dist." shows the distributions of ε_t: $\xi_1 \sim iidN(0,1)$, $\xi_2 \sim ARCH(1)$ and $\xi_3 = \sqrt{0.5}v_t + \sqrt{0.5}\varepsilon_{x,t}$, where $v_t \sim iidN(0,1)$ and $\varepsilon_{x,t}$ is the error term of the predictive variable $x_{n,t}$.

Appendix D

Appendix to Chapter 5

D.1 Results, quarterly VAR

In this section we repeat the analysis carried out above using a sample of quarterly instead of monthly data. This is done for two reasons. First, it should be seen as a robustness exercise, since by using quarterly data we get rid of some noise that can be induced by intra-quarter volatility in the time series. Second, in this way we reproduce the empirical analysis made by **?** in a time-varying parameters context.

Our quarterly time-varying-parameters VAR (TVP-VAR) specification includes a set of six endogenous variables: (1) the unemployment rate; (2) core inflation, measured by the annualized quarterly growth rate of the personal consumption expenditures index; (3) inflation, measured by the annualized quarterly growth rate of the GDP deflator; (4) the log of real per capita GDP; (5) the Fed Funds rate and (6) the Excess Bond Premium (EBP), by **?**. The set of variables mimics the one used by **?**, though we drop some variables they used in their Bayesian VAR to avoid over-fitting and an explosion of the number of parameters to estimate, given the time-varying nature of our model. The quarterly data sample covers the period from 1973Q1 to 2019Q3, for a total of 187 observations. We estimate the VAR with 1 lag, as suggested by the BIC criterion.

Figure D.1 shows the estimated time-varying IRFs to the EBP shock, identified by ordering it last in a Cholesky decomposition. The pattern observed are similar to the results from the monthly sample, though the estimated time variation of IRFs is smoother. The response of unemployment is again strong but less persistent in the first part of the sample, while it is strong and persistent in the last part of the sample (after 2000). A reduction in the magnitude of the unemployment response is observed during the 90's.

The IRFs of inflation is very strong until 1990, very muted afterwards. A similar pattern emerges for the interest rate, again suggesting that the story of a more aggressive response to demand shocks by the Federal Reserve, causing the flattening of the unemployment-inflation relationship, is not supported by the data.

Finally, we implement the **?** technique to estimate the structural parameters of the

NKPC, specified as in equation (4.5.9). Again, as our VAR contains the annualized quarter-on-quarter inflation rate, a proper adjustments of the estimated IRFs had to be made. This time we used horizons $h = 0, 1, 2, \ldots, 17$ to construct the IRF vectors in (4.2.7).

Results are shown in Figure D.2. The resulting evidence is in line with the monthly analysis, and even more clear. The slope of the NKPC reaches value zero already around 1990, and remains zero afterwards. $\gamma_{b,t}$ clearly decreases over time while $\gamma_{f,t}$ increases.

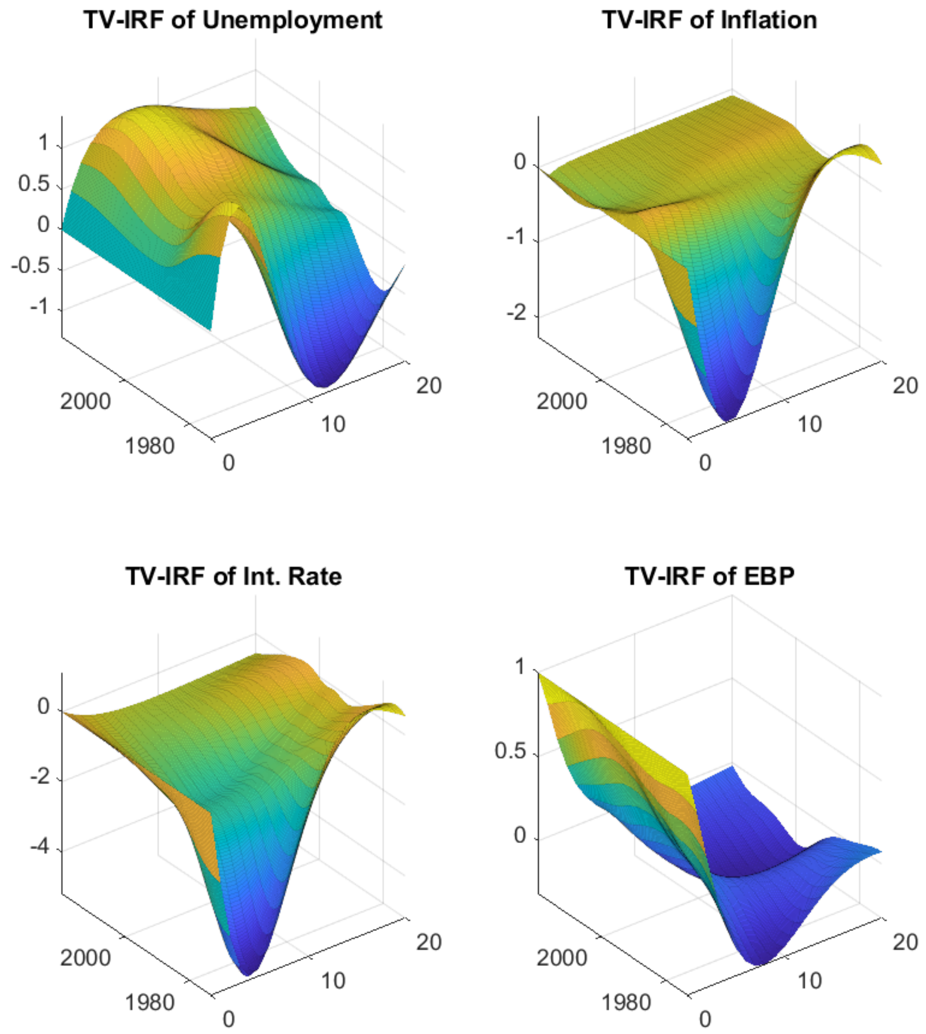


Figure D.1: Time-varying impulse response functions to an EBP shock, for a selected set of variables. VAR at quarterly frequency.

