THREE PAPERS ON WEAK IDENTIFICATION ROBUST BOOTSTRAP INFERENCE

Jose Alfonso Campillo Garcia

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2018

Approved by:

Jonathan B. Hill

Andrii Babii

Peter R. Hansen

Ju Hyun Kim

Valentin Verdier

ABSTRACT

JOSE ALFONSO CAMPILLO GARCIA: THREE PAPERS ON WEAK IDENTIFICATION ROBUST BOOTSTRAP INFERENCE.
(Under the direction of Jonathan B. Hill)

This manuscript is composed of three chapters that develop bootstrap methods in models with weakly identified parameters. In the first chapter, joint with Jonathan B. Hill, we introduce an asymptotically valid wild bootstrapped t-test, which provides robust inference for models with unknown identification category. Under weak identification, the wild bootstrap needs to be constructed using residuals imposing lack of identification, because the usual regression residuals are non-consistent. The robust t-test has better small sample properties compared to the asymptotic approximation counterpart and is simpler to estimate in practice, especially when the underlying Gaussian process has an unknown form and/or is high dimensional. A simulation exercise shows the benefits of using a robust t-test exposing the large-size distortions of the standard t-test when weak identification is present.

In the second chapter, joint with Jonathan B. Hill, we introduce a parametric bootstrap that provides an alternative approach to construct statistical tests when parameters are weakly identified. The method extends the parametric bootstrap in regression models, to cases where some of the parameters cannot be consistently estimated, reducing the number of nuisance parameters that arise in the asymptotic distribution of the test statistic under weak identification. Unlike the known statistical tests in the literature, this parametric bootstrap method can mimic the true distribution without nuisance parameters in some important cases. We establish robust critical values of the t-statistic that lead to correct asymptotic size when the identification category is unknown. The simulation exercise shows that the parametric bootstrap can lead to very accurate test sizes and considerable test power comparable to the (infeasible) test statistic that assumes nuisance parameters are known.

In the final chapter, we consider the mixed data sampling (MIDAS) model proposed by Ghysels, Santa-Clara, and Valkanov (2005) to evaluate the empirical performance of the wild bootstrapped robust t-test of Chapter 1 and the parametric bootstrapped robust t-test of Chapter 2. To test the statistical significance of the MIDAS estimators, we derive the bootstrapped t-test assuming weak identification because the parameters of the MIDAS model cannot be separately identified under the null hypothesis. Contrary to the results by Ghysels et al. (2005), the bootstrapped t-tests suggest that the estimators of the MIDAS model are not statistically significant, implying that the proposed functional form has low explanatory and predictive power in the study of the risk-return trade-off. We extend the empirical results to different sample frequencies to evaluate the small sample performance of the bootstrap methods and propose an alternative MIDAS specification constructed with the absolute value of excess returns.

ACKNOWLEDGMENTS

There is a group of important people whose contribution was essential for the development of this manuscript. First and foremost, I would like to express my gratitude to my advisor, Prof. Jonathan B. Hill whose guidance and knowledge helped me strengthen my research and my enthusiasm for econometrics. His work ethic has inspired me to put my best effort every day, while his motivation and wisdom aided tremendously to advance in the improvement of this work. I would also like to thank Prof. Ju Hyun Kim for being so dedicated and kind. He always had detailed lists of comments and suggestions after every presentation in the UNC Econometrics Workshop, and was always available when I had any question or issue. Thanks to my committee members: Andrii Babii, Peter R. Hansen, and Valentin Verdier, for their valuable feedback and insights.

I would also like to thank all the professors and students that attended the econometrics workshop throughout my doctorate degree. I am particularly thankful to Steven Raymond and Jay Dennis whose motivation and supportive feedback lead me to improve enormously year after year.

The people in my personal life have been just as important to complete my Ph.D. degree. I would like to thank my parents, Jose Ignacio and Maria del Consuelo, for their unconditional love, and persistent support. This would have never been possible if they had not done an extraordinary job teaching me the importance of education and hard work. Thanks to my siblings, Juan Pablo and Flor Belen, who have made me smile when I need it the most, and who I can always count on even if we are thousands of miles apart. Last but not least, I would also like to thank Katy Ascanio for her support during the writing of this manuscript. Whenever I felt exhausted, she was there for me to offer encouragement to go the extra mile. I will always be grateful for her patience and support, which has not only helped me in my academic life, it has also made me a better person.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

A WILD BOOTSTRAPPED T-TEST ROBUST TO ALL IDENTIFICATION CATEGORIES

1.1 Introduction

Conducting inference in econometric models is essential to evaluate the policy implications of the results obtained from model estimation. One of the most important assumptions needed to obtain valid inference is parameter identification of the particular model being studied. When parameters are weakly identified, the strong identification asymptotic results are are not valid in general because the parameters cannot be consistently estimated. A robust test is helpful in situations where strong identification cannot be assumed because it provides valid inference regardless of the identification case. Unfortunately, hypothesis tests robust to cases where parameters are weakly identified are scarce in the econometrics literature. The objective of this paper is to construct a t-test that allows us to perform hypothesis testing even when the parameter of interest cannot be uniquely recovered with the existing data, and provide valid inference for all identification cases.

Andrews and Cheng (2012, 2014, 2013) introduced a unified treatment to derive the asymptotic distributions of parameters with different identification category: that is, parameters that can be weak, semi-strong, and strongly identified according to their value along the parameter space. In a nonlinear least squares regression model, for example, the framework of Andrews and Cheng (2012) provides the theoretical results to perform inference on models in which weak identification leads to least squares estimators with a non-invertible Hessian matrix for some values in the parameter space. Because the parameters that are possibly weakly identified are known in a parametric model, we perform quadratic expansions around the points of lack of identification and obtain stochastic processes dependent on the identification category thereby, bridging the gap between weak and strong identification. Following the definitions and framework of Andrews and Cheng (2012), we propose an asymptotically valid, wild bootstrap t-test that encompasses the limit distribution of each identification category. Specifically, we propose a t-test constructing a t-statistic that; 1) is a wild bootstrap based on the multiplier bootstrap structure of Wu (1986) and Liu et al. (1988) allowing for heterogeneously distributed data; 2) is robust to all identification categories because we derive two asymptotically valid bootstrap distributions, one for the nonstandard asymptotic distribution under non and weak identification, and another for the strong and semi-strong identification categories; and 3) has better small sample properties compared to the asymptotic approximation counterpart, and is simpler to estimate in practice; especially when the underlying Gaussian process has an unknown form and/or is high dimensional.

In empirical applications, the identification category is unknown. We expect a loss in size and power in robust statistics that do not assume the identification category of the parameters. To compensate for this loss, we chose to construct a wild bootstrap process, over asymptotic limit approximations, for the following reasons: 1) the bootstrap is more accurate because an asymptotic refinement (faster higher order convergence rates in the Edgeworth expansion) can be achieved in small samples. In particular, the bootstrapped pivotal statistics like the t-statistic achieve asymptotic refinements which usually outperform the asymptotic limit approximations (see Mammen (1993), MacKinnon (2002), see also Horowitz (2001), Härdle, Horowitz, and Kreiss (2003). 2) As the asymptotic distribution of the t-statistic under weak identification is non-standard, the bootstrap is simpler to simulate when we do not have closed-form expressions of the stochastic process, especially when the process is high dimensional. 3) The wild bootstrap has proven to be particularly useful for cases with heteroscedastic data, non-symmetric statistic distributions, and when the parameter dimension is large. As under weak identification the asymptotic distributions are asymmetrical, in the simulation study, we show that an asymmetric two-point distribution distributions perform better than the (symmetric) limit approximation counterpart. To the best of our knowledge, there is no study that develops a bootstrap method to perform parameter inference robust to all identification categories.

In addition to Andrews and Cheng (2012) and the subsequent extensions to Generalized Method of Moments and Maximum Likelihood (Andrews and Cheng 2014, 2013), the results in this study are based on arguments from the bootstrap literature which was introduced by Efron et al. (1979). As the classic resampling bootstrap method of Efron relies on the strong assumption of i.i.d. observations, in this paper, we construct the t-statistic using a wild bootstrap. Following the suggestions of Wu (1986) and Beran (1986), Liu et al. (1988) introduced the wild bootstrap as an alternative bootstrap with heterogeneously distributed data. In regression models, the wild bootstrap is a convenient method to perform inference in models with unknown heteroscedasticity as it avoids the estimation of the variance-covariance matrix of the residuals. The Heteroscedastic Consistent Covariance Matrix Estimator (HC-CME) introduced by White (1980) is perhaps the most popular estimator of the variance of residuals in models with heteroscedasticity. However, it has been shown that the HC-CME estimator suffers from large bias, particularly in small samples and when outliers are present. MacKinnon and White (1985) demonstrate that the bias of the t-statistic can be very large in regression models with heteroscedastic observations. The wild bootstrap has been demonstrated itself to be an advantageous option to reduce the small sample bias. Mammen (1993), Horowitz (2001) show via Edgeworth expansions that asymptotic refinements, of size $n^{-1/2}$, can be achieved using a wild bootstrap, while the simulation exercise of Davidson and Flachaire (2008) demonstrates that the wild bootstrap outperforms asymptotic approximations and the resampling bootstrap of Efron et al. (1979). Even though the wild bootstrap results developed by Liu et al. (1988) assume independence throughout the paper, Shao (2010) extends the wild bootstrap to models with weakly dependent data.

Performing a bootstrap procedure under weak identification is not an immediate extension of the theoretical results of Efron et al. (1979), Liu et al. (1988), etc. The assumptions ensuring uniform convergence of the bootstrapped distribution, which are required for the asymptotic validity of the bootstrap, fail precisely for the parameter values that lead to weak identification (see Efron et al. (1979), Giné and Zinn (1990), Andrews and Guggenberger

(2010), Mammen (1993), among others). For example, a resampling residual bootstrap for a non-linear regression model under weak identification leads to invalid inference because the residuals cannot be resampled when some of the parameters are not consistent. The same problem would apply to the Integrated Conditional Moment test of Bierens and Ploberger (1997), which is comparable to our t-test in the sense that it has a nonstandard distribution and depends on nuisance parameters under the null hypothesis. On the other hand, models with parameters that allow different identification category have an asymptotic distribution that varies within the parameter space. To overcome these issues, the wild bootstrap is constructed using residuals centered at the point of lack of identification, which does not depend on the weakly identified parameters and therefore, relies solely on the available consistent estimators.

The t-test developed in this paper is subject to similar challenges from previous studies which analyze hypothesis testing on statistics that depend on nuisance parameters under the null hypothesis. Under weak identification, the nuisance parameters are precisely those that cannot be estimated consistently but appear in the asymptotic distribution of the test statistic. Testing the null hypothesis $H_0 : \pi = 0$, where π is weakly identified, leads to a t-test whose asymptotic distribution depends on the its true (unknown) parameter π_0 (see Section 1.2 and 1.3). The challenges of hypothesis testing with nuisance parameters under the null or alternative dates back to Chernoff and Zacks (1964) for a sup-Lagrange multiplier and Davies (1977, 1987) for a sup-Likelihood ratio test. More recently, Hansen (1996) studied the effect of testing with nuisance parameters in the null hypothesis, introducing a transformation that eliminates the dependence on the nuisance parameter. Stinchcombe and White (1998), Andrews and Ploberger (1994) develop smoothed test statistics when there is a nuisance parameter under the alternative hypothesis. Hill (2017) developed a conditional mean test with nuisance parameters under the alternative that is consistent against general alternatives in the sense of Bierens (1990), Bierens and Ploberger (1997) and Hill (2008), Hill (2013). To obtain a robust t-statistic, Hill (2017) constructed a test that exploits the p-value occupation time, which does not depend on a nuisance parameter as it integrates over its support.

A smaller number of research papers have analyzed the consequences of hypothesis testing under weak identification in the parameter space. Studies by Antoine and Renault (2009, 2012) and Caner (2009) provide asymptotic results of GMM models with near weak instruments. In particular, Caner (2009) shows parameter inference is possible as the Wald, LR, and LM tests have a standard χ^2 limit distribution when nearly-weak instruments are present. Nelson and Startz (2007) and Ma and Nelson (2008) analyze models in which the asymptotic variance of one parameter depends on the identification of a different parameter. However, none of these studies extend their results to construct tests that are robust for strong and weak identification nor do they employ a bootstrap method.

The weak identification literature is vast and encompasses very different fields and methodologies. A partial list of recent studies include Nelson and Startz (1988), Stock and Wright (2000), Dufour and Taamouti (2005), Staiger and Stock (1994), Kleibergen (2002, 2005), etc. Bootstrapping methods with weak instruments have received less attention; some examples include Davidson and MacKinnon (2014), Moreira, Porter, and Suarez (2009). In contrast to the weak instrument literature, wherein the source of weak identification is treated as exogenous, in the framework of Andrews and Cheng (2012), the source of lack of identification is known to be caused by specific values of the parameter space. For this reason, the approach of Andrews and Cheng (2012) is complementary to the weak instrument literature. The ability to pinpoint the weak identification cases is key to develop a valid bootstrap method.

Robust inference is an important issue in economic applications. Ignoring the consequences of weak identification in hypothesis testing can lead to severely biased statistics. We claim using a robust t-test leads to more accurate asymptotic sizes when weak identification is present, with a minor loss in the power of the test. Examples of parametric models that can encounter weak identification for certain values of the parameter space include; ARMA models (Andrews and Cheng 2012), Maximum Likelihood Estimation (Andrews and Cheng 2013), Generalized Method of Moments (Andrews and Cheng 2014), Dynamic General Equilibrium models (Guerron-Quintana, Inoue, and Kilian 2013; Andrews and Mikusheva

2015), MIDAS regression (Ghysels, Hill, and Motegi 2016), Smooth Transition Autoregressive models (Andrews and Cheng 2013), Probit models (Andrews and Cheng 2014), Regime Switching models (Chen, Fan, and Liu 2016), among others. The t-test proposed in this study allows to perform robust inference against weak identification for all the previously enlisted models.

This paper has the following structure. In Section 1.2, we introduce the framework used throughout the paper and define the identification categories: strong, semi-strong, weak, and non-identification. Section 1.3 shows consistency and derives the limit distribution of the parameters and t-statistics under all identification categories. We introduce the wild bootstrap method and develop the bootstrapped distribution of the parameters and the bootstrapped distribution of the t-statistic in Section 1.4. Simulation results are presented in Section 1.5, while Section 1.6 concludes the paper. All proofs are provided in the Appendix. Let $X_n(\pi) = o_{p,\pi}(1)$ be defined as $\sup_{\pi\in\Pi} ||X_n(\pi)|| = o_p(1)$, where $||\cdot||$ denotes the Euclidean norm. Let $\underline{eig}(A)$ and $\overline{eig}(A)$ denotes the smallest and largest eigenvalues of matrix A. The symbol $\Rightarrow$ denotes weak convergence of a stochastic process indexed by $\pi \in \Pi$ for parameter space Π.

1.2 Identification Categories

In this section, we introduce the framework used throughout the paper and define the identification categories: strong, semi-strong, weak and non-identification. We begin by introducing some examples of data generating processes in which identification depends on the value the parameter takes within the parameter space.

1.2.1 Examples

Example 1. Nonlinear Regression Model

$$y_t = \zeta_0' X_{1,t} + \beta_0' h(X_{2,t}, \pi_0) + \epsilon_t \tag{1.1}$$

The identification of π_0 depends on β_0 being non-zero. If $\beta_0 = 0$, the parameter π_0

is weakly identified. The identification of ζ_0 does not depend on the identification of either β_0 or π_0; it is always strongly identified. Usual examples of the function $h(X_t, \pi_0)$ include a) Exponential $h(X_t, \pi_0) = \exp(-\pi_{0,1}(X_t - \pi_{0,2}))$, and b) Logistic $h(X_t, \pi_0) = 1/(1 + \exp(-\pi_{0,1}(X_t - \pi_{0,2})))$.

Example 2. ARMA(1,1) model

$$y_t = (\pi_0 + \beta_0) y_{t-1} + \epsilon_t - \pi_0 \epsilon_{t-1} \tag{1.2}$$

The parameters β_0 and π_0 are not identified when they have equal numerical value. When $\beta_0 = \pi_0$, the model is observationally equivalent to the model $y_t = \epsilon_t$. Ansley and Newbold (1980) and Nelson and Startz (2007) demonstrate that when the time series y_t is serially uncorrelated, the estimators suffer from substantial bias and large variance and hypothesis testing may suffer from size distortions.

Example 3. MIDAS Regression

Consider a mixed data sampling process $\{y_t, X_{t/m}^{(m)}\}$, where y_t is observable at times $t = 1, ..., n$ and $X_{t/m}^{(m)} = (1, X_{1,t/m}^{(m)}, ..., X_{p,t/m}^{(m)})$, where m is the number of high frequency lags used in the temporal aggregation of $X_{t/m}$. The MIDAS regression sets the higher frequency variable on the right-hand side of a regression equation

$$y_t = \beta_0' X_t(\pi_0) + \epsilon_t \tag{1.3}$$

where $X_t(\pi_0)$ is a nonlinear function that maps the high frequency data into the lower frequency data,

$$X_{k,t}^{(m)}(\pi_{0,k}) = \sum_{j=1}^{m} w_{j,k}(\pi_{0,k}) L^{j/m} X_{k,t/m}^{(m)} \tag{1.4}$$

where L denotes the lag operator. If $\beta_0 = 0$, then π_0 is not identified, and the weight function

$w_{j,k}(\pi_{0,k})$ can potentially take any value.

1.2.2 Definitions and model setup

Before we state the assumptions, we introduce several concepts and notation. Consider a sampling process $W_t = \{y_t, X_t\}$ of observable random variables $X_t \in \mathbb{R}^d, y_t \in \mathbb{R}$, and $\theta = (\zeta', \beta', \pi')'$ denotes the vector of parameters. Throughout this study, we consider a generalized non-linear model:

$$\epsilon_t(\theta) = y_t - \zeta' X_{1,t} - \beta' h(X_{2,t}, \pi) \tag{1.5}$$

where $X_{1,t}$ and $X_{2,t}$ denote elements of the X_t matrix and h is a non-random function. The parameters are defined as elements contained in compact sets of $\mathbb{R}$, $\zeta \in \mathcal{Z} \subset \mathbb{R}^{d_\zeta}$, $\beta \in \mathcal{B} \subset \mathbb{R}^{d_\beta}$ and $\pi \in \Pi \subset \mathbb{R}^{d_\pi}$.

The estimator $\hat{\theta}_n$ minimizes the objective function $Q_n(\theta)$. For simplicity, we use the least squares objective function.

$$\hat{\theta}_n = \inf_{\theta \in \Theta} Q_n(W_t, \theta) = \inf_{\theta \in \Theta} \frac{1}{2n} \sum_{t=1}^{n} \epsilon_t^2(\theta) \tag{1.6}$$

Under weak identification, the limit objective function of the non-linear model does not depend on the weakly identified parameter, i.e., the objective function $Q_n(\theta)$ becomes flatter with respect to π as n grows to infinity and β goes to zero. As a consequence, the second-order derivative of $Q_n(\theta)$ is singular or near singular for some values of the parameter space. In these cases, the delta method cannot be applied because the uniformity conditions required for the bootstrap are not valid. To develop the wild bootstrap method, we derive first-order expansions around the point of lack of identification, which is defined using drifting sequences of true parameters.

We assume drifting sequences of true parameters to define the degree of identification and true process $\epsilon_t(\theta_n)$ for each $n \geq 1$. The drifting sequence of parameters serves as a useful theoretical tool to describe the range of behavior of the asymptotic distribution of parameters under the different identification categories. Suppose the sequences of true parameters are defined by $\theta_n = (\beta'_n, \zeta_n,' \pi'_n)'$ for $n \geq 1$, converging to the limit true parameters defined by $\theta_0 = (\beta'_0, \zeta'_0, \pi'_0)'$ [1]. Table 1.1 illustrates the definitions of identification categories introduced by Andrews and Cheng (2012, 2014, 2013).

Table 1.1: Identification categories

Category	$\{\beta_n\}$ sequence	Identification Property of π
I(a)	$\beta_n = 0 \; \forall n \geq 1$	Unidentified
I(b)	$\beta_n \neq 0$ and $\sqrt{n}\beta_n \to b \in \mathbb{R}^{d_\beta}$	Weakly identified
II	$\beta_n \to 0$ and $\sqrt{\beta_n} \to \infty$	Semi-strongly identified
III	$\beta_n \to \beta_0 \neq 0$	Strongly Identified

For notational convenience, we partition the parameter space according to each parameter's identification category.

$$\theta = (\zeta', \beta', \pi')' = (\psi', \pi')' \tag{1.7}$$

The parameter ψ denotes the strongly identified parameters, which can be estimated consistently, whereas π denotes parameters that are weakly identified. The identification of π depends on the parameter β, whereas ζ denotes all other parameters that do not affect the identification of π.

The speed at which the parameter β_n converges to zero determines the identification category of π and therefore if a consistent estimator of π is attainable. If $n^\alpha ||\beta_n|| = O(1)$ for some $\alpha \in [0, 1/2)$, a consistent estimator of π is feasible because the sequence is converging to zero at a slower rate than $\sqrt{n}$. Subsequently, if the speed is larger or equal to 1/2, that is $n^\alpha ||\beta_n|| = O(1)$ for some $\alpha \geq 1/2$, no consistent estimate of π_0 is available because the elements of the first order expansion with respect to π and the noise process ϵ_t have the same order of magnitude. The non-linear model studied in this paper defines the true error

[1] For more details on drifting sequences of distributions see Staiger and Stock (1994) and Stock and Wright (2000).

process for each sequence of drifting true parameters n, which we denote by $\epsilon_t(\theta_n)$, with $\theta_n = (\zeta_n', \beta_n', \pi_n')'$

$$\epsilon_t(\theta_n) = y_t - \zeta_n' X_{1,t} - \beta_n' h(X_{2,t}, \pi_n) \tag{1.8}$$

Accordingly, the limit error process is denoted by $\epsilon_t(\theta_0) = y_t - \zeta_0' X_{1,t} - \beta_0' h(X_{2,t}, \pi_0)$, where $\theta_0 = (\zeta_0', \beta_0', \pi_0')'$, while the finite sample and limit variance of errors are defined by $\sigma_t^2(\theta_n)$ and $\sigma_t^2(\theta_0)$ respectively.

Semi-strong identification bridges the gap between weak identification and strong identification. In Section 1.3 we show π cannot be estimated consistently when $\beta_n \to 0$ and $\sqrt{n}\beta_n \to b$ (i.e., weak identification). Nonetheless, we can obtain an expression of the asymptotic distribution of π that depends on functionals of an empirical process. Moreover, the weak identification of π has consequences on the parameter β, which although it can be consistently estimated, has a non-standard asymptotic distribution because β depends on the random draw of the distribution of π. On the other hand, under semi-strong identification, β and π can be consistently estimated, and inference is standard under the proper normalization which avoids the singularity of the second order term in the Taylor expansion.

We define the degree of identification into three non-exclusive cases of sequences of true parameters:

$$\Theta(\theta_0) = \{\{\theta_n \in \Theta : n \geq 1\} : \theta_n \to \theta_0 \in \Theta\} \tag{1.9}$$

$$\Theta(\theta_0, 0, b) = \{\{\theta_n\} \in \Theta(\theta_0) : \beta_0 = 0 \text{ and } \sqrt{n}\beta_n \to b \in \mathbb{R}\} \tag{1.10}$$

$$\Theta(\theta_0, \infty, \omega_0) = \{\{\theta_n\} \in \Theta(\theta_0) : \sqrt{n}||\beta_n|| \to \infty \text{ and } \beta_n/||\beta_n|| \to \omega_0 \in \mathbb{R}^{d_\beta}\} \tag{1.11}$$

Unless we explicitly differentiate within identification categories, we write "under $\Theta(\theta_0, 0, b)$" to refer to cases under weak identification, while we write "under $\Theta(\theta_0, \infty, \omega_0)$" when referring to strongly identified parameters. The asymptotic distribution of the estimators and the t-statistics turn out to be standard under $\Theta(\theta_0, \infty, \omega_0)$ and non-standard under $\Theta(\theta_0, 0, b)$.