D.2 Sign restrictions

In this section, we revisit our empirical analysis, employing an alternative identification strategy for the demand shock. Departing from the Cholesky ordering of variables, we

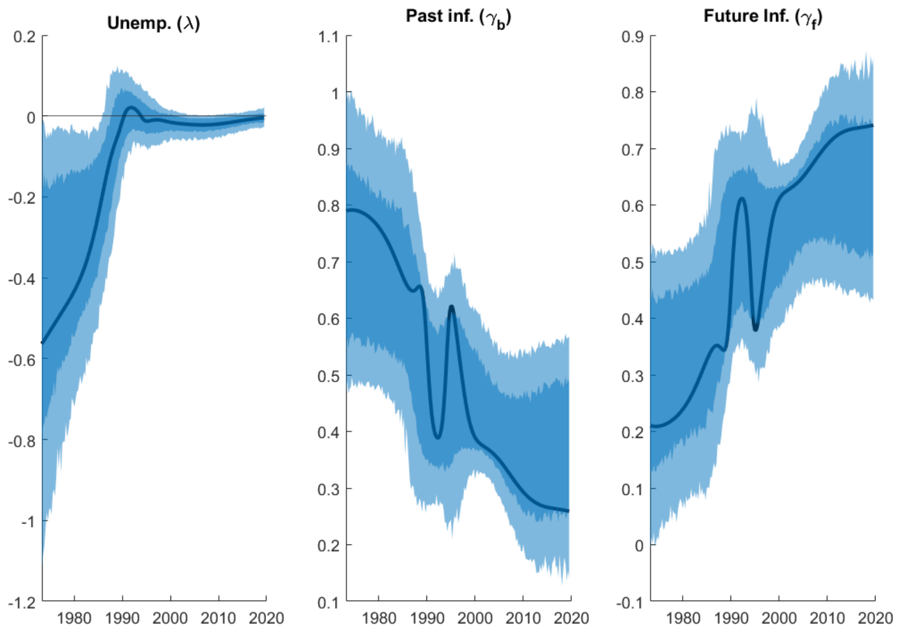


Figure D.2: Time-varying estimates of the NKPC parameters, from the regression in impulse responses estimated on quarterly data. Blue areas show 68% and 90% wild bootstrap confidence intervals.

adopt a sign-restrictions approach.

The motivation behind this choice stems from its dual utility. Not only does it provide insights into the dependence of our results on the identification strategy, but it also facilitates a meaningful juxtaposition with studies utilizing a sign-restriction approach for estimating structural PC coefficients, as elaborated earlier. Notably, the literature employing this methodology often leans towards supporting the hypothesis of no-structural change, in contrast to studies utilizing external instrument identification or Cholesky ordering. Consequently, our analysis aims to elucidate whether such disparities persist even when incorporating our fully time-varying methodology.

Our algorithm works slightly different from the typical sign-restrictions method used in the SVARs literature after the seminal paper of **?**. Notice that for the identification of NKPC parameters we need to identify any demand shock, rather than a specific type of demand shock (e.g. a monetary policy shock). Consequently, our approach consists in exploiting information from all identified shocks that look like a demand type shock.

Our definition of a demand shock is based on a minimal set of sign restrictions, namely that it should induce an opposite contemporaneous response of unemployment and core inflation.

More precisely, consider our estimate for the $N \times N$ time-varying variance-covariance of VAR residuals $\hat{\Sigma}_t$. It is known that there are infinite many matrices B_t such that $B_t B_t' =$

$\hat{\Sigma}_t$, where B_t is the matrix of contemporaneous responses to the structural shocks, i.e:

$$u_t = B_t \varepsilon_t \tag{D.2.1}$$

Our sign-restriction algorithm works as follows. For each $t = 1, 1 \ldots, T$ and $s = 1, \ldots, S$:

1. We randomly generate a candidate matrix $B_{t,s}$ that satisfies the condition $B_{t,s} B'_{t,s} = \hat{\Sigma}_t$. This is done as usual by multiplying the Cholesky decomposition of $\hat{\Sigma}_t$, $\hat{P}_t$, by a randomly generated orthonormal matrix $H_s : H_s H'_s = I$. Hence $B_{t,s} = \hat{P}_t H_s$.
2. For each generated $B_{t,s}$, we check whether any of the identified shocks in $\hat{\varepsilon}_{t,s} = B_{t,s}^{-1} \hat{u}_t$ satisfy our sign-restrictions for being classified as a demand shock. Denote by $N_{d,s,t} \in [0, N]$ the number of shocks satisfying the restriction.
3. If $N_{d,s} = 0$, we go back to step 1 and generate a new matrix. If $N_{d,s} > 0$, we compute impulse responses for all the shocks satisfying the restriction. Denote by $\varepsilon_{j,t,s}$, $j = 1, \ldots, N_{d,s,t}$. Also denote by $\tilde{\Theta}_{Y,t,j,s}$ the $(H+1) \times 1$ vector of impulse responses of inflation to the shock $\varepsilon_{j,t,s}$, hence:

$$\tilde{\Theta}_{Y,t,j,s} = \begin{bmatrix} \widehat{Irf}_{0,j,s}(\pi) \\ \widehat{Irf}_{1,j,s}(\pi) \\ \vdots \\ \widehat{Irf}_{H,j,s}(\pi) \end{bmatrix} \tag{D.2.2}$$

Denote by $\bar{\Theta}_{Y,t,s} = \left[\tilde{\Theta}_{Y,t,1,s} \tilde{\Theta}_{Y,t,2,s} \ldots \tilde{\Theta}_{Y,t,N_{d,s,t},s} \right]$ the $(H+1) \times N_{d,s,t}$ matrix collecting the response of inflation for all demand-like shocks in $\hat{\varepsilon}_{t,s}$. Finally, denote by $\Theta_{Y,t,s} = \text{vec}\left(\bar{\Theta}_{Y,t,s}\right)$ the $(H+1)N_{d,s,t} \times 1$ vector collecting all impulse responses of inflation. Similarly define:

$$\tilde{\Theta}_{X,t,j,s} = \begin{bmatrix} \widehat{Irf}_{0,j,s}(x) & 0 & \widehat{Irf}_{1,j,s}(\pi) \\ \widehat{Irf}_{1,j,s}(x) & \widehat{Irf}_{0,j,s}(\pi) & \widehat{Irf}_{2,j,s}(\pi) \\ \vdots & \vdots & \vdots \\ \widehat{Irf}_{H,j,s}(x) & \widehat{Irf}_{H-1,j,s}(\pi) & \widehat{Irf}_{H+1,j,s}(\pi) \end{bmatrix} \tag{D.2.3}$$

that is the matrix collecting the impulse responses of unemployment and the lagged and forwarded impulse responses of inflation. $\Theta_{X,t,s}$ collects the three responses for all the $N_{d,s,t}$ shocks that satisfy the restrictions:

$$\Theta_{X,t,s} = \left[\tilde{\Theta}'_{X,t,1,s} \, \tilde{\Theta}'_{X,t,2,s} \, \ldots \, \tilde{\Theta}'_{X,t,N_{d,s,t},s} \right]' \tag{D.2.4}$$

4. Hence, our estimate for the NKPC parameters at time t for the sth model is given by:

$$\left[\hat{\lambda}_{t,s} \hat{\gamma}_{b,t,s} \hat{\gamma}_{f,t,s} \right]' = \left(\Theta'_{X,t,s} \Theta_{X,t,s} \right)^{-1} \Theta'_{X,t,s} \Theta_{Y,t,s} \tag{D.2.5}$$

Finally we get the identified set of admissible values for the NKPC parameters at time

t:

$$\mathcal{I}_{\lambda,t} = \left[\hat{\lambda}_{t,1} \dots \hat{\lambda}_{t,s} \dots \hat{\lambda}_{t,S}\right]$$
$$\mathcal{I}_{\gamma_b,t} = [\hat{\gamma}_{b,t,1} \dots \hat{\gamma}_{b,t,s} \dots \hat{\gamma}_{b,t,S}] \tag{D.2.6}$$
$$\mathcal{I}_{\gamma_f,t} = [\hat{\gamma}_{f,t,1} \dots \hat{\gamma}_{f,t,s} \dots \hat{\gamma}_{f,t,S}]$$

Two caveats apply here. First, as we did in the baseline specification with Cholesky ordering, we adjust impulse responses of inflation to the year-on-year and quarter-on-quarter definitions in order to follow the specification of the NKPC used by **?** (see equation (4.5.9)). Also we use only horizons $h = 0, 3, 6, \dots, 33$. Second, we impose the restriction $\gamma_{b,t} + \gamma_{f,t} = 1$ when running the OLS regression in equation (D.2.5).

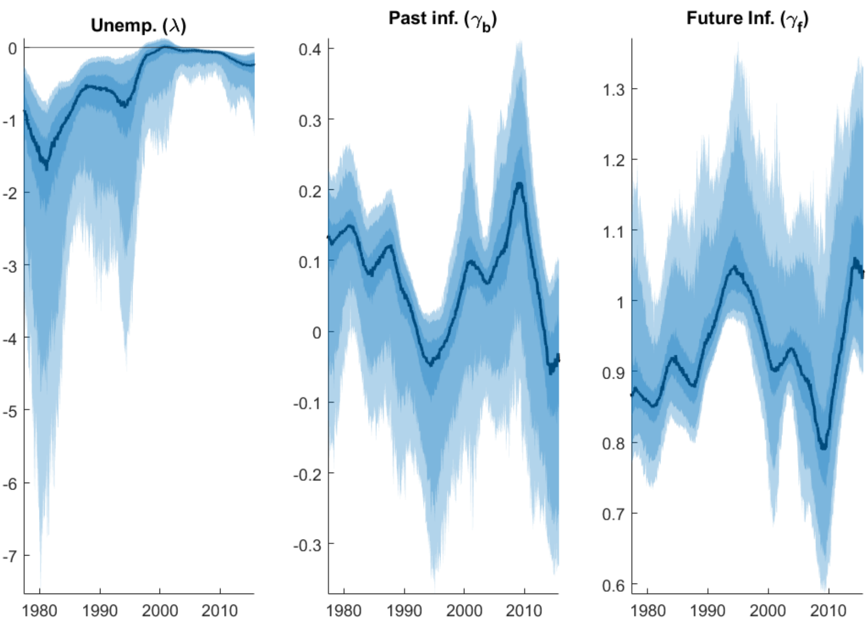


Figure D.3: Sign-restrictions admissible sets for the three KNPC parameters over time. Black lines denote the median values, over time, of the identified sets. Blue areas show the $5 - 10 - 32 - 68 - 90 - 95$ percentiles.