1.3 Asymptotic results under weak and strong identification

In this section, we show consistency and convergence in distribution for each identification category. In particular, we show that under weak identification, $\hat{\pi}_n$ converges to a random variable which leads to non-standard asymptotics of $\hat{\psi}_n$.

1.3.1 Asymptotic distributions of the estimators

Let $\hat{\theta}_n$ be the estimator that minimizes the objective function $Q_n(\theta)$ over the parameter space Θ. We assume that the true parameters β_n lie in the interior of the parameter space Θ as boundary effects are not the main focus of this paper (see Andrews (1999, 2001)). In the weak identification case, it is necessary to obtain the asymptotic results of ψ and π separately as the distribution of ψ depends on the (random) draw of π. We define the extremum estimator $\hat{\psi}_n(\pi)$, which characterizes the minimizer of the objective function for each $\pi \in \Pi$ as:

$$Q_n(\hat{\psi}_n(\pi), \pi) = \inf_{\psi \in \Psi(\pi)} Q_n(\psi, \pi) + o(n^{-1}) \tag{1.12}$$

To estimate the possibly weakly identified parameter π, we let $Q_n^c(\pi)$ denote the concentrated sample objective function $Q_n(\hat{\psi}_n(\pi), \pi)$. The parameter $\hat{\pi}_n$ is defined as the extremum estimator that minimizes the concentrated estimator sample objective function.

$$Q_n^c(\hat{\pi}_n) = \inf_{\pi \in \Pi} Q_n^c(\pi) + o(n^{-1}) \tag{1.13}$$

We derive the second-order expansion of the objective function uniformly on π, meaning the quadratic expansion is made around ψ for each π. Define $d_\psi(\theta) = \frac{\partial}{\partial \psi}\epsilon_t(\theta)$ as the derivative of ϵ_t with respect to the strongly identified parameters. Using least squares estimation, the gradient of the objective function with respect to ψ uniformly on π takes the following form:

$$\frac{\partial}{\partial \psi} Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t(\theta) d_{\psi,t}(\pi) \tag{1.14}$$

Under semi-strong identification we consider the gradient of the objective function with respect to all parameters because, in this case, a normalization matrix is be enough to obtain convergence in distribution. The gradient is denoted by, $\frac{\partial}{\partial\theta}Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\theta)d_{\theta,t}(\theta)$, with $d_\theta(\theta) = \frac{\partial}{\partial\theta}\epsilon_t(\theta)$. Without loss of generality, we express the parameter space dependent on π to work with the concentrated estimator.

$$\Theta = \{\theta = (\psi', \pi')' : \psi \in \Psi(\pi), \pi \in \Pi\}, \text{where} \tag{1.15}$$

$$\Pi = \{\pi : (\psi', \pi')' \in \Theta \quad \text{for some} \quad \psi\} \tag{1.16}$$

$$\Psi(\pi) = \{\psi : (\psi', \pi')' \in \Theta\} \text{ for } \pi \in \Pi \tag{1.17}$$

The following stochastic processes are important, as they define the asymptotic distribution of ψ and π.

$$H_n(\psi_{0,n}, \pi) = \frac{\partial^2}{\partial\psi\partial\psi'}Q_n(\psi_{0,n}, \pi) \tag{1.18}$$

$$G_{\psi,n}(\pi) = \sqrt{n}[\frac{\partial}{\partial\psi}Q_n(\psi_{0,n}, \pi) - \mathbb{E}_{\theta_n}(\frac{\partial}{\partial\psi}Q_n(\psi_{0,n}, \pi))] \tag{1.19}$$

$$K_n(\psi_{0,n}, \pi; \theta_n) = \frac{\partial}{\partial\beta_n}\mathbb{E}_{\theta_n}(\frac{\partial}{\partial\psi}Q_n(\psi_{0,n}, \pi)) \tag{1.20}$$

where $\frac{\partial}{\partial\beta_n}$ denotes a partial derivative with respect to the true parameter β_n. In Lemma A.2.3 we show that the limit distribution of $G_{\psi,n}(\pi)$ converges to a zero mean Gaussian process $G_\psi(\pi)$ with bounded continuous sample paths and covariance kernel $\Omega(\pi, \tilde{\pi}; \theta_0)$ for $\pi, \tilde{\pi} \in \Pi$. The process $H_n(\theta)$ is standard and converges uniformly to a probability limit. Conversely, the process $K_n(\pi)$ is a first-order bias that arises when we center the stochastic quadratic expansion around the point of lack of identification.

Assumption A *Identification of data generating process*

(i) $\epsilon_t(\theta_n)$ *is a martingale difference sequence and* L_p *bounded for* $p = 4 + \iota$ *for small* ι.

(ii) $\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n)|X_t) = 0$ *a.s.*

(iii) $Var_{\theta_n}(\epsilon_t(\theta_n)|X_t) = \mathbb{E}_{\theta_n}(\epsilon_t^2(\theta_n)|X_t) = \sigma_t^2(\theta_n) > 0$ *a.s.*

(iv) Under weak identification $\mathbb{E}_{\theta_n}(\epsilon_t(\psi,\pi)d_{\psi,t}(\pi)) = 0$ *for unique* $\psi_n = (\zeta_n', \beta_n')'$ *in the interior of* $\Psi^*(\pi)$ *and under strong identification* $\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n)d_{\theta,t}(\theta_n)) = 0$ *for unique* $\theta_n = (\zeta_n', \beta_n', \pi_n')'$ *in the interior of* Θ^*.

Assumption B *Properties of data generating process*

(i) The random variables $W_t = \{y_t, X_t\}$ *are* $\alpha-$*mixing of size* $-r/(r-2)$ *for some* $r > 2$ *and* X_t *is* L_p *bounded for* $p = 4 + \iota$ *for small* ι.

(ii) The stochastic processes $\{d_{\psi_t}(\pi) : \pi \in \Pi\}$ *and* $\{d_{\theta_t}(\theta) : \theta \in \Theta\}$ *are* L_p *bounded, for* $p = 4 + \iota$ *with tiny* ι *and all* n, *that is* $\mathbb{E}_{\theta_n}(||(d_{\psi,t}(\pi)||^{4+\iota}) < C$ *and* $\mathbb{E}_{\theta_n}(||(d_{\theta,t}(\theta)||^{4+\iota}) < C$ *for some constant* C.

(iii) The stochastic processes $\{d_{\psi_t}(\pi) : \pi \in \Pi\}$ *and* $\{d_{\theta_t}(\theta) : \theta \in \Theta\}$ *are Lipschitz, that is* $\forall \pi, \tilde{\pi} \in \Pi$, $\exists C_n$ *such that* $||n^{-1}\sum_{t=1}^n d_{\psi_t}(\pi) - n^{-1}\sum_{t=1}^n d_{\psi_t}(\tilde{\pi})|| \leq C_n(W_t)||\pi - \tilde{\pi}||$, *and* $\forall \theta, \tilde{\theta} \in \Theta$, $||n^{-1}\sum_{t=1}^n d_{\theta_t}(\theta) - n^{-1}\sum_{t=1}^n d_{\theta_t}(\tilde{\theta})|| \leq C_n(W_t)||\theta - \tilde{\theta}||$ *for some random variable such that* $C_n(W_t) = O_p(1)$ *for all* n.

(iv) The function $h(\cdot)$ *is a Borel measurable function for each* $\pi \in \Pi$, *twice continuously differentiable in* Π, *non-degenerate and bounded for all values of* $X_{2,t}$ *and* π.

Assumption C *Long Run Variances*

(i) The limit variance of the stochastic processes $G_{\psi,n}$ *and* $G_{\theta,n}$ *are positive definite and finite, that is* $\forall \pi \in \Pi$, $\mathbb{E}_{\theta_n}(G_{\psi,n}(\pi)G_{\psi,n}(\pi)')$ *and* $\forall \theta \in \Theta$, $\mathbb{E}_{\theta_n}(G_{\theta,n}(\theta)G_{\theta,n}(\theta)')$ *are positive definite and finite.*

(ii) The uniform limits of the stochastic process and random matrix are positive definite and finite, that is $\forall \pi \in \Pi$, $\mathbb{E}_{\theta_n}(d_{\psi,t}(\pi)d_{\psi,t}(\pi)')$ *and* $\forall \theta \in \Theta$, $\mathbb{E}_{\theta_n}(d_{\theta,t}(\theta)d_{\theta,t}(\theta)')$ *are positive definite and finite.*

Assumption D *Parameter Spaces*

(i) Θ *is a compact set of* $\mathbb{R}^{d_\theta}$, *where* $d_\theta = d_\zeta + d_\beta + d_\pi$

(ii) $\Theta^* = \{(\beta^{*'}, \zeta^{*'}, \pi^{*'})' : \beta \in \mathcal{B}^*, \zeta \in \mathcal{Z}^*, \pi \in \Pi^*\}$ *is a compact set and* $\Theta^* \subset int(\Theta)$ *and* $0_{d_\beta} \in int(\mathcal{B}^*)$

(iii) $||\pi^*|| > \epsilon$ *for some* $\epsilon > 0$, $\pi^* \in \Pi$

Remark 1. Assumption A imposes correct specification, i.e., conditional mean zero of ϵ_t, for the whole sequence of drifting true parameters θ_n. Under strong identification, the conditional mean zero condition applies to the gradient with respect to all parameters, while under weak identification the mean zero expectation applies to the gradient with respect to ψ uniformly over π. The variance of ϵ_t is assumed to be time-varying, allowing for stochastic volatility models such as ARCH and GARCH.

Remark 2. We focus on a data generating process that allows time series data satisfying a strong mixing decay rate stated in Assumption B. We assume $d_{\psi,t}$ and $d_{\theta,t}$ are Lipschitz, which can be relaxed using the results of Newey (1991). The measurability and boundedness of function $h(\cdot)$ are standard in parametric models with known objective function.

Remark 3. Assumption C specifies that the limit variances must be nondegenerate and positive definite. These variance-covariance conditions do not apply to the weakly identified parameters because the second order derivative is singular by construction.

Remark 4. Assumption D is equivalent to the parameter space assumptions of Andrews and Cheng (2012). The compactness assumption is standard in the econometrics literature to show weak convergence. Moreover, as boundary effects are not the focus of this paper, we assume that the true parameters are in the interior of the parameter space and refer to Andrews (2001, 1999) for more details on the limit theory with boundary constraints.

Now we state the consistency result.

Proposition 1.3.1 *Suppose that Assumptions (A) to (D) hold. Under* $\{\theta_n\} \in \Theta(\theta_0)$,

(a) When $\beta_0 = 0$, *then* $\sup_{\pi \in \Pi} ||\hat{\psi}_n(\pi) - \psi_n|| \xrightarrow{p} 0$, *in particular* $\hat{\psi}_n(\hat{\pi}_n) - \psi_n \xrightarrow{p} 0$.

(b) When $\beta_0 \neq 0$, *then* $||\hat{\theta}_n - \theta_n|| \xrightarrow{p} 0$.

The theorem states that consistency under strong identification is achieved for all parameters, whereas under weak identification $\hat{\pi}_n$, is inconsistent. Even though we are not able to estimate π consistently, we can derive its random probability limit in Proposition 1.3.2.

Before we state the convergence in distribution result, we provide some intuition on how the result is derived (see further details in Appendix A.1). It is key to express the Taylor expansion around the point of lack of identification, i.e., setting $\beta = 0$. Taking a first-order condition of the objective function with respect to the strongly identified parameter ψ and letting $\psi_{0,n} = (\zeta_n, 0)$ yields:

$$
\begin{aligned}
0 = \frac{\partial}{\partial \psi} Q_n(\hat{\psi}_n(\pi), \pi) &= \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) + \frac{\partial^2}{\partial \psi \partial \psi'} Q_n(\psi_{0,n}, \pi)(\hat{\psi}_n - \psi_{0,n}) + o_p(1) \qquad (1.21)\\
\Leftrightarrow \frac{\partial^2}{\partial \psi \partial \psi'} Q_n(\psi_{0,n}, \pi)\sqrt{n}(\hat{\psi}_n - \psi_{0,n}) &= \sqrt{n}[\frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) - \mathbb{E}_{\theta_n} \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) + \mathbb{E}_{\theta_n} \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi)]
\end{aligned}
$$

The previous expression follows uniformly over π using the Mean Value Theorem at the point of lack of identification. At $\psi_{0,n}$, the objective function does not depend on π, and valid residuals can be estimated using $\hat{\psi}_{0,n}$. Using Equation (1.21) and the results from Appendix A.1, we derive the asymptotic distribution of $\hat{\psi}_n(\pi)$ under weak identification:

$$
\tau(\pi; \theta_0, b) = -H^{-1}(\pi; \theta_0)(G_\psi(\pi; \theta_0) + K(\pi; \theta_0)b) - (b, 0_{d_\zeta}) \qquad (1.22)
$$

The stochastic process $\tau(\pi; \theta_0, b)$ establishes that the asymptotic distribution of $\hat{\psi}_n$ depends on the value of π, which does not converge to a fixed constant under weak identification. Obtaining the limit distribution of π requires a quadratic expansion of the concentrated estimator. The limit distribution of π is a non-central χ^2 process:

$$
\xi(\pi; \theta_0, b) = -\frac{1}{2}(G_\psi(\pi; \theta_0) + K(\pi; \theta_0)b)' H^{-1}(\pi; \theta_0)(G_\psi(\pi; \theta_0) + K(\pi; \theta_0)b) \qquad (1.23)
$$

We assume that for each random sample path of $\xi(\pi; \theta_0, b)$, there exists a unique minimizer $\pi^*(\theta_0, b) = \arg\min_{\pi \in \Pi} \xi(\pi; \theta_0, b)$. The minimizer $\pi^*(\theta_0, b)$ defines the distribution function of $\hat{\psi}_n(\pi)$. Two more assumptions are stated to prove weak convergence of $\hat{\theta}_n$

Assumption E *Identification of* π

(i) Each sample path of the stochastic process $\{\xi(\pi;\theta_0,b) : \pi \in \Pi\}$ *is minimized over* Π *at a unique point denoted* $\pi^*(\theta_0,b)$ $\forall\theta_0 \in \Theta$ *in some set* $A(\theta_0,b)$ *with* $\mathbb{P}_{\theta_0}(A(\theta_0,b)) = 1$ *with* $\beta_0 = 0$ *and* $||b|| < \infty$.

(ii) $\mathbb{P}_{\theta_0}(\frac{\partial}{\partial\beta}\tau(\pi^*(b),b) = 0) = 0$

Remark 6. Assumption E is equivalent to Assumption C6 of Andrews and Cheng (2012). The Assumption guarantees that, although we cannot estimate the weakly identified parameter consistently, the random probability limit of π is a uniquely identified random variable, because $\pi^*(\theta_0,b)$ is unique for each sample path of $\xi(\pi;\theta_0,b)$.

Under weak identification, the Hessian matrix with respect to θ converges to a non-singular matrix for all identification cases in which $\beta \to 0$. Furthermore, the parameter ψ achieves a root-n limit distribution, while π has a different rate of convergence. Nonetheless, under semi-strong identification, a normalization matrix is enough to obtain a normal asymptotic distribution of $\hat{\theta}_n$ because π can be consistently estimated. To obtain a finite limit of the gradient and Hessian of the objective function, we define the normalization matrix as follows:

$$B(\beta) = \begin{bmatrix} I_{d_\psi} & 0_{d_\psi \times d_\pi} \\ 0_{d_\pi \times d_\psi} & \iota(\beta) \end{bmatrix} \tag{1.24}$$

where $\iota(\beta) = \beta$ is β is scalar and $\iota(\beta) = ||\beta||$ if β is a vector. Now define the variance-covariance matrix of the estimators as $\hat{\Sigma}_n = \hat{J}_n^{-1}\hat{V}_n\hat{J}_n^{-1}$, and the limit variance-covariance matrix $\Sigma(\theta_0) = J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0)$. The matrix $B(\beta)$ is used to obtain non-singular expressions of the variance-covariance matrix.

$$J_n = B^{-1}(\beta_n)\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_n)B^{-1}(\beta_n) \xrightarrow{p} J(\theta_0) \tag{1.25}$$

$$V_n = \sqrt{n}B^{-1}(\beta_n)\frac{\partial}{\partial\theta}Q_n(\theta_n) \xrightarrow{d} N(0, V(\theta_0)) \tag{1.26}$$

The next assumption imposes regularity conditions on the limit functions J and V.

Assumption F *Continuity and non-singularity of variance-covariance matrix*

(i) $J(\theta;\theta_0)$ *and* $V(\theta;\theta_0)$ *are continuous in* $\theta \in \Theta$, $\forall\theta_0 \in \Theta$ *with* $\beta_0 = 0$.

(ii) $J(\pi;\theta_0)$ *and* $V(\pi;\theta_0)$ *are positive definite matrices* $\forall\pi \in \Pi, \theta_0 \in \Theta$ *with* $\beta_0 = 0$. *(that is, the max-min eigenvalues are finite,* $\underline{eig}(J(\pi;\theta_0)), \underline{eig}(V(\pi;\theta_0)) > 0$ *and* $\overline{eig}(J(\pi;\theta_0)), \overline{eig}(V(\pi;\theta_0)) < \infty$*)*

Remark 5. The variance-covariance matrix of the estimators in Assumption F must be consistent regardless of the identification case. Under weak identification, consistency must be uniform over π because the asymptotic distribution of $\hat{\psi}_n(\pi)$ is a function of π. Assumption F guarantees that the limit exists and that it is positive semi-definite for all identification categories.

We state the convergence in distribution result for weak and strong identification.

Proposition 1.3.2 *Suppose that Assumptions (A) to (E) hold.*

(a) *Under* $\{\theta_n\} \in \Theta(\theta_0, 0, b)$ *with* $||b|| < \infty$, *the following holds*

$$\begin{pmatrix} \sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n) \\ \hat{\pi}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tau(\pi^*(\theta_0, b);\theta_0, b) \\ \pi^*(\theta_0, b) \end{pmatrix}$$

(b) *Under* $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, *the following holds*

$$\sqrt{n}B(\beta_n)(\hat{\theta}_n - \theta_n) \xrightarrow{d} N(0, J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0))$$

In sum, the strongly identified parameters can be consistently estimated for all identification categories, while under weak identification we cannot consistently estimate π. The asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_n)$ is asymptotically normal under semi-strong identification when we multiply it by the normalizing matrix $B(\beta)$. The asymptotic distribution of $\sqrt{n}(\hat{\psi}_n - \psi_n)$ is a functional of a Gaussian processes by the inconsistency of $\hat{\pi}_n$ and depends on the minimizer of the functional $\xi(\pi)$. The distribution of $\hat{\pi}_n$ is non-standard and influences the distribution of $\hat{\psi}_n$.

Even though we have a consistent estimator for β, we cannot estimate the parameter b consistently. As the process $\tau(\pi; \theta_0, b)$ is a function of b, the t-statistic we develop in Section 1.3.2 includes b as a nuisance parameter under the null hypothesis.

1.3.2 The t-statistic under weak and strong identification

We have derived the asymptotic distribution of $\hat{\theta}_n$ for all identification categories. In this subsection, we derive the distribution of the t-statistic. For the remainder of this subsection, we assume the parameter β is a scalar for ease of exposition. The asymptotic distribution of the t-statistic differs slightly when β is a vector (see Appendix A.3).

The null hypothesis of the t-test, which formally tests all values of the true sequences of parameters, is denoted by a restriction function $r : \Theta \to \mathbb{R}$ as follows:

$$H_0 : r(\theta_n) = q_n \tag{1.27}$$

Under weak identification, the asymptotic distribution of the t-statistic depends on which parameters are tested. Testing with respect to the parameter ζ is standard in a non-linear regression model; the classic t-statistic under strong identification is valid because the parameter's distribution is not affected nor does it affect the distribution of the weakly identified parameter π. The interesting cases are the t-statistics with restrictions with respect to β and π because the limit distribution is non-standard.

Let $\dim(r_\psi(\theta))$ and $\dim(r_\pi(\theta))$ denote the dimension of the derivative of the restriction function with respect to the parameters ψ and π. If the proposed null hypothesis solely concerns ψ, then $\dim(r_\psi(\theta)) = 1$, whereas, if the restriction concerns π, then $\dim(r_\pi(\theta)) = 1$.

We now consider the regularity assumptions required to obtain the asymptotic distribution of the t-statistic under all identification categories.

Assumption G *Properties of the restrictions function*

(i) $r(\theta) \in \mathbb{R}$ *is continuously differentiable on* Θ

(ii) $r_\theta(\theta) \neq 0 \ \forall \theta \in \Theta$

(iii) $rank(r_\pi(\theta)) = d^*_\pi$ *for some constant* $d^*_\pi \leq min(d_r, d_\pi), \forall \theta \in \Theta_\delta = \{\theta \in \Theta : ||\beta|| < \delta\}$

Remark 7. The conditions on Assumption G are non-restrictive. Differentiability is required to use the delta method of the restriction function.

It is well known that the standard sample t-statistic converges to a standard normal under strong identification. Under semi-strong and weak identification, the standard t-statistic does not have a finite probability limit as the variance-covariance matrix is singular. The sample t-statistic must be modified to obtain valid critical values under weak and semi-strong identification. The statistic requires a non-singular variance-covariance matrix of $\hat{\theta}_n$, which can be obtained using the normalization matrix $B(\beta)$ because the values that lead to lack of identification are known. Specifically, the robust t-statistic (robust in the sense of converging to a distribution for all identification categories) is defined as:

$$T_n = \frac{\sqrt{n}(r(\hat{\theta}_n) - v)}{[r_\theta(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)r_\theta(\hat{\theta}_n)']^{1/2}} \tag{1.28}$$

Under weak identification, the asymptotic distribution of the t-statistic is non-standard and varies according to the parameter tested. Testing with respect to the π modifies the limit distribution as the estimators $\hat{\psi}_n(\pi)$ and $\hat{\pi}_n$ have different rates of convergence. Specifically, when the null hypothesis tests a restriction on π, the randomness of $\hat{\pi}_n$ dominates the randomness of $\hat{\psi}_n$. When the restrictions of the t-test are with respect to the strongly identified parameters ψ, the limit distribution of the t-statistic T^ψ takes the following form:

$$T^\psi(\pi) = \frac{r_\psi(\pi)\tau(\pi; \theta_0, b)}{[r_\psi(\pi)\Sigma(\pi; \theta_0, b)r_\psi(\pi)']^{1/2}} \tag{1.29}$$

Conversely, when the t-test restrictions are with respect to π, i.e., $d_\pi^* = 1$, we must normalize the asymptotic distribution with respect to τ_β to obtain a finite limit distribution. Let $\tau_\beta(\pi;\pi_0,\sigma_t,b) = S_\beta \tau(\pi;\pi_0,\sigma_t,b) + bI_{d\beta}$ and $S_\beta = [I_{d\beta} : 0_{d_\beta \times d_\zeta}]$ be the selector matrix that selects β out of ψ. Define the asymptotic distribution of the t-statistic when we test with respect to π as T^π

$$T^\pi(\pi) = \frac{||\tau_\beta(\pi;\pi_0,b)||(r(\psi_0,\pi) - r(\psi_0,\pi_0))}{[r_\pi(\pi)\Sigma(\pi;\theta_0,b)r_\pi(\pi)']^{1/2}} \tag{1.30}$$

Finally, for the semi-strong identification case, $\{\theta_n\} \in \Theta(\theta_0,\infty,\omega_0)$, it is not surprising that, because the asymptotic distribution of $\hat{\theta}_n$ is normal under the proper normalization, the t-statistic has a standard normal distribution. Additionally, as π is consistent, the convergence result is pointwise and a standard central limit theorem applies. The following theorem formalizes the results of the asymptotic distribution of the t-statistic for all identification cases.