Figure D.3 shows the identified sets for the three KNPC parameters obtained by using the above sign restrictions procedure, setting $S = 1000$. The three panels show the median value of the identified set and the $5\%, 10\%, 32\%, 68\%, 90\%$ and 95% percentiles. Results for the slope λ_t are supportive of the hypothesis of a strong flattening of the structural PC over the years. Differently from the Cholesky estimates, the decline is more gradual and the slope seems to be not different from zero only starting from year 2000 approximately. However, it is noticeable that our approach confirms the structural change hypothesis even by using a very different identification strategy. This may suggests that the literature using sign-restrictions to identify NKPC parameters, failed to identify the structural source of the PC flattening because of the arbitrary sample splitting that is necessary if one does not make use of a fully time-varying specification.

The results for the $\gamma_{b,t}$ and $\gamma_{f,t}$ parameters, however, differ substantially with respect to the baseline Cholesky estimation. The identified set for $\gamma_{b,t}$ suggest a very small value for "persistence" parameter, with the zero almost always included all over the sample, and no big change over the years. This mirrors in the results for $\gamma_{f,t}$ (which recall that it has been restricted to be equal to $1 - \gamma_{b,t}$), which point to a very high and close to 1 value for the "expectation" parameter.

D.3 Results for the Euro Area

We repeat our empirical analysis by applying our methodology to data for the Euro Area, to see whether the results observed for the US extend to other countries. Our baseline VAR specification mirrors the one used for the US. We have a set of seven endogenous variables: (1) the unemployment rate for the Euro Area; (2) core inflation, measured by the annualized monthly growth rate of the seasonally-adjusted HICP index, excluding energy, food and tobacco; (3) the 12-month growth rate of the manufacturing PPI index for the Euro Area ; (4) the log change of total industrial production, excluding construction, in the Euro Area; (5) the 10-year German bond rate; (6) the 3-month interbank rate and (7) the Credit Risk Premium (CRP).

The CRP is the Euro Area version of the EBP used above and it is constructed by **?** as the spread between a measure of cost of financing for a large set of European non-financial corporations and the German bund rate. Compared to EBP, it is a less clean measure of credit frictions. However it is often used for this purpose (see, for example, **?**).

As usual, we face the issue of having to work with a much shorter time-span of observations. We have a total of 252 monthly observations, from 1999M1 to 2019M12.[1]

Figure D.4 shows the estimated values of the three NKPC parameters over time, using the Cholesky identification of the risk-premium shock. In the Euro Area sample, the slope λ_t is never significantly different from zero, suggesting a flat NKPC all over the existence period of the Euro Area. As for the US estimates, we document an increase over time of the expectation coefficient $\gamma_{f,t}$ and a consequent decline over time of $\gamma_{b,t}$.

Figure D.5 shows the admissible set of values for the three NKPC parameters resulting from the sign-restrictions identification strategy applied to the Euro Area sample. Here the results are strikingly different from the ones produced by Cholesky ordering. The slope λ_t is estimated to increase (in absolute value) over time, suggesting a steeper NKPC in the last part of the sample. As for the $\gamma_{b,t}, \gamma_{f,t}$ parameters, the trend resulting from Figure D.4 is confirmed: a decreasing role of past inflation together with an increasing role of inflation expectations.

[1]We exclude observations from the Covid19 recession.

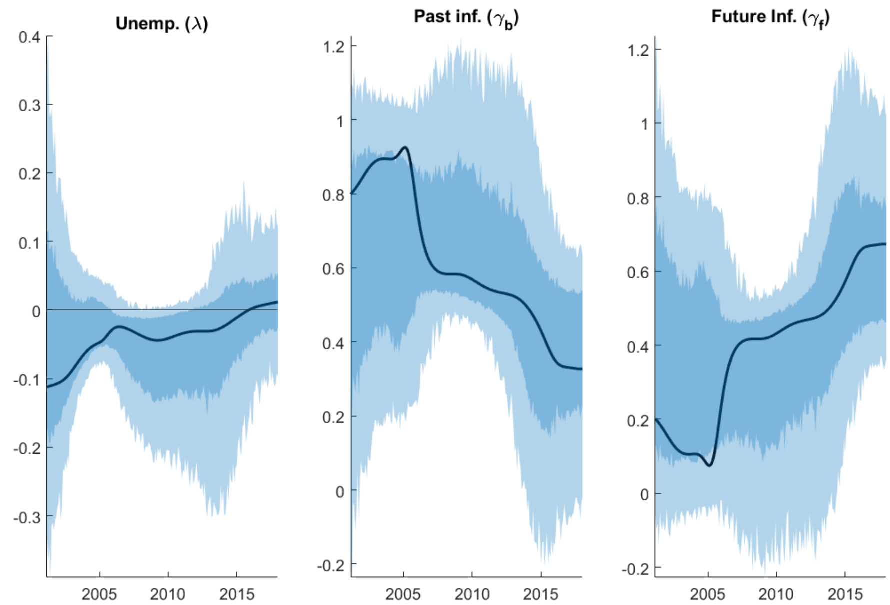


Figure D.4: Time-varying estimates of the NKPC parameters for the Euro Area, from the regression in impulse responses estimated on monthly data. Blue areas show 68% and 90% wild bootstrap confidence intervals.

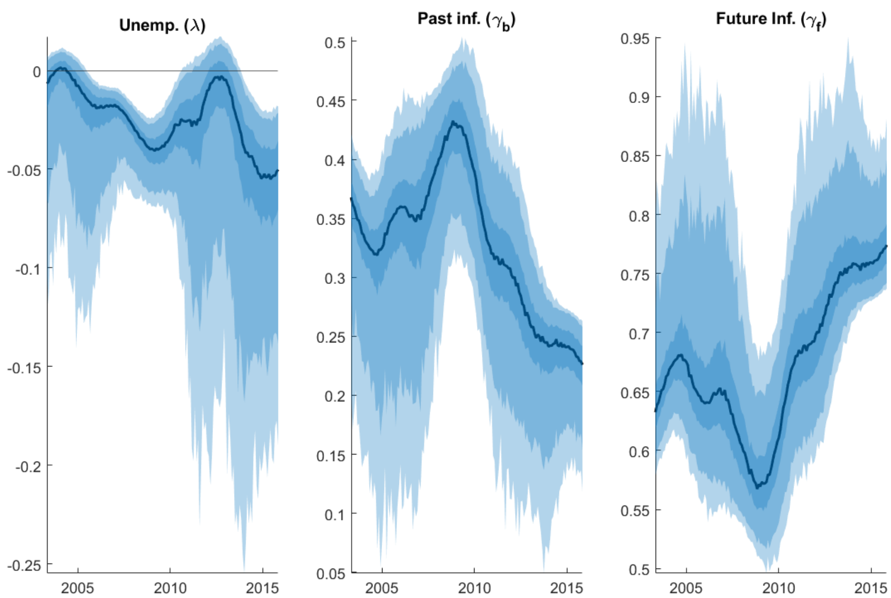


Figure D.5: Sign-restrictions admissible sets for the three KNPC parameters for the Euro Area. Black lines denote the median values, over time, of the identified sets. Blue areas show the $5-10-32-68-90-95\%$ percentiles.

Appendix E

Appendix to Chapter 5

Proof of Theorem 1. We start by considering B_n. Let us define $\sigma_x^2 := \mathbb{E}(x_1^2)$ have that:

$$
\begin{aligned}
B_n &:= \hat{\sigma}_{x,n}^{-1} \sum_{j=1}^{n} k_{tj}(\beta_j - \beta_t)x_j^2 \\
&= \sigma_x^{-1} \frac{1}{\sqrt{H}} \sum_{j=1}^{n} k_{tj}(\beta_j - \beta_t)x_j^2 + o_p(1)
\end{aligned}
$$

where the second equality follows from $\mathbb{E}[H^{-1}\sum_{j=1}^n k_{tj}x_j^2] = \sigma_x^2 H^{-1} \sum_{j=1}^n k_{tj} = \sigma_x^2$ and $\frac{1}{H}\hat{\sigma}_{x,n}^2 = \frac{1}{H}\sum_{i=1}^n k_{tj}\left(x_j^2 - \mathbb{E}(x_j^2)\right) + \mathbb{E}(x_j^2) = \mathbb{E}(x_j^2) + o_p(1)$, where $\frac{1}{H}\sum_{i=1}^n k_{tj}\left(x_j^2 - \mathbb{E}(x_j^2)\right) = o_p(1)$ from Lemma 1. Furthermore, we can note that, by Lemma 2:

$$
\sigma_x^{-1} \frac{1}{\sqrt{H}} \sum_{j=1}^{n} k_{tj}(\beta_j - \beta_t)x_j^2 = \sigma_x \frac{1}{\sqrt{H}} \sum_{j=1}^{n} k_{tj}(\beta_j - \beta_t) + o_p(1) \tag{E.0.1}
$$

We are now deriving the following proposition as an intermediate result.

Proposition 1. Let the conditions of Theorem 1 hold, then:

$$
S_n := \frac{\sigma_x}{\sqrt{H}} \sum_{j=1}^{n} k_{tj}(\beta_j - \beta_t) \xrightarrow{d} N\left(0, \kappa\sigma_\nu^2\sigma_x^2 \int K^2(u)u^2 du\right) \tag{E.0.2}
$$

where $\kappa := \lim_{n\to\infty} H^2/n$.

Proof of Proposition 1. The proof is tri

Lemma 1. Let $\{z_j\}_{j=1}^n$ be a covariance stationary process with autocovariance function $\gamma(k)$ satisfying $\sum_{k=1}^{\lfloor ar_n \rfloor} |\gamma(k)| = o(r_n)$, for every $a > 0$ and $r_n^2 = O(n)$, $\{w_j\}_{j=1}^n$ be a sequence of bounded real numbers such that $w_j = 0$ if $j \notin S_{w,n}$ for some closed interval

$S_{w,n} = [s_{l,n}, s_{u,n}]$ such that $|s_{u,n} - s_{l,n}| = br_n$, for some $b > 0$ then

$$r_n^{-1} \sum_{j=1}^{n} w_j \left[z_j - \mathbb{E}\left[z_j \right] \right] = o_p(1) \tag{E.0.3}$$