Proposition 1.3.3 *Suppose assumptions (A) to (H) hold*

(a) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$ *and* $\dim(r_\pi(\theta)) = 0$, $T_n \xrightarrow{d} T^\psi(\pi^*(\theta_0,b);\theta_0,b)$

(b) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$ *and* $\dim(r_\pi(\theta)) = 1$, $T_n \xrightarrow{d} T^\pi(\pi^*(\theta_0,b);\theta_0,b)$

(c) Under $\{\theta_n\} \in \Theta(\theta_0,\infty,\omega_0)$, $T_n \xrightarrow{d} N(0,1)$

The asymptotic distribution under semi-strong identification of the t-statistic is equivalent to the asymptotic distribution under strong identification. The robust t-statistic critical values involve exclusively two cases: the weak/non-identification (non-standard) distribution, and the semi-strong/strong identification standard normal distribution. For this reason, unless the distinction is necessary, we use the term "weak identification" to characterize non and weak identification categories, and we use the term "strong identification" for the strong and semi-strong identification. As the asymptotic distribution only depends on these two grouped identification categories, we develop a bootstrap process for weak/non-identification and another for the strong/semi-strong identification.

1.4 Wild Bootstrap under Weak and Semi-strong Identification

Constructing critical values of the t-test using Proposition 1.3.3 involves simulating a stochastic process dependent on the (unknown) randomness structure of π (see Equation (1.32)). We propose using a wild bootstrap approach which multiplies by a random variable z_t^m, the process $G_{\psi,n}(\pi)$ to separate the effects from the randomness structure of π and the randomness from the data probability space. This bootstrapped process is asymptotically valid while it performs better in small samples. For example, if the asymptotic distribution is known to be symmetric but the small sample distribution is skewed; a wild bootstrap with a skewed process z_t^m would lead to better inference. As under weak identification, no consistent estimator of π is available, the wild bootstrap cannot be constructed using the residuals $\epsilon_t(\hat{\theta}_n)$. Nonetheless, we can construct the bootstrap using the errors centered at the point of lack of identification $\epsilon_t(\psi_{0,n})$ since they are consistent estimates of $\epsilon_t(\theta_0)$.

1.4.1 Wild bootstrapped limit distributions

Two bootstrapped processes are necessary to construct the bootstrap distribution of the t-test. Let $G^m_{\psi,n}$ denote the bootstrapped process under weak identification, and let $G^m_{\theta,n}$ be the process under strong identification. First, we focus on the weak identification case and establish the steps to derive the wild bootstrapped process $G^m_{\psi,n}$. From the definition of $G_{\psi,n}$, centering ϵ_t at the point $\psi_{0,n}$, we obtain the following weak convergence result:

$$
\begin{aligned}
G_{\psi,n}(\psi_{0,n},\pi) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\frac{\partial}{\partial\psi}\epsilon_t^2(\psi_{0,n},\pi) - \mathbb{E}_{\theta_n}\frac{\partial}{\partial\psi}\epsilon_t^2(\psi_{0,n},\pi)] \\
&= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[(y_t-\zeta_n' X_{1,t})d_{\psi,t}(\pi) - \mathbb{E}_{\theta_n}((y_t-\zeta_n' X_{1,t})d_{\psi,t}(\pi))] \\
&= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t-\zeta_n' X_{1,t}-\beta_n' h(X_{2,t},\pi_n))d_{\psi,t}(\pi) \\
&+ b\ n^{-1}\sum_{t=1}^{n}[h(X_{2,t},\pi_n)d_{\psi,t}(\pi) - \mathbb{E}_{\theta_n} h(X_{2,t},\pi_n)d_{\psi,t}(\pi)] \\
&= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_n)d_{\psi,t}(\pi) + o_{p,\pi}(1) \Rightarrow G_\psi(\pi)
\end{aligned}
\tag{1.31}
$$

where $G_\psi(\pi)$ is a zero mean Gaussian process with covariance kernel:

$$\Omega(\pi, \tilde{\pi}; \theta_0) = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathbb{E}_{\theta_n}(\epsilon_t^2(\theta_n) d_{\psi,t}(\pi) d_{\psi,t}(\tilde{\pi})) \tag{1.32}$$

where Equation (1.32) follows as $\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n) d_{\psi,t}(\pi)) = 0, \ \forall \pi \in \Pi$ by Assumption A, while the weak convergence result follows from Lemmas A.2.1 to A.2.3 in the Appendix. We can construct a multiplier bootstrap to generate random draws of Equation (1.31). We define $G^m_{\psi,n}(\pi)$ as a wild bootstrapped stochastic process, centered at the point of lack of identification $\psi_{0,n}$,

$$G^m_{\psi,n}(\pi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_t^m f(\epsilon_t(\psi_{0,n})) d_{\psi,t}(\pi) \tag{1.33}$$

For this study, we let $f(\epsilon_t(\psi_{0,n})) = \epsilon_t(\psi_{0,n})$ because $\psi_{0,n}$ can be estimated consistently, and we assume ϵ_t is a martingale difference sequence. Under more stringent assumptions such as ϵ_t being i.i.d., a more convenient choice of $f(\cdot)$ would be $f(\epsilon_t(\psi_{0,n})) = \sigma^2_\epsilon$, which is equivalent to performing the resampling bootstrap of Efron et al. (1979). The assumptions required for z_t^m to converge weakly to the Gaussian process limit distribution are $\mathbb{E}(z_t^m) = 0$ and $Var(z_t^m) = 1$ and $\mathbb{E}|z_t^m|^{2+\iota} \leq C < \infty$. A common choice in the literature is $z_t^m \sim N(0,1)$, which is convenient when the small sample distribution is close to the asymptotic distribution. An option to obtain small sample improvements over the asymptotic approximation is to use the Rademacher distribution suggested by Liu et al. (1988).

$$z_t^m = \begin{cases} 1 & \text{with probability} \quad 1/2 \\ -1 & \text{with probability} \quad 1/2 \end{cases} \tag{1.34}$$

The simulation results of Davidson, Monticini, and Peel (2007) and Davidson and Flachaire (2008) show that the Rademacher distribution performs better than other distributions if the conditional distribution of the errors is symmetric and suggest it should be preferred in practice. When the distribution of the errors is asymmetric, a more convenient choice is the two point distribution proposed by Mammen (1993).

$$z_t^m = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability} \quad (\sqrt{5}+1)/(2\sqrt{5}) \\ (\sqrt{5}+1)/2 & \text{with probability} \quad (\sqrt{5}\text{ - }1)/(2\sqrt{5}) \end{cases} \tag{1.35}$$

Lemma A.2.9 in the Appendix shows $G^m_{\psi,n} \Rightarrow G_\psi(\pi)$ is asymptotically valid for any of the options of z_t^m described above. The choice of z_t^m does not affect the first-order terms and consequently does not affect the asymptotic distribution. Nonetheless it can lead to very different critical values in small samples. The bootstrapped process $G^m_{\psi,n}(\pi)$ is used to construct the bootstrapped distributions of $\hat{\pi}_n$ and $\hat{\psi}_n$ in the same fashion as in Equation (1.18) to (1.23). Specifically, we construct the bootstrapped stochastic processes under weak identification according to the following expressions:

$$H_n(\pi) = \frac{\partial^2}{\partial\psi\partial\psi'}Q_n(\psi_{0,n},\pi) = n^{-1}\sum_{t=1}^n d_{\psi,t}(\pi)d_{\psi,t}(\pi)' \tag{1.36}$$

$$K_n(\pi,\pi_n) = \frac{\partial}{\partial\beta_n}\mathbb{E}_{\theta_n}\frac{\partial}{\partial\psi}Q_n(\theta) = -n^{-1}\sum_{t=1}^n h(X_{2,t},\pi_n)d_{\psi,t}(\pi) \tag{1.37}$$

$$\xi_n^m(\pi,\pi_n,b) = -\frac{1}{2}[G^m_{\psi,n}(\pi) + K_n(\pi,\pi_n)b]'H_n(\pi)^{-1}[G^m_{\psi,n}(\pi) + K_n(\pi,\pi_n)b] \tag{1.38}$$

$$\pi_n^m = \underset{\pi\in\Pi}{\arg\min}\ \ \xi_n^m(\pi,\pi_n,b) \tag{1.39}$$

$$\tau_n^m(\pi_n^m,\pi_n,b) = -H_n^{-1}(\pi_n^m)(G^m_{\psi,n}(\pi_n^m) + K_n(\pi_n^m,\pi_n)b) - (b,0) \tag{1.40}$$

Next, we construct the bootstrapped distribution of the parameters $\hat{\psi}_n$ and $\hat{\pi}_n$ under strong identification. The major difference with respect to the weak identification case hinges on the construction of a bootstrapped process G that does not depend on π. This construction leads to theoretically and practically simpler expressions that can be simulated using a sequence of independent draws instead of sample paths depending on a grid of π. When we let $\{\theta_n\} \in \Theta(\theta_0,\infty,\omega_0)$, the wild bootstrapped process $G^m_{\theta,n}$ under semi-strong identification takes the following form:

$$G^m_{\theta,n}(\theta_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^n z_t^m\epsilon_t(\theta_n)d_{\theta,t}(\theta_n) \tag{1.41}$$

Moreover, the normalizing matrix $B(\beta)$ is required to obtain a non-singular variance-covariance matrix in the limit. With the bootstrapped stochastic process $G^m_{\theta,n}$ and the consistent probability limits J_n, V_n, we can derive the asymptotic distribution of $\hat{\theta}_n$ under strong identification which we denote by $\tau^m_{\theta,n}$

$$\tau^m_{\theta,n}(\theta) = [B^{-1}(\beta_n)\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta)B^{-1}(\beta_n)]^{-1}B^{-1}(\beta_n)\sqrt{n}\frac{\partial}{\partial\theta}Q^m_n(\theta) \tag{1.42}$$

$$= J^{-1}_n(\theta)G^m_{\theta,n}(\theta) \tag{1.43}$$

where $J_n(\theta) = n^{-1}\sum_{t=1}^n d_{\theta,t}(\theta)d_{\theta,t}(\theta)'$ and $\sqrt{n}\frac{\partial}{\partial\theta}Q^m_n(\theta) = G^m_{\theta,n}(\theta)$. The bootstrapped t-statistic distribution under semi-strong identification is replicated using Equations (1.41) to (1.43) as well as Equation (1.29) and (1.30).

The following consistency theorem requires the introduction of some notation. Define $c^\tau_{1-\alpha} = \inf\{z \in \mathbb{R}^{d_\psi} : \mathbb{P}_{\theta_0}(\tau(\pi^*(\theta_0,b);\theta_0,b) \leq z) \geq 1-\alpha\}$ and let $c^\pi_{1-\alpha} = \inf\{z \in \mathbb{R}^{d_\pi} : \mathbb{P}_{\theta_0}(\pi^*(\theta_0,b) \leq z) \geq 1-\alpha\}$ be the $1-\alpha$ asymptotic critical values of $\tau(\pi^*(\theta_0,b);\theta_0,b)$ and $\pi^*(\theta_0,b)$ respectively. To obtain the critical values of the bootstrapped statistics, let M_n denote the number of bootstrapped samples. For the sequence of bootstrapped distributions $\{\hat{\tau}^m_n\}^{M_n}_{m=1} = \{\tau^m_n(\hat{\pi}^m_n;\pi_0,b)\}^{M_n}_{m=1}$, denote the order statistics $\hat{\tau}^{[1]}_n \leq \hat{\tau}^{[2]}_n \leq ... \leq \hat{\tau}^{[M_n]}_n$. The approximated $1-\alpha$ critical value of $\{\hat{\tau}^m_n\}^{M_n}_{m=1}$ is defined by $c^{m,\tau}_{n,1-\alpha} = \hat{\tau}^{[(1-\alpha)M_n]}_n$. For $\{\hat{\pi}^m_n\}^{M_n}_{m=1}$, define $c^{m,\pi}_{n,1-\alpha} = \hat{\pi}^{[(1-\alpha)M_n]}_n$. Equivalently, under strong identification let $c^{\tau_\theta}_{1-\alpha}$ be the $1-\alpha$ critical value of the distribution $N(0, J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0))$. Also, let $\{\hat{\tau}^m_{\theta,n}\}^{M_n}_{m=1} = \{\tau^m_{\theta,n}(\hat{\theta}_n)\}^{M_n}_{m=1}$ be the bootstrapped samples with order statistics $\hat{\tau}^{[1]}_n \leq \hat{\tau}^{[2]}_n \leq ... \leq \hat{\tau}^{[M_n]}_n$ and $1-\alpha$ critical values $c^{m,\tau_m}_{n,1-\alpha} = \hat{\tau}^{[(1-\alpha)M_n]}_{\theta,n}$. The next theorem shows that the bootstrap procedure is valid.

Theorem 1.4.1 *Suppose that Assumptions (A) to (E) hold. Under weak identification, let $\hat{\pi}^m_n$ be constructed using Equation* (1.39) *and $\hat{\tau}^m_n(\hat{\pi}^m_n;\pi_0,b)$ be constructed using Equation* (1.40)*, while under strong identification let $\hat{\tau}^m_{\theta,n}(\hat{\theta}_n)$ be constructed using Equation* (1.42)*. Denote $c^{m,a}_{n,1-\alpha}$ and $c^a_{1-\alpha}$ with $a = \pi, \tau, \tau_\theta$ be the critical values of the bootstrapped and asymptotic distributions, respectively. Letting $M_n \to \infty$ as $n \to \infty$ then* *(a) Under* $\{\theta_n\} \in \Theta(\theta_0, 0, b)$ *with* $||b|| < \infty$, $|c^{m,\tau}_{n,1-\alpha} - c^{\tau}_{1-\alpha}| \xrightarrow{p} 0$ *and* $|c^{m,\pi}_{n,1-\alpha} - c^{\pi}_{1-\alpha}| \xrightarrow{p} 0.$

(b) Under $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, $|c^{m,\tau_\theta}_{n,1-\alpha} - c^{\tau_\theta}_{1-\alpha}| \xrightarrow{p} 0.$

It is worth mentioning that contrary to $G_{\psi,n}$, the bootstrapped stochastic process $G^m_{\psi,n}$ does not have to be demeaned by a non-zero expectation. In practice, the unknown expectation of $G_{\psi,n}(\pi)$, namely $\mathbb{E}_{\theta_n}(\epsilon_t(\psi_{0,n})d_{\psi,t}(\pi))$ is not required to construct the bootstrap process. The bootstrapped stochastic process $G^m_{\psi,n}$ is mean zero by construction. In contrast, the sample process $G_{\psi,n}(\pi)$ derived has a non-zero expectation $\mathbb{E}_{\theta_n}\epsilon_t(\psi_{0,n}d_{\psi,t}(\pi))$ that has to be estimated. This expectation not only depends on the unknown true parameters θ_n but also depends on the expectation of the induced measure with respect to the true parameters, which is not simple to estimate. The estimation error that can arise from the estimation of this expectation is not present in the wild bootstrap estimation, suggesting that bootstrapped distributions are more precise in practice.

1.4.2 The bootstrapped t-statistic

The bootstrapped t-test is derived using a two-step procedure. In the first step, we estimate $\hat{\pi}^m_n$, $\hat{\tau}^m_n$ and $\hat{\tau}^m_{\theta,n}$ and the relevant probability limits using Equation (1.31) to (1.43). The second step consists of constructing the asymptotic distribution of Proposition 1.3.3 using the bootstrapped processes. Specifically, let $T^m_{\psi,n}$, $T^m_{\pi,n}$ be bootstrapped asymptotic distributions under weak identification incorporating the bootstrapped processes $\hat{\pi}^m_n, \hat{\tau}^m_n, \hat{\tau}^m_{\theta,n}$.

$$T^m_{\psi,n} = T^{\psi}(\hat{\pi}^m_n) = \frac{r_\psi(\hat{\pi}^m_n)\tau^m_n(\hat{\pi}^m_n)}{[r_\psi(\hat{\pi}^m_n)\hat{\Sigma}_n(\hat{\pi}^m_n)r_\psi(\hat{\pi}^m_n)']^{1/2}} \tag{1.44}$$

$$T^m_{\pi,n} = T^{\pi}(\hat{\pi}^m_n) = \frac{||\tau_\beta(\hat{\pi}^m_n)||(r(\hat{\psi}_n(\hat{\pi}^m_n), \hat{\pi}^m_n) - r(\hat{\psi}_n(\hat{\pi}^m_n), \pi_0))}{[r_\pi(\hat{\pi}^m_n)\hat{\Sigma}_n(\hat{\pi}^m_n)r_\pi(\hat{\pi}^m_n)']^{1/2}} \tag{1.45}$$

As the parameters ψ and π have different rates of convergence, the asymptotic distribution of the t-statistic depends on the identification category and the parameters that the null hypothesis is testing.

Under strong identification, the t-statistic is evaluated at $\hat{\theta}_n$ and takes the form

$$T^m_{\theta,n}(\hat{\theta}_n) = \frac{r_\theta(\hat{\theta}_n)\tau^m_{\theta,n}(\hat{\theta}_n)}{[r_\theta(\hat{\theta}_n)\hat{\Sigma}_n(\hat{\theta}_n)r_\theta(\hat{\theta}_n)']^{1/2}} \tag{1.46}$$

Using the definitions in Theorem 1.4.1, let $c^{m,a}_{n,1-\alpha}$ with T^ψ, T^π, T^θ denote the $1-\alpha$ critical value of the bootstrapped t-statistics $T^m_{\psi,n}, T^m_{\pi,n}, T^m_{\theta,n}$. Similarly, let $c^a_{1-\alpha}$ with $a = T_\psi, T_\pi, N$ denote the $1-\alpha$ critical value of T^ψ, T^π and $N(0,1)$. The next theorem demonstrates that the bootstrapped statistics converge to the same limit distributions and therefore can be used for hypothesis testing.

Theorem 1.4.2 *Suppose assumptions (A) to (H) hold. Let $T^m_{\psi,n}$, $T^m_{\pi,n}$ be the bootstrapped t-statistics of Equation* (1.44) *to* (1.46). *Denote $c^{m,a}_{n,1-\alpha}$ and $c^a_{1-\alpha}$ with $a = T^\psi, T^\pi, T^\theta, N$ the $1-\alpha$ critical values of the bootstrapped and asymptotic distributions, respectively. Let $M_n \to \infty$ as $n \to \infty$*

(a) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ with $||b|| < \infty$ and $\dim(r_\pi(\theta)) = 0$, $|c^{m,T^\psi}_{n,1-\alpha} - c^{T^\psi}_{1-\alpha}| \xrightarrow{p} 0$

(b) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ with $||b|| < \infty$ and $\dim(r_\pi(\theta)) = 1$, $|c^{m,T^\pi}_{n,1-\alpha} - c^{T^\pi}_{1-\alpha}| \xrightarrow{p} 0$

(c) Under $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, $|c^{m,T^\theta}_{n,1-\alpha} - c^{N}_{1-\alpha}| \xrightarrow{p} 0$

Under weak identification, the bootstrapped distributions $T^m_{\psi,n}$, $T^m_{\pi,n}$ depend on the nuisance parameters. The nuisance parameters in the non-linear model are (b, π_0). To recover the critical values, we consider a grid of nuisance parameters $p \in \mathcal{P}$[2] and use the supremum over that grid. Under strong identification, we do not have to deal with nuisance parameters because we have a consistent estimator of π_0 and the value b has no role outside of weak identification.

After deriving the bootstrapped distribution under weak and strong identification, the salient question is "which critical values should be chosen if the identification category is

[2]The nuisance parameters can be reduced if we use null imposed critical values. For example, if we are testing $\beta_n = 0$ and use null imposed critical values, by construction the number of nuisance parameters is reduced as $b = 0$ under the null hypothesis.

unknown"?. We consider two specifications of robust critical values which incorporate the critical values under weak and strong identification: 1) Least Favorable Critical Value (LF) and 2) Identification Category-Selection Type 1 (ICS_1).

The LF critical value, as the name suggests, adopts the largest critical value regardless of the true identification category. Let $c^m_{n,1-\alpha}(p)$, $p \in \mathcal{P}$ denote the critical values of the weakly identified bootstrap statistic and $c^m_{n,\theta,1-\alpha}$ denote the critical values of the semi-strong identified bootstrap statistic. The LF critical value is defined as

$$c^{LF,m}_{n,1-\alpha} = \max\{\sup_{p\in\mathcal{P}} c^m_{n,1-\alpha}(p), c^m_{n,\theta,1-\alpha}\} \tag{1.47}$$

The LF critical value is a naive selection as we use the larger critical values regardless of the identification case. To improve the size and power of the test, Andrews and Cheng (2012) propose a data-driven critical value, ICS_1, that relies on a first-step test to determine whether b is finite, implying weak identification. If we cannot reject the null hypothesis of finite b, the LF critical value is selected; otherwise, the strong identification bootstrapped critical value is used. The ICS_1 critical values are defined by

$$c^{ICS_1,m}_{n,1-\alpha} = \begin{cases} c^{LF,m}_{n,1-\alpha} & \text{if} \quad A_n \le \kappa_n \\ c^m_{n,\theta,1-\alpha} & \text{if} \quad A_n > \kappa_n \end{cases} \tag{1.48}$$

$$A_n = (n\hat{\beta}_n\hat{\Sigma}^{-1}_n\hat{\beta}_n)^{1/2} \tag{1.49}$$

where A_n is a first step statistic and κ_n is a sequence of positive constants such that $\kappa_n \to \infty$ and $\kappa_n/n^{1/2} = o(1)$. For example, $\kappa_n = (\log n)^{1/2}$ is analogous to the penalty term used in the Bayesian Information Criterion. The null hypothesis $H_0 : ||b|| < \infty$ is accepted in favor of weak identification when $A_n \le \kappa_n$, while the null hypothesis is rejected in favor of strong identification if $H_1 : ||b|| = \infty$ when $A_n > \kappa_n$. We need one more assumption about the validity of the critical values proposed. Let $c_{a-\alpha}(\infty)$ denote the $1-\alpha$ quantile of a normal standard distribution and let $\hat{\Sigma}_n^{-1}$ be the upper left variance matrix size $d_\beta \times d_\beta$ in Assumption F.

Assumption H *Properties of LF and ICS_1 critical values*

(i) The distribution function of T^ψ and T^π are continuous at $c_{1-\alpha}(p)$, $\forall p \in \mathcal{P}$. If $c_{1-\alpha}^{LF} > c_{1-\alpha}(\infty)$, $c_{1-\alpha}^{LF}$ is attained at some $p_{max} \in \mathcal{P}$.

(ii) $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$

(iii) The distribution function of T^ψ and T^π are continuous at the critical values $c_{1-\alpha}$, $\forall p \in \mathcal{P}$.

Remark 8. Assumption H is equivalent to Assumptions LF, K and V3 of Andrews and Cheng (2012).

To prove that the robust critical values lead to the correct asymptotic size, Andrews and Cheng (2012) introduce the asymptotic size of the t-test under different identification categories. We are interested in the effects of the Confidence Sets from various null hypotheses of θ. Recall the null hypothesis for any element of the drifting sequence is $H_0 : r(\theta) = q$. The Confidence Sets are obtained by inverting a test. For example, the t-statistic $1 - \alpha$ Confidence Set under $r(\theta)$ is defined as:

$$CS_n = \{q : T_n(q) \leq c_{n,1-\alpha}(q)\} \tag{1.50}$$

Notice that the Confidence Sets are a function of the value q as well as sample size n. The coverage probability of a Confidence Set for $r(\theta)$ is

$$CP_n(\theta) = \mathbb{P}_\theta(r(\theta) \in CS_n) = \mathbb{P}_\theta(T_n(r(\theta) \leq c_{n,1-\alpha}(r(\theta)))) \tag{1.51}$$

An important measure of the t-test is the maximum null rejection probability as it is equivalent to the asymptotic size of the test. The test null rejection probability is defined by

$\mathbb{P}_\theta(T_n(r(\theta)) > c_{n,1-\alpha}(r(\theta)))$ with asymptotic size

$$AsySz = \liminf_{n\to\infty} \inf_{\theta\in\Theta:r(\theta)=q} \mathbb{P}_\theta(T_n(r(\theta)) \leq c_{n,1-\alpha}(r(\theta))) \tag{1.52}$$

We need to introduce one more assumption need to assume the convergence of the Coverage probabilities for all nuisance parameters. For more details on asymptotic size, see Section 2 of Andrews and Cheng (2012), in particular Lemma 2.1.

Assumption I *Properties of the Coverage Probabilities*

(i) Under weak identification for any $\theta_n \in \Theta(\theta_0, 0, b)$ with $b < \infty$, $CP_n(\theta_n) \to CP(p)$ for some $CP(p) \in [0,1]$, where $p = (b, \theta_0) \in \mathcal{P}$.

(ii) Under strong identification for any $\theta_n \in \Theta(\theta_0, \infty, \omega_0)$, $\liminf_{n\to\infty} CP_n(\theta_n) \geq CP_\infty$ for $CP_\infty \in [0,1]$ and for some θ_n, $CP_n(\theta_n) \to CP_\infty$.

(iii) For some $\delta > 0$, $\theta = (\zeta', \beta', \pi')' \in \Theta$ with $0 < ||\beta|| < \delta$ implies that $\tilde{\theta} = (\zeta', \tilde{\beta}', \pi')' \in \Theta$ for all $\tilde{\beta} \in R^{d_\beta}$ with $0 < ||\tilde{\beta}|| < \delta$.

Remark 9. Assumption I is equivalent to Assumption ACP of Andrews and Cheng (2012).

To obtain robustness against all identification categories, the critical values must either change according to the unknown identification category or be the largest critical values according to the asymptotic distribution obtained for each case. The ICS_1 and LF critical values fulfill these requirements, as proven in the following theorem. Further, the power of the test approaches 1 in the limit. For our purposes, we focus on the asymptotic power of the test, instead of the Confidence Sets, which are defined for each $n \geq 1$.

Theorem 1.4.3 *Let $M_n \to \infty$ as $n \to \infty$. Under the null hypothesis $H_0 : r(\theta) = q$, the LF and ICS_1 critical values of the t-test have the correct asymptotic size with probability approaching one,*

(a) $AsySz^{LF,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta:r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c^{LF,m}_{n,1-\alpha}(r(\theta))) = 1 - \alpha$

(b) $AsySz^{ICS_1,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta:r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c_{n,1-\alpha}^{ICS_1,m}(r(\theta))) = 1-\alpha$

(c) If H_0 *is false, that is* $r(\theta_n) \neq q$*, then* $T_n(r(\theta)) \xrightarrow{p} \infty$

1.5 Simulation Results

The simulation exercise compares the performance of the robust t-test which has correct asymptotic size with respect to all identification categories, and the standard t-test, which invariably assumes strong identification. The data generating process assumes strong, weak or non-identified according to the value of β_n. Moreover, we compare the differences in the size and power of using the wild bootstrapped t-statistic against the asymptotic approximation distribution of Andrews and Cheng (2012, 2014, 2013).

We assume a non-linear model with an autoregressive exponential smoothing specification as in Cheng (2015). The data generating process of the residuals takes the following form:

$$\epsilon_t(\theta_n) = y_t - \zeta_0 x_{1,t} - \beta_n y_{t-1}(1 - \exp(-c(x_{2,t} - \pi_0)^2)) \qquad (1.53)$$

We require a drifting sequence of true parameters for β_n only, albeit ζ and π have a unique true value ζ_0 and π_0. The simulations are constructed by assuming the following true values: $\zeta_0 = 1, \pi_0 = 0$ and $b = 0.9$. The drifting sequence of true parameters β_n takes the values $\beta_n = b$ under strong identification, $\beta_n = b/\sqrt{n}$ under weak identification and $\beta_n = 0$ under non-identification. We assume $x_{1,t}$, $x_{2,t} \sim N(0,1)$ and set $c = -1$ to reduce the optimization parameter space. Four distributions for the errors ϵ_t are considered: standard normal $N(0,1)$, t-distribution with 4 degrees of freedom $t(4)$ and GARCH(1,1) errors with parameters $\omega = 0.1, \alpha = 0.6, \beta = 0.3$ and $\omega = 0.1, \alpha = 0.3, \beta = 0.6$. The results are estimated for the sample sizes $n =$ 100, 250 and 500. We use two tailed critical values and assume z_t^m as in Mammen (1993). Under weak identification the asymptotic distribution is asymmetric and can have large skewness and kurtosis, which justifies the use of two tailed critical values. The number of simulations is $1,000$. For each simulation, we construct a bootstrapped and asymptotic approximation distribution to derive their critical values using 500 samples. For brevity we report the size and power of the test for sample size $n = 250$ and $\epsilon_t \sim N(0,1)$ and GARCH(1,1). The Appendix contains the remaining tables.

We estimate the model by Least Squares. Let $\epsilon_t(\theta) = y_t - \zeta x_{1,t} - \beta y_{t-1}(1 - \exp(-c(x_{2,t} - \pi)^2))$, and $\theta = (\zeta, \beta, \pi)'$. In the simulation exercise, $h(X_{2,t}, \pi)) = y_{t-1}(1 - \exp(-c(x_{2,t} - \pi)^2))$.

$$
\begin{aligned}
d_{\psi,t} &= (X_{1,t}, h(X_{2,t}, \pi))' & d_{\theta,t} &= (X_{1,t}, h(X_{2,t}, \pi), h_\pi(X_{2,t}, \pi))' \\
Q_n(\theta) &= \frac{1}{2n}\sum_{t=1}^{n} \epsilon_t^2(\theta) & \frac{\partial}{\partial \theta} Q_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n} \epsilon_t(\theta) B(\beta) d_{\theta,t}(\pi) \\
\frac{\partial}{\partial \psi} Q_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n} \epsilon_t(\theta) d_{\psi,t}(\pi) & \frac{\partial^2}{\partial \psi \partial \psi'} Q_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} d_{\psi,t}(\pi) d_{\psi,t}(\pi)'
\end{aligned}
\tag{1.54}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} [B(\beta) d_{\theta,t}(\theta) d_{\theta,t}(\theta)' B(\beta) - \epsilon_t(\theta) D_t(\theta)] \\
D_t(\theta) &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & h_\pi(X_{2,t}, \pi) \\ 0 & h_\pi(X_{2,t}, \pi) & \beta h_{\pi,\pi}(X_{2,t}, \pi) \end{pmatrix}
\end{aligned}
\tag{1.55}
$$

For ease of exposition, we state the steps required to derive the bootstrapped t-test. These steps follow the results of Section 1.4. For more details on the estimation see Appendix A.4.

Step 1. Construct the following matrices

$$
\begin{aligned}
H_n(\pi, \tilde{\pi}) &= n^{-1}\sum_{t=1}^{n} d_{\psi,t}(\pi) d_{\psi,t}(\pi)' & J_n(\theta) &= n^{-1}\sum_{t=1}^{n} d_{\theta,t}(\theta) d_{\theta,t}(\theta)' \\
K_n(\pi, \pi_n) &= -n^{-1}\sum_{t=1}^{n} h(X_t, \pi_n) d_{\psi,t}(\pi) & V_n(\theta) &= n^{-1}\sum_{t=1}^{n} \epsilon_t^2(\theta) d_{\theta,t}(\theta) d_{\theta,t}(\theta)' \\
H(\pi; \theta_0) &= \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \mathbb{E}_{\theta_n}(d_{\psi,t}(\pi) d_{\psi,t}(\pi)') & J(\theta_n) &= n^{-1}\sum_{t=1}^{n} \mathbb{E}_{\theta_n}(d_{\theta,t}(\theta_n) d_{\theta,t}(\theta_n)') \\
K(\pi, \pi_n; \theta_0) &= \lim_{n\to\infty} -n^{-1}\sum_{t=1}^{n} \mathbb{E}_{\theta_n}(h(X_t, \pi_n) d_{\psi,t}(\pi)) & V(\theta_n) &= n^{-1}\sum_{t=1}^{n} \mathbb{E}_{\theta_n}(\epsilon_t^2 d_{\theta,t}(\theta_n) d_{\theta,t}(\theta_n)') \\
\Sigma(\theta_0) &= J^{-1}(\theta_0) V(\theta_0) J^{-1}(\theta_0) & \hat{\Sigma}_n &= J_n^{-1}(\hat{\theta}_n) V_n(\hat{\theta}_n) J_n^{-1}(\hat{\theta}_n)
\end{aligned}
\tag{1.56}
$$

Note that the only variables that depend on the nuisance parameter π_n are the stochastic processes $K_n(\pi)$ and $K(\pi)$.

Step 2. Construct the weak identification bootstrap process $G^m_{\psi,n}(\psi_{0,n},\pi)$ and the strong identified process $G^m_{\theta,n}(\hat{\theta}_n)$ developed in Section 1.4 where $\hat{\epsilon}_t(\hat{\psi}_{0,n})$ denote the residuals centered at the point of lack of identification and $\hat{\psi}_{0,n}$ is a consistent estimator of ψ_0

$$\begin{aligned}
G^m_{\psi,n}(\pi) &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t^m \epsilon_t(\hat{\psi}_{0,n}) d_{\psi,t}(\pi) \\
\xi_n^m(\pi,\pi_n,\hat{\sigma}_t,b) &= -\frac{1}{2}[G^m_{\psi,n}(\pi)+K_n(\pi,\pi_n)b]' H_n(\pi)^{-1}[G^m_{\psi,n}(\pi)+K_n(\pi,\pi_n)b] \\
\hat{\pi}_n^m &= \arg\min \ \xi_n^m(\pi,\pi_n,\hat{\sigma}_t,b) \\
\tau_n^m(\hat{\pi}_n^m;\pi_n,b) &= -H_n^{-1}(\pi)(G^m_{\psi,n}(\hat{\pi}_n^m)+K_n(\hat{\pi}_n^m,\pi_n)b)-(b,0) \qquad (1.57)\\
G^m_{\theta,n}(\hat{\theta}_n) &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t^m \epsilon_t(\hat{\psi}_{0,n}) d_{\theta,t}(\hat{\theta}_n) \\
\tau^m_{\theta,n}(\hat{\theta}_n) &= [n^{-1}\sum_{t=1}^{n} d_{\theta,t}(\hat{\theta}_n) d_{\theta,t}(\hat{\theta}_n)']^{-1} G^m_{\theta,n}(\hat{\theta}_n)
\end{aligned}$$

As the error process ϵ_t is an martingale difference sequence, the residuals $\epsilon_t(\hat{\psi}_{0,n})$ are heterogeneously distributed. The bootstrapped process τ_n^m depends on the minimizer $\hat{\pi}_n^m$, which is derived using a grid of 1001 values of π within the interval $[\pi_0-2,\pi_0+2]$. It is worth noting that the wild bootstrap procedure presented here avoids the inconvenience of simulating the (unknown) stochastic process $G_\psi(\pi)$.

Step 3. Define the null hypothesis $r(\theta_n)=q$. If the null hypothesis sets restrictions with respect to π, we use the asymptotic process defined by T^π, whereas if the restrictions depend solely on ψ, we construct the statistic T^ψ.

$$T^m_{\psi,n} = T^\psi(\hat{\pi}_n^m) = \frac{r_\psi(\hat{\pi}_n^m)\tau_n^m(\hat{\pi}_n^m)}{[r_\psi(\hat{\pi}_n^m)\hat{\Sigma}_n(\hat{\pi}_n^m)r_\psi(\hat{\pi}_n^m)']^{1/2}} \qquad (1.58)$$

$$T^m_{\pi,n} = T^\pi(\hat{\pi}_n^m) = \frac{||\tau_\beta(\hat{\pi}_n^m)||(r(\hat{\psi}_n(\hat{\pi}_n^m),\hat{\pi}_n^m)-r(\hat{\psi}_n(\hat{\pi}_n^m),\pi_0))}{[r_\pi(\hat{\pi}_n^m)\hat{\Sigma}_n(\hat{\pi}_n^m)r_\pi(\hat{\pi}_n^m)']^{1/2}} \qquad (1.59)$$

The semi-strong identification case t-statistic is simpler because we do not need to simulate the random paths of a stochastic process.

$$T^{m}_{\theta,n}(\hat{\theta}_n) = \frac{r_\theta(\hat{\theta}_n)\tau^{m}_{\theta,n}(\hat{\theta}_n)}{[r_\theta(\hat{\theta}_n)\hat{\Sigma}_n(\hat{\theta}_n)r_\theta(\hat{\theta}_n)']^{1/2}} \tag{1.60}$$

We consider three null hypotheses in this simulation exercise. The first null hypothesis sets the parameters equal to their true values. The second and third consider the false null hypotheses of parameters within one and three standard deviations of the true value. The false null hypotheses are important to analyze the loss of power of the robust t-test compared to the standard t-test.

$$\begin{aligned} &H^{\beta}_{0,1}: \beta = \beta_n && H^{\pi}_{0,1}: \pi = \pi_0 \\ &H^{\beta}_{0,2}: \beta = \beta_n + \sigma_\beta && H^{\pi}_{0,2}: \pi = \pi_0 + \sigma_\pi \\ &H^{\beta}_{0,3}: \beta = \beta_n + 3\sigma_\beta && H^{\pi}_{0,3}: \pi = \pi_0 + 3\sigma_\pi \end{aligned} \tag{1.61}$$

The robust sample t-statistic is constructed by

$$T_n = \frac{\sqrt{n}(r(\hat{\theta}_n) - v)}{[r_\theta(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)r_\theta(\hat{\theta}_n)']^{1/2}} \tag{1.62}$$

while the standard t-statistic takes the form

$$T^{s}_n = \frac{\sqrt{n}(r(\hat{\theta}_n) - v)}{[r_\theta(\hat{\theta}_n)\hat{\Sigma}_n(\hat{\theta}_n)r_\theta(\hat{\theta}_n)']^{1/2}} \tag{1.63}$$

The critical values of the bootstrapped t-statistic are computed using order statistics. Let $\{T^{m}_{a,n}(\pi)\}^{m}_{j=1}$ be a sequence of independent draws of the t-statistic with $a = \psi, \pi, \theta$. Denote the order statistics by $T^{m}_{a,n,[1]} \leq T^{m}_{a,n,[2]}, ...$, etc. The LF and ICS_1 critical values are computed using $c^{a,m}_{n,1-\alpha/2} = \inf\{c \geq 0 : \mathbb{P}(T^{m}_n \leq c) \geq 1 - \alpha/2\}$ and $c^{a,m}_{n,\alpha/2} = \inf\{c \geq 0 : \mathbb{P}(T^{m}_n \leq c) \geq \alpha/2\}$, as we use two tailed critical values. To construct the ICS_1 critical value, we use $\kappa_n = (\ln(n))^{1/2}$, as suggested by Andrews and Cheng (2012).

All critical values of the t-statistic are simulated for the asymptotic approximation t-tests of Andrews and Cheng (2012), the robust wild bootstrap t-test and the standard t-test. For ease of comparison, the p-values of the tests are compared using the distribution with known nuisance parameters and the distribution with unknown nuisance parameters.

The results in Tables 1.2 and 1.3 show that the bootstrapped critical values behave better than the asymptotic approximation critical values, particularly in small samples, but the difference is small. The benefits of bootstrapping are overshadowed by the loss in size and power resulting from constructing a grid of nuisance parameters. The loss of power of the robust t-test is small compared to the important gains in size. As we expected, the standard t-test works well only when the data generating process is strongly identified. When the process is weakly identified the size distortions are significant leading, to p-values of 0.30 to 0.40 for asymptotic sizes of 0.10.

Figure 1.1 and Figure 1.2 compare the large sample asymptotic distribution of the t-statistic under strong and weak identification. The figures show that there are large size distortions of the parameters $\hat{\zeta}_n$ and $\hat{\beta}_n$ when weak identification is present, which would lead to over-rejection of the null hypothesis when the standard t-statistic is employed. It is worth noticing that the rejection rates of the t-statistic of $\hat{\zeta}_n$ can be severely large, especially when errors have a $t(4)$ distribution, which is surprising because ζ is always strongly identified.

1.6 Conclusions

In this paper, we introduce an asymptotically valid, wild bootstrap t-test robust to all identification categories, in the sense of Andrews and Cheng (2012). The wild bootstrap is developed using a multiplier bootstrap at the point of lack of identification, instead of around the true parameter. The robust t-test properties include the asymptotic refinements of the bootstrap and the simplicity to estimate when the asymptotic Gaussian process has unknown form and/or is high dimensional. The simulations affirm that the wild bootstrap performs better than the finite sample counterpart, while it exposes the large-size distortions of the standard t-test when weak identification is present.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 250$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.006	0.059	0.005	0.017	0.106	0.107	0.106	0.107	0.106	0.074	0.087	0.000	0.006	0.093	0.089	0.093	0.089	0.093
5%	0.004	0.027	0.003	0.008	0.063	0.067	0.063	0.067	0.063	0.035	0.043	0.000	0.000	0.045	0.044	0.045	0.044	0.045
1%	0.001	0.005	0.001	0.002	0.010	0.017	0.010	0.017	0.010	0.010	0.013	0.000	0.000	0.011	0.014	0.011	0.014	0.011
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.032	0.098	0.024	0.076	0.259	0.259	0.259	0.259	0.259	0.233	0.260	0.001	0.023	0.271	0.266	0.271	0.266	0.271
5%	0.017	0.054	0.010	0.046	0.169	0.171	0.169	0.171	0.169	0.149	0.167	0.000	0.008	0.174	0.170	0.174	0.170	0.174
1%	0.005	0.020	0.002	0.014	0.060	0.071	0.060	0.071	0.060	0.047	0.059	0.000	0.001	0.062	0.063	0.062	0.063	0.062
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.495	0.785	0.435	0.754	0.930	0.929	0.930	0.929	0.930	0.900	0.901	0.098	0.467	0.913	0.904	0.913	0.904	0.913
5%	0.336	0.692	0.258	0.628	0.875	0.883	0.875	0.883	0.875	0.823	0.842	0.029	0.319	0.840	0.844	0.840	0.844	0.840
1%	0.128	0.473	0.060	0.374	0.715	0.719	0.715	0.719	0.715	0.617	0.643	0.001	0.082	0.648	0.658	0.648	0.658	0.648
Weakly identified $n = 250$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.098	0.110	0.054	0.056	0.125	0.124	0.115	0.117	0.276	0.056	0.060	0.000	0.000	0.098	0.098	0.028	0.027	0.062
5%	0.056	0.057	0.026	0.017	0.073	0.071	0.064	0.054	0.155	0.018	0.021	0.000	0.000	0.039	0.039	0.017	0.019	0.018
1%	0.007	0.005	0.002	0.003	0.009	0.011	0.007	0.010	0.046	0.002	0.003	0.000	0.000	0.002	0.004	0.002	0.004	0.002
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.305	0.303	0.294	0.293	0.306	0.303	0.295	0.294	0.385	0.425	0.430	0.113	0.125	0.482	0.488	0.142	0.151	0.430
5%	0.276	0.269	0.253	0.245	0.276	0.270	0.253	0.246	0.344	0.322	0.323	0.065	0.062	0.405	0.392	0.095	0.089	0.327
1%	0.168	0.171	0.127	0.125	0.168	0.171	0.127	0.125	0.279	0.189	0.193	0.014	0.013	0.283	0.281	0.053	0.053	0.193
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.856	0.861	0.847	0.842	0.895	0.901	0.886	0.884	0.974	0.801	0.807	0.616	0.606	0.829	0.832	0.619	0.609	0.805
5%	0.805	0.809	0.779	0.775	0.822	0.831	0.797	0.800	0.925	0.761	0.756	0.554	0.547	0.795	0.791	0.557	0.550	0.761
1%	0.679	0.680	0.624	0.623	0.686	0.688	0.632	0.633	0.826	0.680	0.681	0.427	0.426	0.731	0.722	0.432	0.431	0.683
Non-identified $n = 250$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.110	0.099	0.033	0.038	0.112	0.099	0.053	0.054	0.242	0.053	0.059	0.000	0.000	0.098	0.094	0.025	0.024	0.058
5%	0.054	0.050	0.013	0.015	0.065	0.058	0.053	0.053	0.124	0.022	0.024	0.000	0.000	0.040	0.045	0.019	0.017	0.025
1%	0.008	0.013	0.002	0.002	0.028	0.031	0.028	0.031	0.028	0.001	0.001	0.000	0.000	0.005	0.004	0.001	0.001	0.001
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.418	0.411	0.299	0.287	0.423	0.414	0.303	0.291	0.482	0.349	0.349	0.083	0.088	0.425	0.418	0.095	0.097	0.350
5%	0.341	0.332	0.196	0.199	0.342	0.333	0.198	0.201	0.444	0.276	0.276	0.044	0.045	0.363	0.352	0.060	0.055	0.279
1%	0.192	0.176	0.076	0.071	0.198	0.187	0.080	0.071	0.266	0.158	0.158	0.008	0.012	0.235	0.240	0.031	0.031	0.158
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.933	0.933	0.859	0.863	0.938	0.937	0.866	0.868	0.965	0.796	0.801	0.562	0.549	0.842	0.841	0.567	0.553	0.798
5%	0.888	0.891	0.803	0.798	0.889	0.894	0.806	0.802	0.942	0.734	0.741	0.493	0.477	0.808	0.804	0.498	0.483	0.734
1%	0.784	0.784	0.663	0.663	0.787	0.796	0.665	0.668	0.844	0.629	0.623	0.331	0.332	0.715	0.712	0.340	0.340	0.631

Table 1.2: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with Normal(0,1) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 250$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1}: \beta = \beta_n$									$H_{0,1}: \pi = \pi_0$								
10%	0.093	0.082	0.061	0.054	0.136	0.143	0.136	0.143	0.136	0.106	0.108	0.000	0.000	0.108	0.117	0.108	0.117	0.108
5%	0.058	0.050	0.039	0.035	0.074	0.095	0.074	0.095	0.074	0.063	0.064	0.000	0.000	0.067	0.069	0.067	0.069	0.067
1%	0.019	0.014	0.010	0.010	0.026	0.048	0.026	0.048	0.026	0.022	0.029	0.000	0.000	0.024	0.037	0.024	0.037	0.024
	$H_{0,2}: \beta = \beta_n + \sigma_\beta$									$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.223	0.226	0.205	0.209	0.339	0.314	0.339	0.314	0.339	0.334	0.336	0.009	0.010	0.337	0.339	0.337	0.339	0.337
5%	0.170	0.176	0.138	0.162	0.236	0.208	0.236	0.208	0.236	0.239	0.247	0.008	0.006	0.244	0.254	0.244	0.254	0.244
1%	0.074	0.074	0.059	0.069	0.102	0.099	0.102	0.099	0.102	0.103	0.112	0.000	0.000	0.107	0.127	0.107	0.127	0.107
	$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$									$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.892	0.903	0.885	0.895	0.933	0.920	0.933	0.920	0.933	0.920	0.922	0.495	0.465	0.923	0.924	0.923	0.924	0.923
5%	0.853	0.864	0.829	0.853	0.896	0.873	0.896	0.873	0.896	0.890	0.886	0.345	0.298	0.892	0.889	0.892	0.889	0.892
1%	0.742	0.742	0.682	0.725	0.803	0.760	0.803	0.760	0.803	0.787	0.791	0.110	0.079	0.790	0.796	0.790	0.796	0.790
Weakly identified $n = 250$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1}: \beta = \beta_n$									$H_{0,1}: \pi = \pi_0$								
10%	0.124	0.103	0.067	0.061	0.140	0.128	0.122	0.120	0.264	0.066	0.067	0.003	0.003	0.093	0.094	0.041	0.041	0.082
5%	0.077	0.059	0.034	0.031	0.081	0.086	0.071	0.076	0.168	0.037	0.041	0.000	0.001	0.047	0.052	0.029	0.030	0.047
1%	0.025	0.020	0.011	0.011	0.028	0.045	0.025	0.042	0.051	0.010	0.010	0.000	0.000	0.012	0.021	0.012	0.019	0.013
	$H_{0,2}: \beta = \beta_n + \sigma_\beta$									$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.327	0.331	0.319	0.317	0.330	0.337	0.324	0.325	0.432	0.420	0.415	0.102	0.101	0.481	0.478	0.122	0.122	0.431
5%	0.293	0.290	0.265	0.269	0.293	0.295	0.267	0.275	0.380	0.319	0.329	0.052	0.057	0.396	0.396	0.081	0.086	0.346
1%	0.190	0.196	0.144	0.154	0.190	0.199	0.144	0.157	0.298	0.170	0.172	0.024	0.028	0.238	0.247	0.056	0.059	0.185
	$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$									$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.909	0.920	0.894	0.910	0.924	0.929	0.911	0.922	0.962	0.815	0.810	0.609	0.605	0.842	0.838	0.620	0.616	0.826
5%	0.875	0.890	0.851	0.870	0.888	0.899	0.870	0.881	0.939	0.763	0.767	0.539	0.530	0.795	0.803	0.548	0.540	0.776
1%	0.784	0.803	0.739	0.765	0.798	0.811	0.758	0.780	0.871	0.664	0.668	0.397	0.421	0.710	0.717	0.405	0.428	0.682
Non-identified $n = 250$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1}: \beta = \beta_n$									$H_{0,1}: \pi = \pi_0$								
10%	0.147	0.121	0.064	0.052	0.147	0.121	0.080	0.076	0.277	0.062	0.065	0.000	0.001	0.090	0.085	0.028	0.027	0.084
5%	0.084	0.069	0.032	0.028	0.087	0.077	0.074	0.074	0.155	0.027	0.029	0.000	0.000	0.042	0.042	0.018	0.017	0.040
1%	0.027	0.019	0.005	0.005	0.046	0.065	0.046	0.065	0.046	0.002	0.004	0.000	0.000	0.005	0.005	0.002	0.004	0.008
	$H_{0,2}: \beta = \beta_n + \sigma_\beta$									$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.460	0.456	0.377	0.376	0.463	0.464	0.381	0.384	0.516	0.378	0.375	0.102	0.098	0.442	0.447	0.116	0.112	0.391
5%	0.404	0.405	0.285	0.290	0.405	0.410	0.286	0.293	0.474	0.296	0.295	0.054	0.051	0.373	0.376	0.071	0.069	0.308
1%	0.257	0.269	0.135	0.132	0.261	0.275	0.136	0.134	0.336	0.175	0.177	0.009	0.012	0.247	0.251	0.042	0.042	0.189
	$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$									$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.939	0.945	0.909	0.920	0.943	0.950	0.917	0.928	0.960	0.793	0.797	0.567	0.576	0.837	0.839	0.574	0.583	0.801
5%	0.918	0.928	0.868	0.876	0.919	0.931	0.874	0.882	0.945	0.754	0.754	0.499	0.499	0.797	0.798	0.508	0.508	0.762
1%	0.853	0.867	0.780	0.801	0.861	0.880	0.793	0.813	0.887	0.646	0.658	0.345	0.367	0.716	0.718	0.356	0.375	0.657