Proof of Lemma 1. *We are going to prove this result through L^2-convergence together with Chebychev's inequality. We write $\tilde{z}_j := z_j - \mathbb{E}[z_j]$ and note that:*

$$\begin{aligned}
\mathbb{E}\left[\left(r_n^{-1} \sum_{j=1}^{n} w_j \left[z_j - \mathbb{E}\left[z_j \right] \right] \right)^2 \right] &= r_n^{-2} \sum_{j=1}^{n} \sum_{j'=1}^{n} w_j w_{j'} \mathbb{E}\left[\tilde{z}_j \tilde{z}_{j'} \right] \\
&= r_n^{-2} \sum_{k=0}^{br_n} \gamma(k) \sum_{j'=s_{l,n}}^{s_{u,n}} w_j w_{j+k} \\
&= O\left(\frac{1}{r_n} \right) \sum_{k=0}^{br_n} |\gamma(k)| = O\left(\frac{1}{r_n} \right) o(r_n) = o(1)
\end{aligned}$$

where the penultimate inequality follows from setting $a = b$ and the result proves by Chebychev's inequality.

Lemma 2. Let $\{z_j\}_{j=1}^n$ be a covariance stationary process with autocovariance function $\gamma(k)$ satisfying $\sum_{k=1}^{\lfloor ar_n \rfloor} |\gamma(k)| = o(r_n)$, for every $a > 0$ and $r_n^2 = O(n)$, $\{w_j\}_{j=1}^n$ be a sequence of bounded real numbers such that $w_j = 0$ if $j \notin S_{w,n}$ for some closed interval $S_{w,n} = [s_{l,n}, s_{u,n}]$ such that $|s_{u,n} - s_{l,n}| = br_n$, for some $b > 0$ and $\sum_{j=1}^n w_j = O(r_n)$, and $\{y_j\}_{j=1}^n$ be a process satisfying $\sup_{j,j' \in S_{w,n}} \mathbb{E}[y_j y_{j'}] = O_p(r_n/n)$, then

$$r_n^{-1/2} \sum_{j=1}^{n} w_j \left[z_j - \mathbb{E}\left[z_j \right] \right] y_j = o_p(1) \tag{E.0.4}$$

Proof of Lemma 2. *We are going to prove this result through L^2-convergence together* *with Chebychev's inequality. We write* $\tilde{z}_j := z_j - \mathbb{E}[z_j]$ *and note that:*

$$\begin{aligned}
\mathbb{E}\left[\left(r_n^{-1/2}\sum_{j=1}^{n} w_j\left[z_j - \mathbb{E}\left[z_j\right]\right] y_j\right)^2\right] &= r_n^{-1}\sum_{j=1}^{n}\sum_{j'=1}^{n} w_j w_{j'}\mathbb{E}\left[\tilde{z}_j\tilde{z}_{j'}\right]\mathbb{E}\left[y_j y_{j'}\right] \\
&= O(n^{-1})\sum_{j=1}^{n}\sum_{j'=1}^{n} w_j w_{j'}\mathbb{E}\left[\tilde{z}_j\tilde{z}_{j'}\right] \\
&= O(n^{-1})\sum_{j=s_{l,n}}^{s_{u,n}}\sum_{j'=s_{l,n}}^{s_{u,n}} w_j w_{j'}\gamma(|j-j'|) \\
&= O(n^{-1})\sum_{k=0}^{br_n}\gamma(k)\sum_{j'=s_{l,n}}^{s_{u,n}} w_j w_{j+k} \\
&= O\left(\frac{r_n}{n}\right)\sum_{k=0}^{br_n}|\gamma(k)| = O\left(\frac{r_n}{n}\right)o(r_n) = o(1)
\end{aligned}$$

where the penultimate inequality follows from setting $a = b$ *and the result proves by Chebychev's inequality.*

References

References of Chapters 1-3

Andrews, D.W.K. (1999): Estimation when a parameter is on a boundary, *Econometrica* 67, 1341–1383.

—— (2000): Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space, *Econometrica* 68, 399–405.

Angrist, J.D., & Lavy, V. (1999): Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114, 533–575.

Arcones, M.A., and E. Gine (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 25, 457–481.

Bartalotti, O., Calhoun, G., & He, Y. (2017): Bootstrap confidence intervals for sharp regression discontinuity designs. In *Regression Discontinuity Designs: Theory and Applications*, 421-453. Emerald Publishing Limited.

Bartalotti, O. (2019): Regression discontinuity and heteroskedasticity robust standard errors: evidence from a fixed-bandwidth approximation. *Journal of Econometric Methods*, 8(1), 20160007.

Beran, R. (1987). Prepivoting to reduce level error in confidence sets. *Biometrika* 74, 457–468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–97.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002): Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97(457), 160-169.

Berti, P., L. Pratelli and P. Rigo (2006): Almost sure weak convergence of random probability measures, *Stochastics: an International Journal of Probability and Stochastic Processes* 78, 91–97.

Black, S. (1999): Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114, 577–599.

Calonico, S., M.D. Cattaneo, and M.H. Farrell (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113, 767–779.

Calonico, S., M.D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82, 2295–2326.

Canova, F. & Sala, L. (2009): Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449.

Castelnuovo, E. (2012): Estimating the Evolution of Money's Role in the U.S. Monetary Business Cycle. *Journal of Money, Credit and Banking*, 44(1):23–52.

Cattaneo, M.D., and M. Jansson (2018). Kernel-based semiparametric estimators: small bandwidth asymptotics and bootstrap consistency. *Econometrica* 86, 955–995.

Cattaneo, M.D., and M. Jansson (2022). Average density estimators: efficiency and bootstrap consistency. *Econometric Theory* 38, 1140–1174.