Table 1.3: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

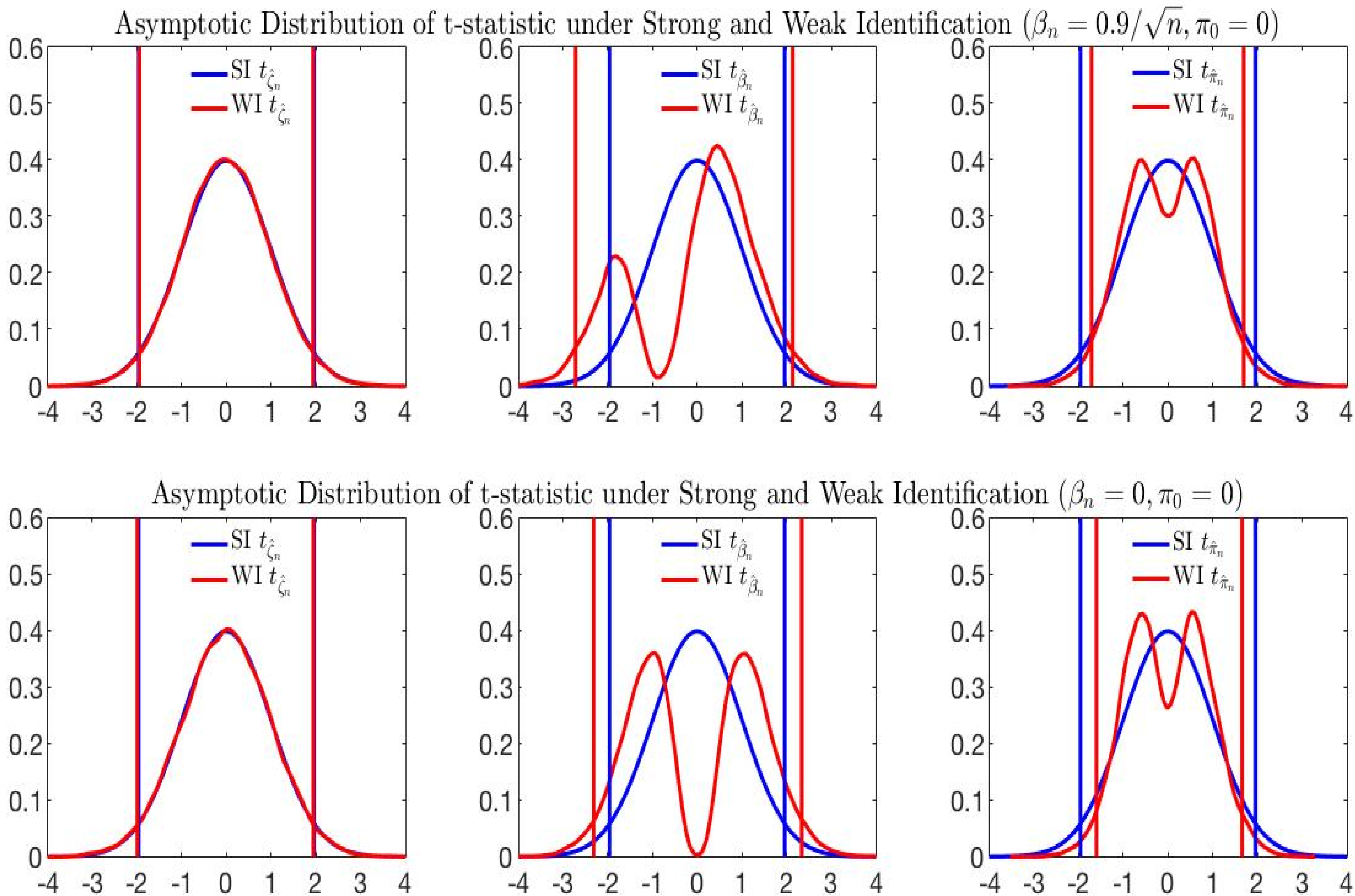

Figure 1.1: Asymptotic distribution of the t-statistic for the Smoothing Exponential model with Normal(0,1) errors. The blue line denotes the asymptotic distribution of the t-statistic under strong identification (standard normal), while the red line denotes the asymptotic distribution under weak identification (see Proposition 1.3.3). Vertical lines illustrate the 95% rejection critical values for each distribution. Each column illustrates the distribution for $\hat{\zeta}_n$, $\hat{\beta}_n$ and $\hat{\pi}_n$ respectively, while the rows illustrate the change in the parameter b. Sample size $n = 500$, simulations $S = 1000$, bootstrapped samples $M_n = 500$.

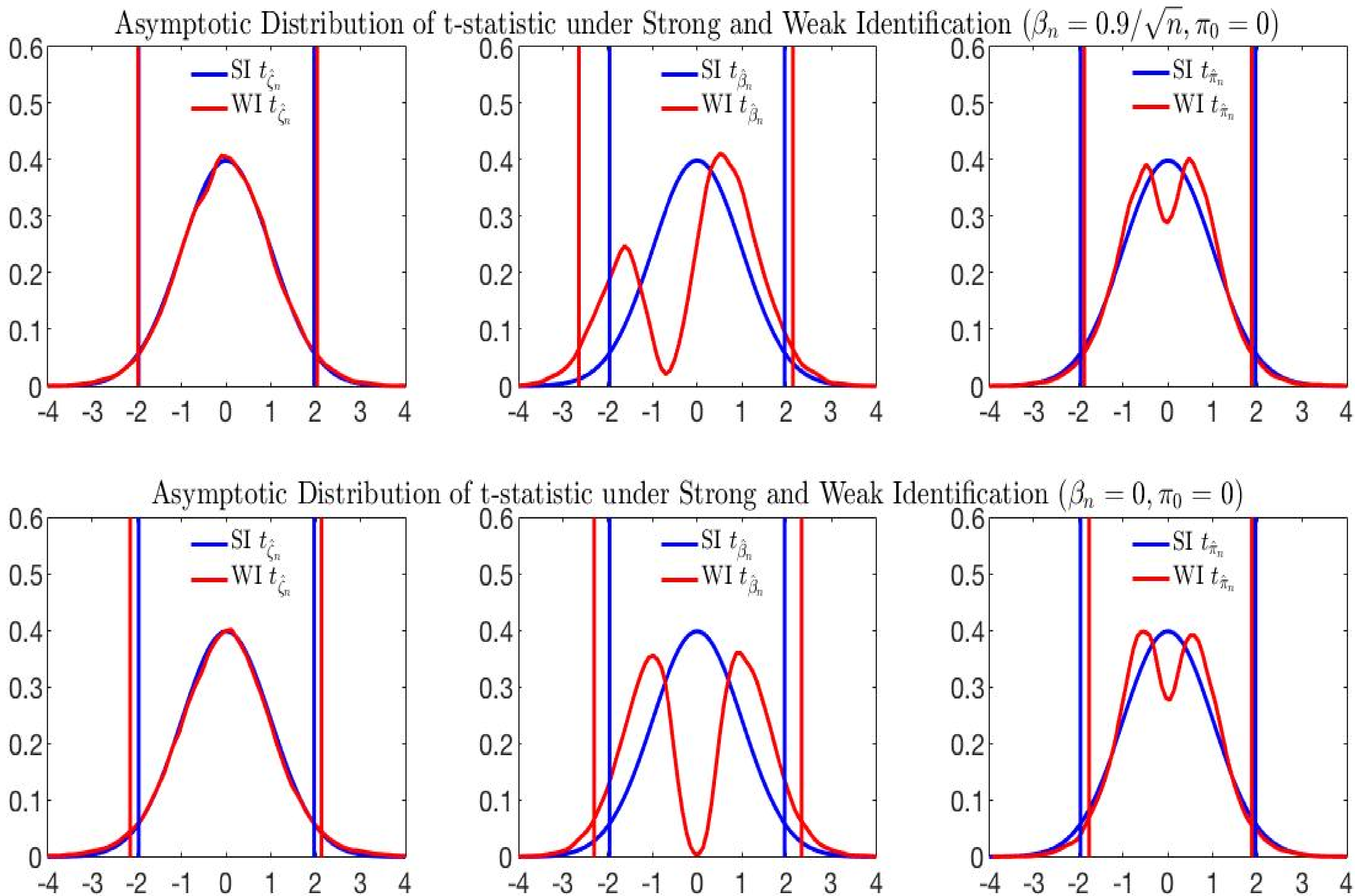

Figure 1.2: Asymptotic distribution of the t-statistic for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. The blue line denotes the asymptotic distribution of the t-statistic under strong identification (standard normal), while the red line denotes the asymptotic distribution under weak identification (see Proposition 1.3.3). Vertical lines illustrate the 95% rejection critical values for each distribution. Each column illustrates the distribution for $\hat{\zeta}_n$, $\hat{\beta}_n$ and $\hat{\pi}_n$ respectively, while the rows illustrate the change in the parameter b. Sample size $n = 500$, simulations $S = 1000$, bootstrapped samples $M_n = 500$.

CHAPTER 2

PARAMETRIC WILD BOOTSTRAP INFERENCE WITH WEAKLY IDENTIFIED PARAMETERS

2.1 Introduction

Identification is one of the most important assumptions in econometric modeling. Being able to correctly specify the model is essential to obtain estimates that are informative. Identification is usually assumed but the validity of this assumption is rarely tested. Implementing standard hypothesis tests on models that are unidentified would usually lead to false positives. This paper expands the parameter identification literature by proposing a parametric bootstrap to conduct hypothesis testing for any identification category.

Heuristically, the parameters in an econometric model are identified if a unique value of the parameter can be obtained with a sufficient amount of data. If this is the case, we say that the parameter is strongly identified. In contrast, when the parameter value does not have enough signal to noise ratio and a unique value cannot be obtained regardless of the amount of data at our disposal, we say that the parameter is weakly identified.

In principle, the identification category of parameters is unknown. Not surprisingly, the asymptotic distribution of the estimators, as well as their convergence rate, depends on being either strongly or weakly identified. In consequence, assuming strong identification in models where weak identification is present leads to erroneous statistical testing. In the case of the t-statistic, the Hessian converges to a singular matrix when parameters are weakly identified. The t-statistic is growing without bound which leads to an eventual rejection of the null hypothesis, i.e. false positives. This is conceptually similar to the spurious regression phenomenon introduced by Granger and Newbold (1974) because the t-statistic of a regression constructed with (drift-less) unit root variables does not have a standard normal distribution and is growing at rate $O_p(n^{1/2})$.

In Chapter 1 we propose a wild bootstrapped robust t-test that can be employed in models with strongly or weakly identified parameters. The authors derive a bootstrap method to generate the asymptotic distribution and generate robust testing by combining the critical values. In this paper, we propose an alternative bootstrap method to construct the t-test. The parametric bootstrap is developed for models in which consistent estimators of some parameters are not available.

One of the most important properties of the parametric bootstrap is that it reduces the number of nuisance parameters of the test statistic under weak identification. When a parameter in a model cannot be consistently estimated (i.e., it is weakly identified), the t-test proposed by Andrews and Cheng (2012) and the wild bootstrapped t-test proposed in Chapter 1 include nuisance parameters under the null hypothesis. To be able to construct the statistic in practice, the usual approach is to generate a grid of possible nuisance parameters and obtain the supremum of the critical values as in Davies (1977, 1987). This approach would lead to test statistics with correct level but not correct size. If the rejection rate of the test is 5%, the statistic rejects at a 5% level or lower. Unfortunately, for a large enough grid, some test statistics lose all statistical power. The simulation study of Chapter 1 shows that the t-statistic with respect to the weakly identified parameter π has no statistical power for a large enough grid as the distribution is centered around a nuisance parameter. Reducing the number of nuisance parameters in the limit distribution leads to statistical tests with better performance.

This paper proposes a parametric bootstrap influenced by the residual bootstrap employed in strongly identified regression models e.g. Freedman (1981); Mammen (1993); Liu et al. (1988); Horowitz (2001); Davidson and Flachaire (2008). The residual bootstrap is used to generate bootstrap samples of the original data, by resampling residuals and generating new draws imposing the regression model. When parameters are weakly identified, this procedure cannot be employed as residuals depend on estimators that are inconsistent. Nonetheless, our parametric bootstrap method generates valid bootstrap samples by resampling residuals over a grid of potential values for the weakly identified parameters. More specifically, the parametric bootstrap under weak identification follows these steps: 1) obtain residuals centered at the point of lack of identification which are valid to perform a bootstrap, 2) generate bootstrapped residual draws using a Wild bootstrap multiplicative approach as in Wu (1986); Liu et al. (1988); Shao (2010), 3) obtain bootstrapped data draws by imposing the null hypothesis over a grid of parameters, and 4) minimize the objective function with the bootstrapped samples to retrieve bootstrapped estimators. The main difference of our parametric bootstrap hinges on step 3, in which the bootstrap is performed along a grid of potentially weakly identified parameters, from which a consistent estimator cannot be obtained.

Andrews and Cheng (2012, 2014); Andrews and Mikusheva (2015) introduce a unified framework of identification categories in which the definitions of identification depend on the true values that parameters take along the parameter space. In other words, for some values along the parameter space, the parameters are strongly identified while for other values the parameters are weakly identified. The authors introduce four different identification categories that comprise all possible cases. As all identification categories can be grouped into a t-test with two distributional cases, we regroup the identification categories of Andrews and Cheng (2012) into two groups: weak and strong identification. For simplicity, as the four distributions can be grouped into two cases that enclose all other cases. Specifically, we refer to weak identification for the weak and non-identification categories of Andrews and Cheng (2012), while we refer to strong identification the semi-strong and strong identification categories.

One of the most important differences between the identification categories is that under weak identification, the parameter $\hat{\pi}_n$ converges to a random variable. The random outcome of π will determine the value of the other estimator $\hat{\psi}_n$. The distribution of these parameters is non-standard and usually very different from a normal distribution. To obtain critical values of these distributions we can rely on simulation methods that depend on nuisance parameters. The parametric bootstrap method will simplify the construction of these distributions, avoiding the simulation of these non-standard distributions.

The bootstrap method introduced in this paper relies on bootstrapping residuals and generating valid draws of the model at hand. With validity, we mean that the bootstrapped distribution is equivalent to the asymptotic distribution of the statistic generated with the true data generating process. In this paper we bootstrap the residuals using a wild bootstrap approach (Wu 1986; Liu et al. 1988; Shao 2010) as we work with heteroscedastic and dependent data. The wild bootstrap has important advantageous properties: 1) using a pivotal statistic it leads to an asymptotic refinement in small samples (Mammen 1993; Horowitz 2001; Härdle et al. 2003); 2) the wild bootstrap is useful to approximate asymmetric distributions with fat tails, and 3) the multiplicative form of the wild bootstrap extends easily to multivariate settings. The resampling residual method proposed by Efron et al. (1979) can be applied if the data are independent and identically distributed. We do not pursue this approach because economic and financial data is usually heteroscedastic and dependent.

Extending bootstrap methods to models where weak identification is present is not a trivial extension. The uniformity assumptions required for any bootstrap method, are not satisfied in this setting. As weak identification is present for some values of the parameter space, the distribution of the estimators will be different according to the true value the parameter takes. In particular, the delta method cannot be used for the weakly identified parameters as the objective function does not have a clean minimum (see Assumption J). The non-uniformity of the bootstrap is comparable to the case studied by the integrated conditional test of Bierens and Ploberger (1997). This lead to technical difficulties that are solved using stochastic expansions centered at the point of lack of identification (Andrews and Cheng 2012; Andrews and Mikusheva 2015; Andrews and Cheng 2014). Consequently, if we wish to replicate the distributions under weak and strong identification, we must perform two bootstrap procedures to construct the two distributions separately. The bootstrap procedures for both identification categories are introduced in Section 2.4.

The most relevant contribution of this paper is the ability to construct t-tests that do not depend on nuisance parameters. Hypothesis testing with nuisance parameters dates back to Chernoff and Zacks (1964). Davies (1977, 1987) studies the consequences of having nuisance parameters under the alternative hypothesis and proposes using the sup-Likelihood ratio test over a grid of potential nuisance parameters. This approach leads to a test with the correct level, but not the correct size as the correct distribution can only be constructed if the nuisance parameter is known. In other words, if the test has a 5% size, the sup test would asymptotically have an empirical size of 5% or less. Stinchcombe and White (1998); Andrews (1994) develop smoothed test statistics with nuisance parameters under the alternative hypothesis. Hansen (1996) derives a sup-test to eliminate the nuisance parameter under the null hypothesis, constructing the asymptotic distribution using a local-to-null reparametrization. Contrary to these studies, this paper proposes a bootstrap method valid for all identification categories. To the best of our knowledge, this is the first research paper that proposes a parametric bootstrap method to construct robust critical values using the identification category definitions of Andrews and Cheng (2012).

The paper has the following structure. Section 2.2 introduces the notation of the model and the identification categories framework. Section 2.3 introduces the assumptions used to prove the main results of the paper. Section 2.4 introduces the parametric bootstrap for the strong and weak identification categories, while the following section discusses how to impose the null hypothesis in the parametric bootstrap. Section 2.6 derives the bootstrapped distribution of the t-statistic employed in the simulation exercise. Simulation results are analyzed in Section 2.7. The last section concludes the paper.

2.2 Model Setup

Identification in this paper refers to the ability to extract the true value of the parameters in a model with a sufficient amount of data. In circumstances where a consistent estimator of the parameter is not available with an infinite amount of data, we say that the parameter is weakly identified. On the other hand, when a consistent estimator of the parameter is available, we say that the parameter is strongly identified. The identification of a parameter will depend on the value that it takes in the parameter space. Before introducing identification categories, we present two examples of identification.

Example 1. Nonlinear Regression Model

$$y_t = \zeta_0' X_{1,t} + \beta_0' h(X_{2,t}, \pi_0) + \epsilon_t \tag{2.1}$$

The identification of π_0 depends on β_0 being non-zero. If $\beta_0 = 0$, the parameter π_0 is weakly identified. The identification of ζ_0 does not depend on the identification of either β_0 or π_0; it is always strongly identified.

Example 2. MIDAS Regression

Consider a mixed data sampling process $\{y_t, X_{t/m}^{(m)}\}$, where y_t is observable at times $t = 1, ..., n$ and $X_{t/m}^{(m)} = (1, X_{1,t/m}^{(m)}, ..., X_{p,t/m}^{(m)})$, where m is the number of high frequency lags used in the temporal aggregation of $X_{t/m}$. The MIDAS regression sets the higher frequency variable on the right-hand side of a regression equation

$$y_t = \beta_0' X_t(\pi_0) + \epsilon_t \tag{2.2}$$

where $X_t(\pi_0)$ is a nonlinear function that maps the high frequency data into the lower frequency data,

$$X_{k,t}^{(m)}(\pi_{0,k}) = \sum_{j=1}^{m} w_{j,k}(\pi_{0,k}) L^{j/m} X_{k,t/m}^{(m)} \tag{2.3}$$

Where L denotes the lag operator. If $\beta_0 = 0$, then π_0 is not identified, and the weight function $w_{j,k}(\pi_{0,k})$ can potentially take any value. Other examples of models that suffer from weak identification for some values in the parameter space include: ARMA models (Andrews and Cheng 2012), Maximum Likelihood Estimation (Andrews and Cheng 2013), Generalized Method of Moments (Andrews and Cheng 2014), Dynamic General Equilibrium models (Guerron-Quintana et al. 2013; Andrews and Mikusheva 2015), MIDAS regression (Ghysels et al. 2016), Smooth Transition Autoregressive models (Andrews and Cheng 2013), Probit models (Andrews and Cheng 2014), Regime Switching models (Chen et al. 2016), among others.

Ghysels et al. (2005); Ghysels, Plazzi, and Valkanov (2016) propose a MIDAS model to analyze the risk-return trade-off at a monthly or quarterly frequency using a volatility estimator constructed with daily excess returns. The authors conclude that the estimators of the MIDAS model are significant and that the risk-return trade-off relationship can be captured using the data-driven polynomial of the MIDAS aggregation function. The hypothesis tests constructed by Ghysels et al. (2005, 2016) ignore the presence of weak identification. The null hypotheses $\beta = 0$ or $\pi = 0$ cannot be tested using the standard t-statistic because we cannot separately identify both parameters under the null and in consequence, would not be normally distributed. Chapter 3 analyses if the parameters of the MIDAS model studied by Ghysels et al. (2005, 2016) maintain their statistical significance when critical values robust to weak identification are used. Using a wild bootstrapped t-test robust, Chapter 3 concludes that the MIDAS representation of the risk-return trade-off is not significant for all samples studied. The author suggests exploring other possibilities to model the risk-return trade-off using a different stochastic discount factor (e.g. the utility function introduced by Epstein and Zin (2013)).

The cases where weak identification is present can be hard to visualize using general extremum estimator notation. For this reason, the results of the paper the non-linear model in which weak identification is easy to portray. Let $\{W_t\}_{t=1}^{n} = \{y_t, X_t\}_{t=1}^{n}$ be observable data, and $\theta = (\zeta', \beta', \pi')'$ denote a vector of parameters. We divide the parameter vector θ into these three groups because each of them describes a different identification category. The parameter π characterizes the parameter that is potentially weakly identified. The identification category of π is determined by the value of β, which is itself always strongly identified. Moreover, the parameter ζ is always strongly identified, and its identification does not affect the identification of other parameters. The non-linear model takes the form:

$$\epsilon_t(\theta) = y_t - \zeta' X_{1,t} - \beta' h(X_{2,t}, \pi) \tag{2.4}$$

where $X_{1,t}$ and $X_{2,t}$ denote elements of the $X_t \in \mathbb{R}^d$ matrix of explanatory variables, $y_t \in \mathbb{R}$

is the dependent variable and $h(\cdot)$ is a non-random function. We define the parameters as elements of compact sets $\zeta \in \mathcal{Z} \subset \mathbb{R}^{\lceil_\zeta}$, $\beta \in \mathcal{B} \subset \mathbb{R}^{\lceil_\beta}$ and $\pi \in \Pi \subset \mathbb{R}^{d_\pi}$. Clearly, the parameters of the non-linear model are always strongly identified except when $\beta = 0$. This value is particularly relevant as we usually want to test if the parameters are significant, which implies testing $\beta = 0$. Under the null, $H_0 : \beta = 0$, π is not identified and inference is non-standard.

The estimator $\hat{\theta}_n$ minimizes an objective function $Q_n(\theta)$. For simplicity we consider the least squares objective function.

$$\hat{\theta}_n = \inf_{\theta \in \Theta} Q_n(W_t, \theta) = \inf_{\theta \in \Theta} \frac{1}{2n} \sum_{t=1}^{n} \epsilon_t^2(\theta) \tag{2.5}$$

The objective function does not depend on π under weak identification. For a smaller and smaller β, the signal that comes from π is diminished more and more. If the signal is smaller than the random noise, we say that the parameter is weakly identified (see next subsection). To simplify the notation, for the rest of the paper we partition the parameters into two groups according to their identification category.

$$\theta = (\zeta', \beta', \pi')' = (\psi', \pi')' \tag{2.6}$$

The parameter ψ denotes the strongly identified parameters, whereas π denotes parameters that are weakly identified. In the next section we argue that the strongly identified parameters ψ can be consistently estimated, while the weakly identified parameters π cannot be consistently estimated.

2.2.1 Drifting Sequences of Distributions

To determine whether the parameters are weakly or strongly identified, we consider drifting sequences of distributions as in Staiger and Stock (1994) and Stock and Wright (2000). The drifting sequences of distributions determine the asymptotic behaviour of the estimators for distinct identification categories, according to the speed in which parameters converge to their true value. Suppose the sequences of true parameters are defined by $\theta_n = (\beta_n', \zeta_n', \pi_n')'$ for $n \geq 1$, converging to the limit true parameters $\theta_0 = (\beta_0', \zeta_0', \pi_0')'$. Table 2.1 illustrates the definitions of identification categories following Andrews and Cheng (2012) which are derived using drifting sequences of parameters.

Table 2.1: Identification categories

Category	$\{\beta_n\}$ sequence	Identification Property of π
I(a)	$\beta_n = 0 \ \forall n \geq 1$	Unidentified
I(b)	$\beta_n \neq 0$ and $\sqrt{n}\beta_n \to b \in \mathbb{R}^{d_\beta}$	Weakly identified
II	$\beta_n \to 0$ and $\sqrt{\beta_n} \to \infty$	Semi-strongly identified
III	$\beta_n \to \beta_0 \neq 0$	Strongly Identified

Heuristically, the speed at which β_n converges to zero will determine if the parameter π is strongly or weakly identified. If β_n is converging to zero at a rate faster or equal to $\sqrt{n}$, then π is weakly identified. On the other hand, if β_n is converging to zero at a slower rate than $\sqrt{n}$ (or if it is not converging to zero at all), then π is strongly identified. The following table exemplifies this argument.

Table 2.2: Alternative Table of Identification categories

$\{\beta_n\}$ sequence	Category	Identification Property of π
If $\beta_n \to 0$ and $\beta_n = 0$, $\forall n \geq 1$,	I(a)	Unidentified
If $\beta_n \to 0$ and $\beta_n = O(n^{-\alpha})$ with $\alpha \geq 1/2$	I(b)	Weakly Identified
If $\beta_n \to 0$ and $\beta_n = O(n^{-\alpha})$ with $\alpha \in [0, 1/2)$	II	Semi-strongly Identified
If $\beta_n \to \beta_0 \neq 0$	III	Strongly Identified

In the following sections, we argue that the asymptotic distributions of the t-test of all four categories can be simplified into two cases, Category I(a)-I(b) and Category II-III. To simplify the terminology we will refer to "weak identification" to discuss categories $I(a)$ and $I(b)$, while we refer to "strong identification" when we discuss categories II and III.

The non-linear model studied in this paper defines the true error process for each sequence of drifting parameters, which we denote by $\epsilon_t(\theta_n)$, $\theta_n = (\zeta_n', \beta_n', \pi_n')'$

$$\epsilon_t(\theta_n) = y_t - \zeta_n' X_{1,t} - \beta_n' h(X_{2,t}, \pi_n) \tag{2.7}$$

The limit error process is denoted by $\epsilon_t(\theta_0) = y_t - \zeta_0' X_{1,t} - \beta_0' h(X_{2,t}, \pi_0)$, where $\theta_0 = (\zeta_0, \beta_0, \pi_0)$, while the finite sample and limit variance of errors are defined by $\sigma_t^2(\theta_n)$ and $\sigma_t^2(\theta_0)$ respectively. It is worth mentioning that the exogeneity assumption is not the subject of our study. The exogeneity assumption $\mathbb{E}(\epsilon_t(\theta_n)|X_t) = 0$ is always satisfied, although not for a unique value of θ_n. Identification in this paper is complementary to the weak instruments literature in which $\mathbb{E}(\epsilon_t(\theta_n)|X_t) \neq 0$.

2.3 Assumptions

In this section, we introduce the assumptions required for the implementation of the parametric bootstrap. The parametric bootstrap is constructed to derive the distribution of the estimators and the t-statistic in Section 2.6.

Assumption J *Identification of data generating process*

(i) $\epsilon_t(\theta_n)$ *is* L_p *bounded for* $p = 4 + \iota$ *for small* ι.

(ii) $\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n)|X_t) = 0$ *a.s.*

(iii) $Var_{\theta_n}(\epsilon_t(\theta_n)|X_t) = \Gamma_0(\theta_n) + 2\sum_{j=1}^{\infty} \Gamma_j(\theta_n)$ *a.s.*

(iv) Under weak identification $\mathbb{E}_{\theta_n}(\epsilon_t(\psi, \pi) d_{\psi,t}(\pi)) = 0$ *for unique* $\psi_n = (\zeta_n', \beta_n')'$ *in the interior of* $\Psi^*(\pi)$ *and under strong identification* $\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n) d_{\theta,t}(\theta_n)) = 0$ *for unique* $\theta_n = (\zeta_n', \beta_n', \pi_n')'$ *in the interior of* Θ^*.

Remark J. Assumption J establishes the valid moment conditions under weak and strong identification. As there is no unique π that satisfies Assumption A(iv), the parameter is weakly identified. All random variables are L_p bounded for $p = 4 + \iota$.

Assumption K *Properties of data generating process* *(i) The random variables* $\{W_t\}_{t=1}^n = \{y_t, X_t\}_{t=1}^n$ *are strictly stationary and* $\alpha-$*mixing of size* $-r/(r-2)$ *for some* $r > 2$ *and* X_t *is* L_p *bounded for* $p = 4 + \iota$ *for small* ι*.*

(ii) The processes $\{y_t(\theta)\}$ *is Lipschitz for all* t *and all values* $\theta \in \Theta$*.*

(iii) The function $h(X_t, \pi)$ *is twice continuously differentiable in* $\pi \in \Pi$*, non-degenerate and bounded for fixed values of* X_t*. Moreover,* $h(X_t, \pi)$ *is a Borel measurable function and* L_p *bounded for* $p = 4 + \iota$ *for tiny* ι *for any fixed* $\pi \in \Pi$*.*

Remark K. Assumption K establishes the dependence and distributional properties of the data. As our focus is on stationary time series data, we assume strictly stationary and α - mixing to establish a law of large numbers and central limit theorems. The Lipschitz is a mild condition used to prove weak convergence. Differentiability and measurability of $h(X_t, \pi)$ are standard.

Assumption L *Parameter Spaces*

(i) Θ *is a compact set of* $\mathbb{R}^{d_\theta}$*, where* $d_\theta = d_\zeta + d_\beta + d_\pi$

(ii) $\Theta^* = \{(\beta^{*'}, \zeta^{*'}, \pi^{*'})' : \beta \in \mathcal{B}^*, \zeta \in \mathcal{Z}^*, \pi \in \Pi^*\}$ *is a compact set and* $\Theta^* \subset int(\Theta)$ *and* $0_{d_\beta} \in int(\mathcal{B}^*)$

(iii) $\tilde{\Pi}$ *is a compact set in* $\mathbb{R}^{d_\Pi}$ *and* $\pi_n \in \tilde{\Pi}$*,* $\forall n \geq 1$*, where* $\tilde{\Pi}$ *denotes a fine grid of elements of* π *used for the parametric bootstrap.*

Remark L. The parameter spaces are compact while the whole sequence of true parameters lies in the interior of the compact set. Condition (ii) emphasizes on true values in the interior of the parameter space to eliminate boundary effects (Andrews 1999, 2001).

Assumption M *Identification of* π

(i) Each sample path of the stochastic process $\{\xi(\pi; \theta_{0,k}, b) : \pi \in \Pi\}$ *is minimized over* Π *at a unique point denoted* $\pi^*(\theta_{0,k}, b)$ $\forall \theta_{0,k} \in \Theta$ *in some set* $A(\theta_{0,k}, b)$ *with* $\mathbb{P}_{\theta_{0,k}}(A(\theta_{0,k}, b)) = 1$ *with* $\beta_0 = 0$*,* $\pi_0 = \pi_k$ *and* $||b|| < \infty$*.*

Remark M. Assumption M establishes an identification condition when parameters cannot be estimated consistently. Weak identification implies that the signal and the noise are growing at the same rate, nonetheless, the signal of π must lead to a unique distribution function. The second condition expands the identification condition of π along the grid of Π used for the parametric bootstrap.

Assumption N *Continuity and non-singularity of variance-covariance matrix*

(i) $J(\theta;\theta_0)$ *and* $V(\theta;\theta_0)$ *are continuous in* $\theta \in \Theta$, $\forall \theta_0 \in \Theta$ *with* $\beta_0 = 0$.

(ii) $J(\pi;\theta_0)$ *and* $V(\pi;\theta_0)$ *are positive definite matrices* $\forall \pi \in \Pi, \theta_0 \in \Theta$ *with* $\beta_0 = 0$. *(that is,* $\underline{eig}(J(\pi;\theta_0))$*,*$\underline{eig}(V(\pi;\theta_0)) > 0$ *and* $\overline{eig}(J(\pi;\theta_0))$*,*$\overline{eig}(V(\pi;\theta_0)) < \infty$*, i.e. the max-min eigenvalues are finite)*

Remark N. The conditions in Assumption N guarantee that the variance and covariance matrices in t-statistic are non-singular. The normalization matrix $B(\beta)$ will play an important role to obtain a non-singular matrix in the limit.

Assumption O *Properties of the restrictions function*

(i) $r(\theta) \in \mathbb{R}$ *is continuously differentiable on* Θ

(ii) $r_\theta(\theta) \neq 0 \ \forall \theta \in \Theta$

(iii) $rank(r_\pi(\theta)) = d_\pi^*$ *for some constant* $d_\pi^* \leq min(d_r, d_\pi), \forall \theta \in \Theta_\delta = \{\theta \in \Theta : ||\beta|| < \delta\}$

Remark O. Conditions in Assumption O are standard for the construction of the t-statistic. The restrictions in the hypothesis test must be linearly independent. As we will discuss in detail in the following sections, testing with respect to ψ or π is relevant because of the asymptotic distribution of the t-statistic changes according to which parameter is being tested.

Let $p \in \mathcal{P}$ be the set of nuisance parameters. In the framework of Andrews and Cheng (2012) the set of nuisance parameters is characterized by $p = (b, \theta_0)$.

Assumption P *Properties of LF and ICS_0 critical values*

(i) The distribution function of T^{ψ} and T^{π} are continuous at $c_{1-\alpha}(p)$, $\forall p \in \mathcal{P}$. If $c^{LF}_{1-\alpha} > c_{1-\alpha}(\infty)$, $c^{LF}_{1-\alpha}$ is attained at some $p_{max} \in \mathcal{P}$.

(ii) $\kappa_n \to \infty$ and $\frac{\kappa_n}{\sqrt{n}} \to 0$

(iii) The distribution function of T^{ψ} and T^{π} are continuous at the critical values $c_{1-\alpha}$, $\forall p \in \mathcal{P}$.

Remark P. The use of Least Favourable and ICS_0 robust critical values lead to correct asymptotic size if they satisfy Assumption P. As we show in the following sections, the parametric bootstrap reduces the number of nuisance parameters of the t-statistic.

Before we introduce the next assumption, we must define the Confidence Sets of a test. Let $H_0 : r(\theta_n) = q$ and T_n be the t-statistic while $c_{n,1-\alpha}$ denotes the $1-\alpha$ critical value. The confidence sets are defined by,

$$CS_n = \{q : T_n(q) \leq c_{n,1-\alpha}(q)\} \tag{2.8}$$

Notice that the Confidence Sets are a function of the value q as well as sample size n. The Coverage Probability of a Confidence Set for $r(\theta)$ is defined by,

$$CP_n(\theta) = \mathbb{P}_{\theta}(r(\theta) \in CS_n) = \mathbb{P}_{\theta}(T_n(r(\theta) \leq c_{n,1-\alpha}(r(\theta)))) \tag{2.9}$$

For more information about confidence sets and coverage probabilities using drifting sequences of parameters see Andrews and Cheng (2012). An important measure of the t-test is the minimum null rejection acceptance as it is equivalent to the asymptotic size of the test. [1] The asymptotic size is defined by,

$$AsySz = \liminf_{n\to\infty} \inf_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_{\theta}(T_n(r(\theta)) \leq c_{n,1-\alpha}(r(\theta))) \tag{2.10}$$

[1]The test null rejection probability is defined analogously, $\mathbb{P}_{\theta}(T_n(r(\theta)) > c_{n,1-\alpha}(r(\theta)))$.

One more assumption about the convergence of the coverage probabilities is introduced.

Assumption Q *Properties of the Coverage Probabilities*

(i) Under weak identification for any $\theta_n \in \Theta(\theta_0, 0, b)$ with $b < \infty$, $CP_n(\theta_n) \to CP(p)$ for some $CP(p) \in [0,1]$, where $p = (b, \theta_0) \in \mathcal{P}$.

(ii) Under strong identification for any $\theta_n \in \Theta(\theta_0, \infty, \omega_0)$, $\liminf_{n\to\infty} CP_n(\theta_n) \geq CP_\infty$ for $CP_\infty \in [0,1]$ and for some θ_n, $CP_n(\theta_n) \to CP_\infty$.

(iii) For some $\delta > 0$, $\theta = (\zeta, \beta, \pi) \in \Theta$ with $0 < ||\beta|| < \delta$ implies that $\tilde{\theta} = (\zeta, \tilde{\beta}, \pi) \in \Theta$ for all $\tilde{\beta} \in R^{d_\beta}$ with $0 < ||\tilde{\beta}|| < \delta$.

Remark Q. Assumption Q is equivalent to Assumption ACP of Andrews and Cheng (2012). See Section 2 and in particular Lemma 2.1 of Andrews and Cheng (2012) for a more complete discussion of Asymptotic size and coverage probabilities.

2.4 Parametric Bootstrap

The parametric bootstrap method introduced in this paper will cover two separate cases. Under strong identification, the parametric bootstrap follows the residual bootstrap procedure as in Freedman (1981); Mammen (1993); Liu et al. (1988); Horowitz (2001); Davidson and Flachaire (2008). Consistent estimators of θ are obtained, to build bootstrapped samples of residuals and generate new data with the regression model that has the same distribution as the underlying data generating process. Under weak identification, the process follows the same idea with a few modifications. We cannot use the standard residuals as the estimator of π is not consistent. To eliminate this dependence, we use the residuals at the point of lack of identification, that is, we set $\beta = 0$, to eliminate the dependence of $\hat{\epsilon}_t$ on π. Using these residuals we generate bootstrap samples using a wild bootstrap fixing the value of π along a grid. Clearly, for one element of the grid, the bootstrapped samples have the same underlying distribution as the true data generating process as long as the true value π_n lies inside the grid. We introduce the bootstrap procedure for each identification category separately in a sequence of steps. The validity of each step is proven in the Appendix.

Let $\hat{\theta}_n$ be the estimator of the original sample $\{W_t\}_{t=1}^n = \{y_t, X_t\}_{t=1}^n$, $\hat{\theta}_n^m$ be the estimator with respect to a bootstrapped sample $\{W_t^m\}_{m=1,t=1}^{M_n,n}$.

The concentrated estimator objective function is defined as

$$Q_n^c(\hat{\pi}_n) = \inf_{\pi \in \Pi} Q_n^c(\pi) + o(n^{-1}) \tag{2.11}$$

The concentrated estimator is used to obtain the signal of π fixing all other parameters. The assumptions introduced in the previous sections guarantee that the gradient has a unique minimum for the derivative of the objective function with respect to θ under strong identification. Under weak identification, the limit objective function does not depend on π and therefore, the gradient is still equal to zero for any value of π. In consequence, we derive the gradient of the objective function as a function of π to obtain the distribution of ψ as a function of π.

$$\frac{\partial}{\partial \psi} Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\theta) d_{\psi,t}(\pi) \tag{2.12}$$

Constructing the t-statistic with respect to β using the parametric bootstrap has some complications. Without imposing a null hypothesis describing the behavior of the complete drifting sequence of parameters, the null hypothesis can admit two different distributions. When we construct the parametric bootstrap, it must be able to replicate the underlying distribution under the null for the weak and strong identification case. This cannot happen when the null hypothesis imposed contradicts the identification category. For example, imposing the null $H_0 : \beta = 0$, the parametric bootstrap cannot generate bootstrap samples under strong identification. For this reason, the parametric bootstrap introduced in this paper will be used to construct inference for all parameters except for β.

We begin by introducing the parametric bootstrap under strong identification to compare its derivation to the weakly identified case. For more information on the parametric bootstrap under weak identification see Horowitz (2001); MacKinnon (2006); Davidson and Flachaire (2008) among many others.

2.4.1 Strong identification

Bootstrapping residuals to generate bootstrapped samples in regression models has been used extensively in the econometrics literature since Freedman (1981). As $\hat{\theta}_n$ is a consistent estimator, and $\mathbb{E}(\epsilon_t(\theta_n)|X_t) = 0$, the standard residuals $\epsilon_t(\hat{\theta}_n)$ are valid to mimic the asymptotic distribution, and therefore we can use them to generate bootstrap samples e.g. MacKinnon and White (1985); Härdle et al. (2003). As the t-statistic is pivotal, the bootstrap method has an asymptotic refinement which leads to better small sample properties compared to the asymptotic approximation counterpart. See Horowitz (2001) for a more detailed explanation on the asymptotic refinements obtained from the higher order terms of the Edgeworth expansion of pivotal statistics. We present the steps of the parametric bootstrap under strong identification as a reference and to compare its differences to the weak identification case.

Parametric Bootstrap under Strong Identification

- **Step 1 SI**. Estimate $\hat{\theta}_n^s$ using the original data $\{W_t\}_{t=1}^n = \{y_t, X_t\}_{t=1}^n$ and the objective function introduced in Section 2.2,

$$\hat{\theta}_n^s = \arg\min_{\theta \in \Theta} Q_n(\theta) \tag{2.13}$$

Under strong identification, the estimator and residuals are consistent. Obtain the residuals.

$$\epsilon_t(\hat{\theta}_n^s) = y_t - \hat{\zeta}_n^{s\prime} X_{1,t} - \hat{\beta}_n^{s\prime} h(X_{2,t}, \hat{\pi}_n^s) \tag{2.14}$$

- **Step 2 SI**. Perform a wild bootstrap using the method of Liu et al. (1988) or the method of Shao (2010). [2]

$$[Independence] \qquad \epsilon_t^m(\hat{\theta}_n^s) = z_t^m \epsilon_t(\hat{\theta}_n^s) \qquad z_t^m \overset{d}{\sim} D(0,1) \tag{2.15}$$

$$[Dependent] \qquad \epsilon_t^m(\hat{\theta}_n^s) = z_t^m \epsilon_t(\hat{\theta}_n^s) \qquad z_t^m \overset{d}{\sim} D(0,\Gamma(z)) \tag{2.16}$$

- **Step 3 SI**. Generate bootstrapped variables $W_t^{m,s} = \{y_t^{m,s}, X_t^{m,s}\}_{m=1}^{M_n}$.

$$y_t^{m,s} = \hat{\zeta}_n^{s\prime} X_{1,t}^{m,s} + \hat{\beta}_n^{s\prime} h(X_{2,t}^{m,s}, \hat{\pi}_n^s) + \epsilon_t^m(\hat{\theta}_n^s) \tag{2.17}$$

- **Step 4 SI**. Estimate the parameters $\{\hat{\theta}_n^{m,s}\}_{m=1}^{M_n}$ for each bootstrapped sample $W_t^{m,s} = \{y_t^{m,s}, X_t^{m,s}\}_{m=1}^{M_n}$.

$$\hat{\theta}_n^{m,s} = \arg\min_{\theta \in \Theta} Q_n(W_t^{m,s}, \theta) \tag{2.18}$$

The next proposition states that the bootstrap process outlined in the previous steps leads to correct inference.

Proposition 2.4.1 *Suppose that Assumptions (J) to (N) hold and suppose that the true data generating process is strongly identified. Let $M_n \to \infty$ as $n \to \infty$. The estimators obtained following Step 1 SI to Step 4 SI have the following distribution,*

$$\sup_{z \in \mathbb{R}^{d_\theta}} \left| \mathbb{P}_m(\sqrt{M_n} B(\hat{\beta}_n^{m,s})(\hat{\theta}_n^{m,s} - \hat{\theta}_n) \leq z) - \mathbb{P}_{\theta_n}(\sqrt{n} B(\hat{\beta}_n)(\hat{\theta}_n - \theta_n) \leq z) \right| \xrightarrow{p} 0 \tag{2.19}$$

The previous proposition follows the results of Wu (1986); Liu et al. (1988) among many others that have demonstrated the validity of the wild bootstrap in regression models. The

[2]Other bootstrap methods are available in the literature which can be employed without changing the results, e.g. Resampling bootstrap Efron et al. (1979), Stationary bootstrap Politis and Romano (1994) Moving Block bootstrap Kunsch (1989), Tapered Block Bootstrap Paparoditis and Politis (2001), among others.

wild bootstrap is particularly useful to mimic the heteroscedasticity and dependence properties of the data (see Shao (2010)) with a simple multiplier bootstrap structure. As the model is known up to the parameter values, the draws of the bootstrapped samples should be as close as possible to the true data generating process when we use consistent estimators.

2.4.2 Weak identification

The idea of the following bootstrap is to notice that if we knew the true value π_n, the model would not be weakly identified, and estimation is standard. This idea will be exploited by generating bootstrapped samples $\{W_t^m\}_{m=1,t=1}^{M_n,n} = \{y_t^m, X_t^m\}_{m=1,t=1}^{M_n,n}$ for each element over a grid of π. Assumption M is important as it states that the true sequence of true parameters $\{\pi_n\}$ exists in the interior of the bootstrap grid, say $[\pi_{min}, ..., \pi_{max}] \in \tilde{\Pi}$. Assuming that π_n is an element of the grid, we know that one of those bootstrapped samples has the same distribution as the original sample $\{W_t\}_{t=1}^n$. Therefore we can obtain a bootstrapped statistic for each bootstrapped sample in the grid and obtain the supremum critical values as in Davies (1977, 1987). This test statistic would have the correct level although not the correct size. Moreover, as the value of π is weakly identified, generating many samples over a grid of π does not generate substantially different random draws. The weak identification guarantees that the signal of π is very weak, and therefore, for all values in the grid of π, the bootstrapped samples $\{y_t^m\}_{m=1}^{M_n}$ are numerically very similar.

Let $\hat{\theta}_n$ be the estimator of the original sample $\{W_t\}_{t=1}^n = \{y_t, X_t\}_{t=1}^n$, $\hat{\theta}_m$ be the estimator with respect to each bootstrapped sample $m = 1, ...M_n$. Let $W_t^m(\pi_k)$ denote the bootstrapped samples of W_t imposing $\pi = \pi_k$ (see Step 3 WI). Let $\hat{\theta}_n^m(\pi_k) = (\hat{\zeta}_n^m(\pi_k)', \hat{\beta}_n^m(\pi_k)', \hat{\pi}_n^m(\pi_k)')'$ denote the estimator of θ obtained using bootstrapped sample $W_t^m(\pi_k)$. Similarly, minimizing the concentrated estimator for each fixed π, let $\hat{\psi}_n^m(\pi, \pi_k) = (\hat{\zeta}_n^m(\pi, \pi_k)', \hat{\beta}_n^m(\pi, \pi_k)')'$ be the estimators using bootstrapped samples $W_t^m(\pi_k)$. Notice that X_t^m depends on π_k when the explanatory variables include lags of y_t.