Cattaneo, M.D., M. Jansson, and X. Ma (2019). Two-step estimation and inference with possibly many included covariates. *Review of Economic Studies* 86, 1095–1122.

Cavaliere, G., and I. Georgiev (2020). Inference under random limit bootstrap measures. *Econometrica* 88, 2547–2574.

Cavaliere, G., H.B. Nielsen, R.S. Pedersen and A. Rahbek (2022): Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models, *Journal of Econometrics* 227, 241–263.

Cavaliere, G., H.B. Nielsen and A. Rahbek (2017): On the consistency of bootstrap testing for a parameter on the boundary of the parameter space, *Journal of Time Series Analysis* 38, 513–534.

Cavaliere, G., I. Perera and A. Rahbek (2024): Specification tests for GARCH processes with nuisance parameters on the boundary, *Journal of Business & Economic Statistics* 42, 197–214.

Chatterjee, A., and S.N. Lahiri (2010). Asymptotic properties of the residual bootstrap for lasso estimators. *Proceedings of the American Mathematical Society* 138, 4497–4509.

Chatterjee, A., and S.N. Lahiri (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association* 106, 608–625.

Chay, K., McEwan, P., & Urquiola, M., (2005): The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review* 95, 1237–1258.

Chen, Q. and Z. Fang (2019): Inference on functionals under first order degeneracy, *Journal of Econometrics* 210, 459–481.

Cogley, T., & Sargent, T. J. (2001): Evolving post-world war II US inflation dynamics. *NBER macroeconomics annual*, 16, 331-373.

Coibion, O., Gorodnichenko, Y., and Kamdar, R. (2018): The Formation of Expectations, Inflation, and the Phillips Curve. *Journal of Economic Literature*, 56(4):1447–1491.

Cornea-Madeira, A., and R. Davidson (2015). A parametric bootstrap for heavy-tailed distributions. *Econometric Theory* 31, 449–470.

Del Negro, M., Lenza, M., Primiceri, G. E., and Tambalotti, A. (2020): What's Up with the Phillips Curve? *Brookings Papers on Economic Activity*, pages 301–373.

Del Negro, M. and Primiceri, G. E. (2015). Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum. *The Review of Economic Studies*, 82(4):1342–1345.

Demetrescu, M., I. Georgiev, A.M.R. Taylor and P.M.M. Rodrigues (2023): Extensions to IVX methods of inference for return predictability, *Journal of Econometrics* 237 (Issue 2, Part C).

Dhaene, G., and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82, 991–1030.

Doko Tchatoka, F., and W. Wang (2021): Uniform inference after pretesting for exogeneity with heteroskedastic data, MPRA paper 106408, University Library of Munich, Germany.

Dümbgen, L. (1993): On nondifferentiable functions and the bootstrap, *Probability Theory and Related Fields* 95, 125–140.

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of American Statistical Association* 78, 316–331.

Eubank, R.L., and P.L. Speckman (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association* 88, 1287–1301.

Fan, J., and Gijbels, I. (1996): Local Polynomial Modelling and Its Applications. *Monographs on Statistics and Applied Probability 66.*

Fang Z. (2014): Optimal plug-in estimators of directionally differentiable functionals. Unpublished manuscript.

Fang Z. and A. Santos (2019): Inference on directionally differentiable functions, *Review of Economic Studies* 86, 377–412.

Fu, W., and K. Knight (2000). Asymptotics for lasso-type estimators. *Annals of Statistics* 28, 1356–1378.

Fu, B. (2020): Is the slope of the Phillips curve time-varying? Evidence from unobserved components models. *Economic Modeling*, 88:320–340.

Gali, J. (2008): Monetary Policy, Inflation, and the Business Cycle. *Princeton University Press*.

Gali, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222.

Galì, J. (2011): The return of the wage phillips curve. *Journal of the European Economic Association*, 9(June):436–461.

Galì, J., Smets, F., and Wouters, R. (2011): Unemployment in an Estimated New Keynesian Model. *NBER Macroeconomics Annual*.

Galí, J. and Gambetti, L. (2019): Has the U.S. Wage Phillips Curve Flattened? A Semi- Structural Exploration. Technical Report w25476, *National Bureau of Economic Research*.

Galvão, A. B., Giraitis, L., Kapetanios, G., & Petrova, K. (2016): A time varying DSGE model with financial frictions. *Journal of Empirical Finance*, 38:690–716.

Georgiev, I., D. Harvey, S. Leybourne and A.M.R. Taylor (2019): A bootstrap stationarity test for predictive regression invalidity, *Journal of Business & Economic Statistics* 37, 528–541.

Geyer, C.J. (1994): On the asymptotics of constrained M-estimation, *Annals of Statistics* 22, 1993–2010.

Gilchrist, S. & Mojon, B. (2014): Credit Risk in the Euro area. *Working paper, Banque de France*.

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

Giraitis, L., Kapetanios, G., & Marcellino, M. (2021): Time-varying instrumental variable estimation. *Journal of Econometrics*, 224(2):394–415.

Giraitis, L., Kapetanios, G., & Yates, T. (2014). Inference on stochastic time-varying coefficient models. Journal of Econometrics, 179(1):46–65.

Giraitis, L., Kapetanios, G., & Yates, T. (2018). Inference on Multivariate Heteroscedastic Time Varying Random Coefficient Models. Journal of Time Series Analysis, 39(2):129– 149.

Goncalves, S., and M. Kaffo (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics* 186, 407–426.

Hahn, J., and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.

Hahn, J., and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.

Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics* 14, 1431–1452.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag: Berlin.