The steps to obtain a bootstrapped asymptotic distribution are listed as follows.

Parametric Bootstrap under Weak Identification

- **Step 1 WI**. Construct a discrete grid of Π, $[\pi_{min}, ..., \pi_{max}] \in \tilde{\Pi}$, and let $d_{\tilde{\Pi}} = \dim(\tilde{\Pi})$. Obtain the extremum estimator with respect to each $\pi_k \in \tilde{\Pi}$, $k = 1, ..., d_{\tilde{\Pi}}$ for the objective function $Q_n(\psi, \pi_k)$. Notice that this problem is standard, all estimators are consistent as and weak identification is not present in this objective function. We fix π_k and obtain $\{\hat{\psi}_n(\pi_k)\}_{k=1}^{d_{\tilde{\Pi}}}$ as follows

$$\hat{\psi}_n(\pi_k) = \underset{\psi \in \Psi(\pi_k)}{\arg\min}\, Q_n(\psi, \pi_k) \tag{2.20}$$

  Obtain the optimal $\hat{\psi}_{0,n}$ setting $\beta = 0$ and obtain the residuals that do not depend on π using $\hat{\psi}_{0,n}$. In the non-linear model, the residuals are constructed as follows.

$$\epsilon_t(\hat{\psi}_{0,n}) = y_t - \hat{\zeta}'_{0,n} X_{1,t} \tag{2.21}$$

  where $\hat{\psi}_{0,n} = (0, \hat{\zeta}_n)$ for this particular model.

- **Step 2 WI**. Using the residuals centered at the point of lack of identification $\epsilon_t(\hat{\psi}_{0,n})$, obtain bootstrap samples $\{\epsilon_t^m(\hat{\psi}_{0,n})\}_{m=1}^{M_n}$ using either the Wild Bootstrap process of Liu et al. (1988) or the Dependent Bootstrap of Shao (2010).

$$[Independence] \qquad \epsilon_t^m(\hat{\psi}_{0,n}) = z_t^m(\epsilon_t(\hat{\psi}_{0,n}) - n^{-1}\sum_{t=1}^{n}\epsilon_t(\hat{\psi}_{0,n})) \qquad z_t^m \overset{d}{\sim} D(0,1) \tag{2.22}$$

$$[Dependent] \qquad \epsilon_t^m(\hat{\psi}_{0,n}) = z_t^m(\epsilon_t(\hat{\psi}_{0,n}) - n^{-1}\sum_{t=1}^{n}\epsilon_t(\hat{\psi}_{0,n})) \qquad z_t^m \overset{d}{\sim} D(0,\Gamma(z)) \tag{2.23}$$

  where D denotes a strictly stationary distribution, mean zero and variance one. Centering of $\epsilon_t(\hat{\psi}_{0,n})$ is required as they are not necessarily mean zero. Two examples are the Normal distribution or the two point distributions of Rademacher or Mammen (1993).

- **Step 3 WI**. Obtain bootstrapped samples $W_t^m(\pi_k) = \{y_t^m(\pi_k), X_t^m(\pi_k),\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$ using the bootstrapped residuals for each extremum estimator obtained in Step 1 with fixed π_k.

$$y_t^m(\pi_k) = \hat{\zeta}_n(\pi_k)'X_{1,t}^m(\pi_k) + \hat{\beta}_n(\pi_k)'h(X_{2,t}^m(\pi_k), \pi_k) + \epsilon_t^m(\hat{\psi}_{0,n}) \tag{2.24}$$

  for matrices $\{y_t^m(\pi_k)\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$ of size $(n \times d_{\tilde{\Pi}})$. The variable $\{X_t^m(\pi_k)\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$ would be equal to X_t, $\forall m$ if X_t does not include lags of y_t.

- **Step 4 WI**. Obtain the bootstrapped estimators for each element of the grid of $\pi_k \in \tilde{\Pi}$ using samples $W_t^m(\pi_k) = \{y_t^m(\pi_k), X_t^m(\pi_k)\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$.

$$\hat{\theta}_n^m(\pi_k) = Q_n^m(W_t^m(\pi_k), \theta) = \arg\min_{\theta \in \Theta} \; n^{-1}\sum_{t=1}^{n}(y_t(\pi_k) - \zeta' X_{1,t}(\pi_k) - \beta' h(X_{2,t}(\pi_k), \pi))^2 \tag{2.25}$$

  The resulting steps leads to a sequence of estimators that depend on the data generating process imposed on Step 3, $\{\hat{\theta}_n^{m,k}(\pi_k) = (\hat{\zeta}_n^{m,k}(\hat{\pi}_n^m(\pi_k))', \hat{\beta}_n^m(\hat{\pi}_n^m(\pi_k))', \hat{\pi}_n^m(\pi_k)')'\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$. Alternatively, the estimators can be obtained applying a two step procedure using the concentrated estimator.

To introduce the asymptotic results of the parametric bootstrap under weak identification, we consider first the case where π_n is assumed to be known. Even though this case is unrealistic, it illustrates the validity of the parametric bootstrap. This assumption will be dropped in the next subsection.

2.4.3 The ideal case: known π_n

When π_n is known, we can estimate ψ consistently and obtain residuals that are valid for bootstrapping. Moreover, a grid of Π is not necessary and the bootstrapped samples $\{W_t^m\}_{m=1,t=1}^{M_n,n}$ can be used to obtain bootstrapped estimators of all parameters θ. That is because if π_n were known or if we were able to consistently estimate it, there would not be a weak identification problem, and the bootstrapped would be similar to the case under strong identification.

The first proposition proved in this paper states that we can mimic the asymptotic distribution using the parametric bootstrap when π_n is known. In this case, we abstract for the grid of $\tilde{\Pi}$ and we do not have an estimator for each π_k. The assumption of known π_n will be dropped in the following subsection and is used as a preamble for the results that follow.

Proposition 2.4.2 *Suppose Assumption J to Assumption N hold. Suppose that the true value of the weakly identified parameter π_n is known and imposed on Step 3 WI. The following holds letting $M_n \to \infty$ as $n \to \infty$,*

$$\sup_{z \in \mathbb{R}^{d_\theta}} \left| \mathbb{P}_m \begin{pmatrix} \sqrt{M_n}(\hat{\psi}_n^m(\hat{\pi}_n^m, \pi_n) - \hat{\psi}_n(\pi_n)) \leq z_1 \\ \hat{\pi}_n^m(\pi_n) \leq z_2 \end{pmatrix} - \mathbb{P}_{\theta_n} \begin{pmatrix} \sqrt{n}(\hat{\psi}_n - \psi_n) \leq z_1 \\ \hat{\pi}_n \leq z_2 \end{pmatrix} \right| \xrightarrow{p} 0 \quad (2.26)$$

where $\mathbb{P}_m$ is the bootstrap induced probability measure.

The distribution on the right-hand side of the proposition id derived by Andrews and Cheng (2012) and a wild bootstrapped method to replicate this distribution is derived in Chapter 1. It is worth noticing that the parametric bootstrap can be used to construct the non-standard distribution without the knowledge of the closed for expression derived by Andrews and Cheng (2012).

2.4.4 The realistic case: unknown π_n

Now we focus on the asymptotic size of the test when we do not know the true value π_n. The parametric bootstrap generates bootstrapped samples that depend on the parameter π_k along the grid, which we write as $W_t^m(\pi_k)$. With the bootstrapped samples we can obtain bootstrapped estimators $\hat{\theta}_n^m(\pi_k)$ and generate the distribution of the t-statistic for each element of the grid. As the true π_n is unknown, we construct the supremum of the critical values as in Davies (1977, 1987).

Let $c_{1-\alpha}^{\psi}(\pi_k)$ denote the one tailed critical values of the asymptotic distribution under weak identification of ψ using a data generating process that sets $\pi = \pi_k$, i.e. the critical values of $\tau(\pi, \theta_0, b)$ generating using the data generating process with π_k. Clearly, the supremum of this critical value along all π_k, is rejected with probability of at least α for all estimators generated with any π_k. Similarly, let $c^{\pi}_{1-\alpha}(\pi_k)$ denote the critical value of the asymptotic distribution under weak identification of π, i.e. $\pi^*(\theta_0, b)$ using the data generating process that assumes $\pi = \pi_k$.

Proposition 2.4.3 *Suppose Assumption J to Assumption N hold. Let π_n be unknown. Under weak identification, the following holds letting $M_n \to \infty$ as $n \to \infty$.*

$$\liminf_{n\to\infty} \ \mathbb{P}_m \begin{pmatrix} \sqrt{M_n}(\hat{\psi}^m_n(\hat{\pi}^m_n, \pi_k) - \hat{\psi}_n(\pi_k)) & \leq & \sup_{\pi_k \in \tilde{\Pi}} c^{\psi}_{1-\alpha}(\pi_k) \\ \hat{\pi}^m_n(\pi_k) & \leq & \sup_{\pi_k \in \tilde{\Pi}} c^{\pi}_{1-\alpha}(\pi_k) \end{pmatrix} \geq 1 - \alpha \qquad (2.27)$$

with probability approaching one.

The proposition above is a first step to prove that the t-statistic introduced in the next section has correct asymptotic size. The t-statistic is based on this asymptotic result and the construction of matrices that converge in probability to a constant. In contrast to the standard t-statistic, the robust t-statistic constructed in Section 2.6 will include a normalization matrix which is necessary to obtain a finite probability limit of the variance-covariance matrix under weak identification.

2.5 Imposing the null hypothesis

2.5.1 Strong Identification, imposing the null hypothesis

The result in Proposition 2.4.1 gives us a bootstrap method to construct the bootstrapped t-statistic under strong identification. If we wish to test $H_0 : \theta = \theta_{H0}$, we impose the null hypothesis and generate bootstrapped random draws of the distribution of y_t under the null. As stated by MacKinnon (2006), ”imposing the restrictions of the null hypothesis yields more efficient estimates of the nuisance parameters upon which the distribution of the test statistic may depend. This generally makes bootstrap tests more reliable, because the parameters of the bootstrap data generating process are estimated more precisely".

The wild bootstrapped t-test developed in Chapter 1 derives the asymptotic distribution under strong and weak identification and subsequently generates the bootstrapped distribution using a wild bootstrap. The parametric bootstrap is constructed without the knowledge of those underlying distributions. Imposing the null hypothesis must be implemented in Step 3 because this step generates bootstrapped samples of the underlying data generating process.

First, we consider imposing the null hypothesis under strong identification. Suppose we implement the parametric bootstrap Step 1 SI to Step 4 SI, replacing Step 3 with the following,

- **Step 3 SI***. Generate bootstrapped random variables $W_t^{m,s} = \{y_t^{m,s}, X_t^{m,s}\}_{m=1,t=1}^{M_n,n}$ imposing the null hypothesis $H_0 : \theta = \theta_{H0}$

$$y_t^{m,s} = \zeta_{H0}^{s'} X_{1,t}^{m,s} + \beta_{H0}^{s'} h(X_{2,t}^{m,s}, \pi_{H0}^{s}) + \epsilon_t^{m}(\hat{\theta}_n^{s}) \tag{2.28}$$

Besides Step 3 SI, the rest of the bootstrap method is equivalent to the parametric bootstrap described in Section 2.4. Proposition 2.4.1 follows when the null is imposed, although the distribution would be centered by θ_{H0} instead of $\hat{\theta}_n$.

2.5.2 Weak Identification, imposing the null hypothesis

Imposing the null hypothesis under weak identification is implemented in Step 3 although this has the potential to disregard the grid of parameters π. This implies that imposing the null hypothesis reduces the number of nuisance parameters and can potentially eliminate all of them in some cases. For example, imposing the null hypothesis $H_0 : \pi = a$ leads to a t-test without nuisance parameters because if the null is true, weak identification is no longer present and all parameters in the estimation are consistent. Fixing the value of π would imply that the estimator $\hat{\pi}_n$ is not required to generate bootstrapped samples.

The number of nuisance parameters in the bootstrapped distribution is different the null is imposed on either ψ or π. First we assume that the null hypothesis imposes the null with respect to the strongly identified parameters ψ. Suppose we implement the parametric bootstrap Step 1 WI to Step 4 WI, replacing Step 3 with the following,

- **Step 3 WI* (ψ).** With the bootstrapped residuals obtained in Step 2 WI, construct bootstrapped samples $W_t^m(\pi_k) = \{y_t^m(\pi_k), X_t^m(\pi_k),\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$ imposing $H_0 : \psi = \psi_{H0}$ for each π_k.

$$y_t^m(\pi_k) = \hat{\zeta}_{H0}(\pi_k)X_{1,t}^m(\pi_k) + \hat{\beta}_{H0}(\pi_k)h(X_{2,t}^m(\pi_k), \pi_k) + \epsilon_t^m(\hat{\psi}_{0,n}) \tag{2.29}$$

for matrices $\{y_t^m(\pi_k)\}_{m=1,k=1}^{M_n,d_{\tilde{\Pi}}}$ of size $(n \times d_{\tilde{\Pi}})$.

In the non-linear model introduced in Section 2.3, imposing the null with respect to β leads to a reduction of the set of nuisance parameters to only one, π_0. As π_0 is the only nuisance parameter, the grid $\tilde{\Pi}$ is used to fix that value to generate bootstrapped samples. On the other hand, imposing the null with respect to parameter π, i.e. when we are testing $H_0 : \pi = a$, leads to a generating bootstrapped samples y_t without nuisance parameters as the model is no longer weakly identified.

- **Step 3 WI* (π).** With the bootstrapped residuals constructed in Step 2 WI, obtain the bootstrapped samples $\{y_t^m(\pi_{H0}), X_t^m(\pi_{H0})\}_{m=1}^{M_n}$ imposing $H_0 : \pi = \pi_{H0}$.

$$y_t^m(\pi_{H0}) = \hat{\zeta}_n(\pi_{H0})'X_{1,t}^m(\pi_{H0}) + \hat{\beta}_n(\pi_{H0})'h(X_{2,t}^m(\pi_{H0}), \pi_{H0}) + \epsilon_t^m(\hat{\psi}_{0,n}) \tag{2.30}$$

for matrices $\{y_t^m(\pi_{H0})\}_{m=1}^{M_n}$ of size $(n \times d_{\tilde{\Pi}})$.

It is worth noticing that imposing the null hypothesis can contradict the identification category we are studying. For example, if we wish to test $H_0 : \beta = 0$, this null hypothesis implies weak identification. Constructing the parametric bootstrap under strong identification imposing this null hypothesis would lead to invalid critical values as both conditions are contradictory. Nonetheless, the parametric bootstrap under weak identification would lead to the correct critical values (surprisingly without nuisance parameters) as weak identification is implied by the null hypothesis.

Generating bootstrapped samples using the parametric bootstrap method can be computationally intensive in a model with a large number of parameters. Nonetheless, the parametric bootstrap abstracts from the derivation of the expressions that compose the non-standard distributions of the estimators and the simulation of these processes. The parametric bootstrap only requires to bootstrap residuals and generate new bootstrap samples. Moreover, it has a computational intensity comparable to other bootstrap methods with dependent data.

2.6 Bootstrapped t-statistic

The bootstrapped distributions of the estimators developed in the previous sections are used to construct the bootstrapped t-statistic under strong and weak identification. The t-statistic not only has a different distribution according to the identification category, it can also have a different convergence rate. For example, $\hat{\pi}_n = O_p(1/\sqrt{n})$ under strong identification but $\hat{\pi}_n = O_p(1)$ under weak identification. To obtain an $O_p(1)$ t-statistic, we will adopt a normalization matrix $B(\beta)$.

Let the null hypothesis be defined as,

$$H_0 : r(\theta) = q \tag{2.31}$$

The robust t-statistic requires a normalization matrix $B(\beta)$, defined as follows,

$$B(\beta) = \begin{bmatrix} I_{d_\psi} & I_{d_\psi \times d_\pi} \\ I_{d_\pi \times d_\psi} & i(\beta) I_{d_\psi} \end{bmatrix} \tag{2.32}$$

$$i(\beta) = \begin{cases} \beta, & \text{if } \beta \text{ is a scalar} \\ ||\beta||, & \text{if } \beta \text{ is a vector} \end{cases}$$

The matrix $B(\beta)$ is necessary to avoid singularity of the variance-covariance matrix which leads to a t-statistic bounded in probability for any parameter. See Andrews and Cheng (2012) for more details on the normalization matrix $B(\beta)$, in particular Theorem 3.3. The robust t-statistic is defined as follows,

$$T_n = \frac{\sqrt{n}(r(\hat{\theta}_n) - q)}{[r_\theta(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n B^{-1}(\beta_n) r_\theta(\hat{\theta}_n))']^{1/2}} \tag{2.33}$$

The robust t-statistic is compared to the bootstrapped distribution for both identification categories constructed using the parametric bootstrap.

2.6.1 Bootstrapped distribution of the robust t-statistic

Theorem 4.1 of Andrews and Cheng (2012) derive the limit distribution of the t-statistic for each identification category. The limit distribution depends on the identification category and on which parameter is being tested. Specifically, under weak identification, the t-statistic takes a particular distribution when we impose a null hypothesis with respect to ψ and another distribution when we test a null hypothesis with respect to π.

The parametric bootstrap is used to replicate this distributions directly using the regression model and residuals centered at the point of lack of identification. We begin by assuming that the model is weakly identified and that π_n is known. We also assume that we are testing a null hypothesis with respect to ψ. The asymptotic distribution of the t-statistic when we test with respect to the strongly identified parameters ψ, is denoted by T^ψ and and can be constructed using the bootstrapped estimators $\{\theta_n^m\}_{m=1}^{M_n}$,

$$T_n^{\psi,m}(\pi) = \frac{r_\psi(\pi)\sqrt{n}(\hat{\psi}_n^m(\pi,\pi_n) - \hat{\psi}_n(\pi_n))}{[r_\psi(\pi)\hat{\Sigma}_n(\pi) r_\psi(\pi)']^{1/2}} \tag{2.34}$$

where r_ψ is the derivative of the restriction matrix and Σ denotes the variance covariance matrix defined on Chapter 1. Notice that $T_n^{\psi,m}$ is the bootstrapped version of T^ψ using the parameter distribution $\sqrt{n}(\hat{\psi}_n^m(\pi,\pi_n) - \hat{\psi}_n(\pi_n)$.

Let $S_\beta = [I_{d\beta} : 0_{d\beta \times d_\zeta}]$ is the selector matrix that selects β out of ψ for parameter β.

The bootstrapped t-statistic takes the following form when the null hypothesis tested is with respect to π,

$$T_n^{\pi,m}(\pi) = \frac{||S_\beta(\sqrt{n}(\hat{\psi}_n^m(\pi,\pi_n))||(r(\hat{\psi}_n^m(\pi,\pi_n),\pi) - r(\hat{\psi}_n^m(\pi,\pi_n),\pi_n))}{[r_\pi(\pi)\hat{\Sigma}_n(\pi)r_\pi(\pi)']^{1/2}} \tag{2.35}$$

When the model is strongly identified, the distribution of the t-statistic does not depend on which parameters we are testing. It is not very surprising that as the asymptotic distribution of $\hat{\theta}_n$ is standard normal under the proper normalization [3],

$$T_n^{\theta,m} = \frac{r_\theta(\hat{\theta}_n^{m,s})(\sqrt{n}B(\hat{\beta}_n^{m,s})(\hat{\theta}_n^{m,s} - \hat{\theta}_n))}{[r_\theta(\hat{\theta}_n^{m,s})\hat{\Sigma}_n(\hat{\theta}_n^{m,s})r_\theta(\hat{\theta}_n^{m,s})]^{1/2}} \tag{2.36}$$

Let T^ψ, T^π denote the asymptotic distribution of the t-statistic under weak identification when the restrictions are imposed on ψ and π respectively. The next theorem states that the bootstrapped distributions approximate the asymptotic distribution for each identification case separately.

Theorem 2.6.1 *Suppose Assumption J to Assumption Q are satisfied. Let π_n be known. Using the parametric bootstrap introduced in from Step 1 WI to Step 4 WI and Step 1 SI to Step 4 WI, and letting $M_n \to \infty$ as $n \to \infty$ for each identification category,*

(a) Under weak identification with $\dim(r_\pi(\theta)) = 0$, $T_n^{\psi,m}(\hat{\pi}_n^m(\pi_n)) \xrightarrow{d} T^\psi(\pi^*(\theta_0,b);\theta_0,b)$

(b) Under weak identification with $\dim(r_\pi(\theta)) = 1$, $T_n^{\pi,m}(\hat{\pi}_n^m(\pi_n)) \xrightarrow{d} T^\pi(\pi^*(\theta_0,b);\theta_0,b)$

(c) Under strong identification, $T_n^{\theta,m} \xrightarrow{d} N(0,1)$

The previous proposition assumes that π_n and the identification category is known. In the following subsection, we relax both of this unrealistic assumptions.

[3] the normalization matrix is necessary for the semi-strong identification category

2.6.2 Robust Critical Values

The previous results state that we can obtain critical values of the distributions under strong and weak identification. Nonetheless, it is not clear which critical values should be employed when the identification category is unknown. The simplest way to obtain the correct level of a test is to be conservative and use the largest critical value of both distributions.

The LF critical value proposed by Andrews and Cheng (2012) follows this approach as it selects the largest critical value. Let $c^{m}_{n,1-\alpha}(p)$, $p \in \mathcal{P}$ denote the critical values of the weakly identified bootstrap statistic constructed using Step 1 WI to Step 4 WI. Recall the set $\mathcal{P}$ denotes the set of nuisance parametrs that arise in the bootstrapped distribution of the t-statistic when the model is weakly identified. Let $c^{m,s}_{n,1-\alpha}$ denote the critical values of the strong identified bootstrap statistic constructed using Step 1 SI to Step 4 SI. The LF critical value is defined as

$$c^{LF,m}_{n,1-\alpha} = \max\{\sup_{\pi_k \in \tilde{\Pi}} c^{m}_{n,1-\alpha}(\pi_k), c^{m,s}_{n,1-\alpha}\} \tag{2.37}$$

The set of nuisance parameters $\mathcal{P}$ would change according to the parameters tested as well as the null hypothesis imposed. The wild bootstrap t-test developed in Chapter 1 and the asymptotic approximations of Andrews and Cheng (2012) have two nuisance parameters under the null (b, π_0). The parametric bootstrap, on the other hand, the nuisance parameters are only π_0 as b is not required to impose the null hypothesis. As only one of the π_k is the correct one, the sup of the critical values along the grid provides with a test of the correct level. Moreover, when the test is with respect to π, imposing the null eliminates all nuisance parameters and in consequence, our statistic has the correct test size. The simulation exercise in the next section shows that is this case the parametric bootstrap has empirical sizes comparable to the infeasible statistic where nuisance parameters are known.

Clearly, the LF critical value is not ideal because it is over-conservative. To improve the size and power of the test, Andrews and Cheng (2012) propose a data-driven critical value, ICS_1, that relies on the first-step test of weak identification. If we cannot reject the null of weak identification, the LF critical value is selected; otherwise, the strong identification bootstrapped critical value is used. We modify the ICS_1 critical values introducing a critical value we name ICS_0. This critical value is based on selecting between weak and strong identification, unlike the ICS_1 critical value which selects between LF and strong identification.

$$c_{n,1-\alpha}^{ICS_0,m} = \begin{cases} \sup\limits_{\pi_k \in \tilde{\Pi}} c_{n,1-\alpha}^{m}(\pi_k) & \text{if} \quad \hat{A}_n \leq \kappa_n \\ c_{n,1-\alpha}^{m,s} & \text{if} \quad \hat{A}_n > \kappa_n \end{cases} \tag{2.38}$$

$$\hat{A}_n = (n\hat{\beta}_n \hat{\Sigma}_n^{-1} \hat{\beta}_n)^{1/2} \tag{2.39}$$

$\hat{A}_n$ is a first step statistic and κ_n is a sequence of positive constants such that $\kappa_n \to \infty$ and $\kappa_n/n^{1/2} = o(1)$. For example, $\kappa_n = (\log n)^{1/2}$ is analogous to the penalty term used in the Bayesian Information Criterion. The null hypothesis $H_0 : ||b|| < \infty$ is accepted in favor of weak identification when $\hat{A}_n \leq \kappa_n$, while the null hypothesis is rejected in favor of strong identification i.e. $H_1 : ||b|| = \infty$ when $\hat{A}_n > \kappa_n$. For large enough n, the statistic $\hat{A}_n$ established if the parameters of the model are weakly identified.