Hall, P. (1992):. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, 675-694.

Hall, P. (1993): On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 55(1), 291-304.

Hall, P., and J. Horowitz (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Annals of Statistics* 41, 1892–1921.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001): Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.

Hansen, B.E. (2007). Least squares model averaging. *Econometrica* 75, 1175–1189.

Härdle, W., and A.W. Bowman (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association* 83, 102–110.

Härdle, W., and J.S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics* 19, 778–796.

He, Y., and Bartalotti, O. (2020): Wild bootstrap for fuzzy regression discontinuity designs: obtaining robust bias-corrected confidence intervals. *The Econometrics Journal*, 23(2), 211-231.

Higgins, A., and K. Jochmans (2022). Bootstrap inference for fixed-effect models. Working paper, Cambridge University.

Hirano, K. and Porter, J.R. (2012): Impossibility results for nondifferentiable functionals, *Econometrica* 80, 1769-.1790.

Hjort, N., and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879–899.

Hong, H. and J. Li (2020): The numerical bootstrap, *Annals of Statistics* 48, 397–412.

Horowitz, J.L. (2001). The bootstrap. In *Handbook of Econometrics* (Heckman, J.J., and E. Leamer, eds.), vol. 5, chp. 52, Elsevier: Amsterdam.

Hounyo, U. and K. Lahiri (2023). Estimating the variance of a combined forecast: Bootstrap-based approach. *Journal of Econometrics* 232, 445–468.

Kaffo, M. (2014). Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects. Manuscript.

Kallenberg O. (1997): *Foundations of Modern Probability*, Springer-Verlag: Berlin.

Kallenberg O. (2017): *Random measures: theory and applications*, Springer-Verlag:

Berlin.

Kapetanios, G., & Tzavalis, E. (2010): Modeling structural breaks in economic relationships using large shocks. *Journal of Economic Dynamics and Control*, 34(3), 417-436.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Annals of Statistics* 17, 1168–1175.

Imbens, G. W., & K. Kalyanaraman (2012): Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *Review of Economic Studies*, 79(3), 933-959.

Imbens, G. W., & T. Lemieux (2008): Regression Discontinuity Designs: A Guide to Practice, *Journal of Econometrics*, 142 (2), 615-635.

Inoue, A., Rossi, B., & Wang, Y. (2024): Local projections in unstable environments. *Journal of Econometrics*, 244(2), 105726.

Justiniano, A. & Primiceri, G. E. (2008): The Time-Varying Volatility of Macroeconomic Fluctuations. *American Economic Review*, 98(3):604–641.

Kapetanios, G., Masolo, R. M., Petrova, K., & Waldron, M.(2019). A time-varying parameter structural model of the UK economy. *Journal of Economic Dynamics and Control*, 106:103705.

Kuttner, K. N. (2001): Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

Lee, D. S. (2008): Randomized experiments from non-random selection in US House elections. *Journal of Econometrics*, 142(2), 675-697.

Lewis, D. J. and Mertens, K. (2022): Dynamic Identification Using System Projections and Instrumental Variables. *Working Paper*.

Li, C., and U. Muller (2021). Linear regression with many controls of limited explanatory power. *Quantitative Economics* 12, 405–442.

Liu, C.-A. (2015). Distribution theory of the least squares averaging estimator. *Journal of Econometrics* 186, 142–159.

Li, Q., & J.S. Racine (2007). *Nonparametric Econometrics.* Princeton, NJ: Princeton University Press.

Mavroeidis, S., Plagborg-møller, M., & Stock, J. H. (2014): Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, 52:124–188.

McLeay, M. & Tenreyro, S. (2020): Optimal inflation and the identifiation of the Philips curve. *NBER Macroeconomics Annual*, 34.

Müller, U.K. & M. Watson (2008): Testing models of low frequency variability, *Econometrica* 76, 979–1016.

Oster, E.(2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37, 187–204.

PHILLIPS, P.C.B. (2014): On confidence intervals for autoregressive roots and predictive regression, *Econometrica* 82, 1177–1195.

PLAGBORG-MØLLER, M. AND WOLF, C. K. (2021): Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, 89(2):955–980.

POLITIS, D.N., J.P. ROMANO, AND M. WOLF (1999). *Subsampling.* New York: Springer.

PRIMICERI, G. E. (2005): Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852.

ROMANO, J.P, A.M. SHAIKH AND M. WOLF (2014): A practical two-step method for testing moment inequalities, *Econometrica* 82, 1979–2002.

ROMER, C. D. & ROMER, D. H. (2004): A New Measure of Monetary Shocks: Derivation and Implications. *American Economic Review*, 94(4):1055–1084. ISBN: 00028282.

RUBBO, E. (2023): Networks, Phillips curves, and monetary policy. *Econometrica*, 91(4), 1417-1455.

SHAO, X., AND D.N. POLITIS (2013). Fixed b subsampling and the block bootstrap: Improved confidence sets based on p-value calibration. *Journal of the Royal Statistical Society B* 75, 161–184.

STOCK, J. H., & WATSON, M. W. (2002): Has the business cycle changed and why? *NBER macroeconomics annual*, 17, 159-218.

SWEETING T.J. (1989): On conditional weak convergence, *Journal of Theoretical Probability* 2, 461–474.

UHLIG, H. (2005): What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

WOODFORD, M. (2003): Interest and prices: Foundations of a theory of monetary policy. *Princeton University Press.*

ZANELLI, E. (2023): Bootstrapping Stochastic Time-varying coefficient models. *Mimeo.*