The following proposition proves that the robust critical values lead to correct asymptotic size. Unlike Theorem 2.6.1, the conditions of the theorem specify that π_n and the identification category are unknown.

Theorem 2.6.2 *Let $M_n \to \infty$ as $n \to \infty$. Under the null hypothesis $H_0 : r(\theta) = q$, the LF and ICS_0 critical values of the t-test have correct asymptotic size w.p.a.1,*

(a) $AsySz^{LF,m} = \liminf\limits_{n\to\infty} \inf\limits_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c_{n,1-\alpha}^{LF,m}(r(\theta))) = 1-\alpha$

(b) $AsySz^{ICS_0,m} = \liminf\limits_{n\to\infty} \inf\limits_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c_{n,1-\alpha}^{ICS_0,m}(r(\theta))) = 1-\alpha$

(c) If H_0 is false, that is $r(\theta_n) \neq q$, then $T_n(r(\theta)) \xrightarrow{p} \infty$

The LF critical values lead to test sizes that can be substantially smaller than the correct test sizes when an incorrect identification category is used. On the other hand, the ICS_0 critical value uses the correct asymptotic distribution as long as the first step estimator $\hat{A}_n$ establishes the identification category. The next section develops a simulation exercise shows that the parametric bootstrap has good properties and performs very well empirically.

2.7 Simulation Exercise

The bootstrapped procedure introduced in this paper is motivated by the loss of size and power from the nuisance parameters that appear in the distribution in Chapter 1. In this section, we test if the parametric bootstrap helps to obtain more accurate test sizes in practice.

For the simulation exercise, we use the exponential smoothing model as in Cheng (2015). The model specification is defined as follows.

$$\epsilon_t(\theta_n) = y_t - \zeta_0 x_{1,t} - \beta_n y_{t-1}(1 - \exp(-c(x_{2,t} - \pi_0)^2)) \tag{2.40}$$

For this model, there is not need to use drifting sequences of true parameters for ζ and π, the speed at which $\beta_n \to \infty$ will determine the identification category. We assume the following true values in the simulations : $\zeta_0 = 1, \pi_0 = 0$ and $b = 1$. In this exercise we compare the identification categories as follows, under strong identification we set $\beta_n = b$, under weak identification we set $\beta_n = b/\sqrt{n}$ and under non-identification (limiting case of weak identification) we set $\beta_n = 0$.

The sample size takes value $n = 100, 250$ and 500. We set are $x_{1,t}, x_{2,t} \overset{d}{\sim} N(0,1)$, $c = -1$. We let the true errors to be either Normal(0,1) or GARCH(1,1) with $\omega = 0.1, \alpha = 0.3$ and $\beta = 0.6$. The wild bootstrap of Liu et al. (1988) using the two point distribution multiplier as in Mammen (1993) to generate the bootstrapped residuals in Step 2. The number of simulations is 1,000. For each simulation, we construct bootstrapped samples using 500 draws. We also refer Chapter 1 for a complete reference of gradient, Hessian and other expressions required for estimation and construction of the bootstrap as well as other simulation details. The remaining tables are presented in the Appendix.

We consider three null hypotheses in this simulation exercise. The first null hypothesis sets the parameters equal to their true values. To evaluate the power of the test, the second and third hypothesis consider the false null hypotheses of parameters within one and three standard deviations of the true value.

$$\begin{aligned}
&H_{0,1}^{\beta}: \beta = \beta_n \qquad && H_{0,1}^{\pi}: \pi = \pi_0 \\
&H_{0,2}^{\beta}: \beta = \beta_n + \sigma_\beta \qquad && H_{0,2}^{\pi}: \pi = \pi_0 + \sigma_\pi \\
&H_{0,3}^{\beta}: \beta = \beta_n + 3\sigma_\beta \qquad && H_{0,3}^{\pi}: \pi = \pi_0 + 3\sigma_\pi
\end{aligned} \tag{2.41}$$

The robust sample t-statistic is constructed by

$$T_n = \frac{\sqrt{n}(r(\hat{\theta}_n) - v)}{[r_\theta(\hat{\theta}_n) B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n(\hat{\theta}_n) B^{-1}(\hat{\beta}_n) r_\theta(\hat{\theta}_n)']^{1/2}} \tag{2.42}$$

while the standard t-statistic takes the form,

$$T_n^s = \frac{\sqrt{n}(r(\hat{\theta}_n) - v)}{[r_\theta(\hat{\theta}_n)\hat{\Sigma}_n(\hat{\theta}_n) r_\theta(\hat{\theta}_n)']^{1/2}} \tag{2.43}$$

The critical values of the bootstrapped t-statistic are computed using order statistics. Let $\{T_{a,n}^m(\pi)\}_{j=1}^m$ be a sequence of independent draws of the t-statistic with $a = \psi, \pi, \theta$. Denote the order statistics by $T_{a,n,[1]}^m \leq T_{a,n,[2]}^m, ...$, etc. The LF and ICS_0 critical values are computed using $c_{n,1-\alpha/2}^{a,m} = \inf\{c \geq 0 : \mathbb{P}(T_n^m \leq c) \geq 1 - \alpha/2\}$ and $c_{n,\alpha/2}^{a,m} = \inf\{c \geq 0 : \mathbb{P}(T_n^m \leq c) \geq \alpha/2\}$, as we use two tailed critical values. To construct the ICS_0 critical value, we use $\kappa_n = (\ln(n))^{1/2}$, as suggested by Andrews and Cheng (2012).

Tables 2.3 and 2.4 compares the results of the parametric bootstrap introduced in this paper and the (unfeasible) asymptotic approximation of Andrews and Cheng (2012). The LF AC and ICS_0 AC critical values are unfeasible because it is assumed that the nuisance parameters are known while the identification category is still unknown. The bootstrapped critical values are feasible as they do not assume that the identification category or the nuisance parameters are known. Even though the testing of β is not valid using the parametric bootstrap, the tables illustrate its performance and its irregular behavior in the weakly identified case.

The results in Tables 2.3 and 2.4 indicate that the parametric bootstrap works exceptionally well for the cases of strong and non-identification. The critical values that are constructed in these cases are numerically close to the infeasible critical values of Andrews and Cheng (2012). When the model is weakly identified, the critical values work well, but not as well as the infeasible case. The difference in accuracy hinges on the inability of the $\hat{A}_n$ statistic to recognize if the parameters are weakly or strongly identified. When the $\hat{A}_n$ leads to the incorrect conclusion, the ICS_0 selects the incorrect critical value, and in consequence rejection rates are usually higher than the correct test size. The least favorable critical values perform better as the simulation exercise shows that the distributions under weak identification have larger critical values that the strong identification case. The critical values for π work particularly well compared to the asymptotic approximations of Andrews and Cheng (2012). In the case of the asymptotic approximations, test sizes close to zero when we test with respect to π. As the t-test is centered at π_0 which is a nuisance parameter, generating a grid and taking the supremum of the critical values does not perform well in practice because the critical values are too wide. Using the parametric bootstrap presented in this paper, we can test π without nuisance parameters, which performs as well as the infeasible critical values of Andrews and Cheng (2012). In summary, the parametric bootstrap works very well as long as the ICS_0 critical values are able to recognize if the parameters of the model are weakly or strongly identified.

2.8 Conclusion

We introduce a parametric bootstrap method for models where parameters are potentially weakly identified. The parametric bootstrap is easy to construct as closed-form expressions of the t-statistic are not required, and can reduce the number of nuisance parameters. The simulation exercise suggests that the t-test constructed using the parametric bootstrap leads to accurate test size and test power compared to the asymptotic approximations of Andrews and Cheng (2012) or the wild bootstrap method of Chapter 1.

	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
Strongly Identified $n = 500$ $(\beta_n = 0.5)$															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.086	0.087	0.095	0.090	0.095	0.042	0.057	0.095	0.094	0.095	0.102	0.103	0.103	0.104	0.103
5%	0.041	0.047	0.043	0.053	0.043	0.020	0.029	0.046	0.049	0.046	0.040	0.048	0.040	0.048	0.040
1%	0.008	0.011	0.008	0.011	0.008	0.004	0.007	0.009	0.012	0.009	0.010	0.009	0.010	0.010	0.010
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.246	0.252	0.264	0.259	0.264	0.108	0.255	0.244	0.259	0.244	0.291	0.278	0.292	0.279	0.292
5%	0.154	0.154	0.162	0.160	0.162	0.061	0.164	0.156	0.166	0.156	0.185	0.178	0.185	0.180	0.185
1%	0.048	0.062	0.057	0.064	0.057	0.022	0.060	0.052	0.061	0.052	0.067	0.062	0.067	0.063	0.067
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.902	0.904	0.906	0.909	0.906	0.774	0.915	0.911	0.915	0.911	0.911	0.907	0.911	0.908	0.911
5%	0.847	0.851	0.856	0.854	0.856	0.679	0.857	0.846	0.857	0.846	0.872	0.853	0.873	0.853	0.873
1%	0.616	0.638	0.654	0.648	0.654	0.459	0.660	0.654	0.664	0.654	0.706	0.672	0.706	0.676	0.706
Weakly Identified $n = 500$ $(\beta_n = 0.5/\sqrt{n})$															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.096	0.109	0.110	0.112	0.103	0.106	0.101	0.107	0.141	0.249	0.055	0.088	0.110	0.098	0.056
5%	0.036	0.043	0.041	0.044	0.044	0.054	0.046	0.066	0.085	0.138	0.025	0.043	0.060	0.051	0.025
1%	0.004	0.009	0.009	0.009	0.006	0.010	0.011	0.017	0.024	0.030	0.003	0.006	0.008	0.008	0.003
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.260	0.269	0.276	0.272	0.275	0.356	0.384	0.358	0.394	0.415	0.361	0.424	0.433	0.436	0.361
5%	0.152	0.158	0.164	0.160	0.164	0.314	0.333	0.315	0.357	0.387	0.277	0.331	0.372	0.346	0.277
1%	0.054	0.070	0.072	0.074	0.060	0.178	0.196	0.178	0.230	0.298	0.146	0.182	0.236	0.199	0.146
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.900	0.905	0.907	0.905	0.909	0.905	0.927	0.913	0.945	0.969	0.783	0.820	0.820	0.828	0.783
5%	0.826	0.832	0.834	0.834	0.837	0.861	0.890	0.870	0.910	0.935	0.724	0.780	0.787	0.786	0.724
1%	0.631	0.656	0.655	0.660	0.665	0.731	0.751	0.735	0.792	0.841	0.630	0.684	0.699	0.693	0.631
Non-Identified $n = 500$ $(\beta_n = 0)$															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.086	0.086	0.093	0.087	0.092	0.092	0.095	0.092	0.098	0.247	0.049	0.088	0.115	0.098	0.049
5%	0.043	0.049	0.049	0.051	0.046	0.057	0.057	0.057	0.058	0.115	0.011	0.031	0.053	0.040	0.011
1%	0.011	0.014	0.013	0.014	0.013	0.009	0.010	0.026	0.015	0.026	0.000	0.002	0.004	0.002	0.000
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.268	0.280	0.278	0.282	0.282	0.458	0.468	0.462	0.468	0.517	0.363	0.424	0.441	0.439	0.363
5%	0.174	0.183	0.188	0.190	0.182	0.374	0.381	0.374	0.384	0.480	0.279	0.331	0.372	0.346	0.279
1%	0.043	0.058	0.057	0.059	0.051	0.203	0.212	0.203	0.221	0.291	0.151	0.192	0.259	0.211	0.151
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.906	0.909	0.908	0.909	0.912	0.935	0.937	0.935	0.937	0.973	0.796	0.836	0.838	0.840	0.796
5%	0.844	0.853	0.853	0.853	0.859	0.899	0.903	0.899	0.906	0.941	0.745	0.788	0.797	0.798	0.747
1%	0.625	0.659	0.664	0.663	0.664	0.783	0.787	0.786	0.793	0.852	0.641	0.702	0.722	0.715	0.641

Table 2.3: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with Normal(0,1) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 500$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

	LF_ζ AC	LF_ζ PB	$ICS_{0,\zeta}$ AC	$ICS_{0,\zeta}$ PB	Std	LF_β AC	LF_β PB	$ICS_{0,\beta}$ AC	$ICS_{0,\beta}$ PB	Std	LF_π AC	LF_π PB	$ICS_{0,\pi}$ AC	$ICS_{0,\pi}$ PB	Std
Strongly Identified $n = 500$ ($\beta_n = 0.5$)															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.099	0.103	0.106	0.114	0.106	0.047	0.055	0.108	0.116	0.108	0.110	0.112	0.121	0.123	0.121
5%	0.046	0.051	0.057	0.066	0.057	0.030	0.028	0.052	0.058	0.052	0.062	0.054	0.067	0.063	0.067
1%	0.012	0.012	0.013	0.015	0.013	0.015	0.013	0.022	0.026	0.022	0.012	0.013	0.016	0.024	0.016
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.284	0.279	0.303	0.291	0.304	0.199	0.319	0.313	0.328	0.315	0.857	0.853	0.866	0.856	0.865
5%	0.179	0.185	0.196	0.211	0.196	0.133	0.229	0.217	0.233	0.218	0.804	0.799	0.812	0.802	0.811
1%	0.067	0.076	0.076	0.097	0.076	0.057	0.102	0.088	0.107	0.088	0.664	0.629	0.668	0.645	0.667
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.915	0.911	0.921	0.915	0.921	0.888	0.919	0.920	0.921	0.920	0.988	0.988	0.989	0.989	0.990
5%	0.855	0.851	0.871	0.865	0.871	0.833	0.876	0.879	0.878	0.879	0.986	0.986	0.988	0.988	0.986
1%	0.690	0.687	0.713	0.726	0.713	0.696	0.736	0.745	0.741	0.746	0.981	0.979	0.984	0.982	0.983
Weakly Identified $n = 500$ ($\beta_n = 0.5/\sqrt{n}$)															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.105	0.106	0.113	0.110	0.111	0.129	0.138	0.129	0.160	0.272	0.064	0.073	0.106	0.112	0.076
5%	0.047	0.051	0.054	0.053	0.051	0.072	0.075	0.081	0.103	0.164	0.033	0.028	0.052	0.044	0.039
1%	0.005	0.006	0.007	0.008	0.007	0.023	0.014	0.031	0.037	0.041	0.000	0.001	0.005	0.005	0.001
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.242	0.252	0.258	0.259	0.255	0.396	0.423	0.401	0.441	0.475	0.392	0.412	0.461	0.471	0.399
5%	0.157	0.162	0.166	0.166	0.164	0.355	0.375	0.355	0.385	0.428	0.320	0.322	0.385	0.372	0.326
1%	0.056	0.065	0.068	0.071	0.061	0.246	0.255	0.247	0.265	0.335	0.175	0.162	0.255	0.228	0.185
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.891	0.894	0.897	0.899	0.899	0.941	0.947	0.948	0.949	0.965	0.810	0.826	0.844	0.856	0.813
5%	0.822	0.826	0.835	0.830	0.833	0.920	0.926	0.925	0.934	0.945	0.754	0.770	0.804	0.807	0.758
1%	0.642	0.653	0.673	0.674	0.658	0.864	0.863	0.872	0.875	0.891	0.656	0.666	0.728	0.724	0.666
Non-Identified $n = 500$ ($\beta_n = 0$)															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.081	0.084	0.087	0.088	0.086	0.126	0.132	0.126	0.138	0.283	0.073	0.083	0.139	0.133	0.078
5%	0.043	0.048	0.050	0.050	0.047	0.064	0.064	0.068	0.072	0.151	0.030	0.031	0.067	0.054	0.032
1%	0.009	0.009	0.013	0.010	0.013	0.013	0.015	0.038	0.027	0.038	0.004	0.002	0.008	0.005	0.004
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.269	0.281	0.286	0.287	0.291	0.473	0.483	0.475	0.484	0.527	0.387	0.401	0.442	0.449	0.391
5%	0.170	0.175	0.182	0.183	0.187	0.428	0.428	0.428	0.430	0.491	0.321	0.327	0.386	0.378	0.328
1%	0.042	0.062	0.062	0.064	0.050	0.305	0.277	0.307	0.288	0.374	0.176	0.166	0.275	0.226	0.184
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.904	0.898	0.908	0.908	0.909	0.944	0.935	0.945	0.941	0.956	0.800	0.814	0.830	0.843	0.804
5%	0.843	0.841	0.849	0.848	0.854	0.923	0.912	0.924	0.917	0.939	0.750	0.775	0.798	0.812	0.757
1%	0.659	0.674	0.689	0.693	0.682	0.880	0.864	0.881	0.875	0.889	0.639	0.644	0.708	0.702	0.645

Table 2.4: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 500$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

CHAPTER 3

THE RISK-RETURN TRADE-OFF UNDER WEAK IDENTIFICATION

3.1 Introduction

The risk-return trade-off is one of the most important relations in the finance literature. The relation characterizes the common conception that potentially high levels of risk of an asset should be followed by potentially high returns. Based on the portfolio selection mean-variance analysis model of Markowitz (1952), Merton (1973) proposed the Intertemporal Capital Asset Pricing Model (ICAPM). The ICAPM is derived by solving a micro-founded optimization problem in which a representative investor maximizes his expected utility investing in a portfolio of risky assets and a risk-free bond. In equilibrium, the ICAPM characterizes the risk-return trade-off, suggesting that conditional excess return of a portfolio should have a linear relationship with respect to its conditional variance.

$$\mathbb{E}_t(R_{t+1}) = \zeta + \beta Var_t(R_{t+1}) \tag{3.1}$$

Here β is the coefficient of relative risk aversion of the representative agent, and ζ should be equal to zero. The expectation and variance are conditional on the information prior to time t. According to the risk-return trade-off relationship obtained by Merton (1973), the value of β must be positive and statistically significant, implying that the conditional variance of an asset has a positive and linear relationship with respect to conditional excess returns.

A significant number of research papers have tested the theoretical implications of the risk-return trade-off using data, which has lead to contradictory results. One of the most important difficulties to estimate the risk-return trade-off empirically hinges on the estimation of the (unobservable) conditional variance. As the measure of conditional mean and variance depends on the model at hand, some authors find a positive risk-return trade-off while other find a negative relationship. Using a GARCH-in-Mean model, Baillie and DeGennaro (1990) find a positive non-significant relationship, while French, Schwert, and Stambaugh (1987) obtain similar results using a rolling window estimator of conditional variance. On the other hand, Campbell (1987) and Nelson (1991) find a negative, statistically significant risk-return trade-off. In particular, Nelson (1991) uses a GARCH model that accounts for volatility and return distributional asymmetries (i.e. leverage effect). Glosten, Jagannathan, and Runkle (1993) and Turner, Startz, and Nelson (1989) argue that the results of risk-return relationship can change from positive to negative by slightly modifying their model specification.

Lettau and Ludvigson (2001b) argue that conditioning on information is crucial to obtain meaningful results for the risk-return relationship. In particular, they find a positive and significant risk-return relationship using conditional returns and a negative non-significant relationship using unconditional returns. Ludvigson and Ng (2007) use a three-factor model with 172 financial indicators and find that these factors lead to a positive and significant risk-return relationship. The three factors contain synthesized information from volatility, risk and "real economy" factors. Lettau and Ludvigson (2001a) propose the residual of a cointegrating relationship, $c\hat{a}y_t$, formed with consumption, wealth and labor income data. The authors argue that $c\hat{a}y_t$ has a strong ability to forecast the return on aggregate stock market indices, which supports the validity of the CCAPM model to explain future market returns.

Given that the different results in the literature depend substantially on model selection, Ghysels et al. (2005) propose an optimally weighted estimator of conditional variance using the mixed data sampling (MIDAS) model. The MIDAS estimator of conditional variance combines high-frequency (e.g. daily) data into a lower frequency (e.g. monthly, quarterly) specification using a data-driven optimized weighting function. The results of Ghysels et al. (2005) suggest that there is a positive and significant risk-return trade-off robust to subsamples. The authors argue that their results emerge because estimating the optimal weights leads to a more accurate estimator of conditional variance. In particular, the optimal weights are key to identify the persistence component of the volatility process. The short-term fluctuations in the conditional variance are mostly driven by negative shocks, while positive shocks have a persistent impact on the variance process. The MIDAS model facilitates the construction of a highly persistent conditional variance process whose persistence differs between positive and negative returns.

In a subsequent paper, Ghysels et al. (2016) present a correction and extension of their 2005 paper. The authors argue that even though the evidence of the risk-return trade-off is low when using the corrected results, the relationship still holds if episodes of financial crises are eliminated from the sample, specifically, the Great Depression, the subprime mortgage financial crisis, and the Great Recession. The authors suggest that during financial crises investment decisions are driven by a flight-to-safety phenomenon, which dominates the long-run risk-return trade-off relationship.

In this paper, we construct the wild bootstrapped and parametric bootstrapped robust t-test proposed in Chapter 1 and Chapter 2 to test the validity of the MIDAS specification proposed by Ghysels et al. (2005, 2016) to characterize the risk-return trade-off. Ghysels et al. (2005, 2016) use the standard t-test to evaluate the explanatory and predictive power of the MIDAS model. Nonetheless, the MIDAS model suffers from weak identification in the sense of Andrews and Cheng (2012, 2014, 2013), which implies that the standard t-test is not valid. For example, testing if $H_0 : \beta = 0$ cannot be established using the standard t-test because under the null hypothesis β and π cannot be separately identified. Testing whether $\beta = 0$ is empirically relevant because it determines the adequacy of the MIDAS specification to model the risk-return trade-off. Moreover, as the rejection rates of the standard t-test are usually larger than the rejection rates of the robust t-test, using the wrong critical values would lead to false positives, i.e. erroneous statistical significance of the estimators composing the MIDAS model. We compare the different outcomes of testing parameter significance and model adequacy using the standard t-statistic, the wild bootstrapped, and parametric bootstrapped robust t-test.

3.2 The MIDAS model under weak identification

Constructing regression models using data sets with variables sampled at different frequencies can be challenging. The usual approach is to adjust the data sets such that all variables are expressed in the duration of the variable with the lowest frequency. Since economic data is commonly sampled at a monthly or quarterly frequency, while financial data is sampled daily or intraday, this approach leads to a loss of a large number of observations. The MIDAS model provides an alternative framework to construct regression models using variables sampled at different frequencies. The model summarizes a large amount of information using a relatively small number of parameters.

The specification proposed by Ghysels et al. (2005) characterizes the risk-return trade-off in monthly or quarterly frequency along with explanatory variables at a daily frequency. The left-hand side of the regression equation contains monthly or quarterly market excess returns, while the right-hand side will be composed by a conditional variance estimator constructed with weighted daily returns. We use monthly and quarterly market excess returns because the risk-return trade-off is considered a long-run relationship that daily and intraday returns cannot capture.

Following Ghysels, Santa-Clara, and Valkanov (2004) we introduce the MIDAS model. Let R_t denote monthly (or quarterly) returns, between the month $t-1$ and t, and r_t denote the daily return at time t, and let $V_t(\pi)$ denote the MIDAS model conditional variance estimator. Following Ghysels et al. (2005, 2016) we construct the conditinoal variance estimator $V_t(\pi)$ using an Almon-lag weighting function,

$$V_t(\pi) = A \sum_{d=0}^{D-1} w(d, \pi_1, \pi_2) r_{t-d}^2 \tag{3.2}$$

$$w(d, \pi_1, \pi_2) = \frac{\exp(\pi_1 d + \pi_2 d^2)}{\sum_{k=0}^{D-1} \exp(\pi_1 k + \pi_2 k^2)} \tag{3.3}$$

By construction, the Almon-lag weighting function is always positive and sums to one. Moreover, the behavior of the weighting function is completely determined by π_1 and π_2.

It can be either increasing or decreasing as well as concave or convex for different values of π_1 and π_2. We set $D = 252$ to construct the conditional variance with the 252 daily observations prior to day t, which is roughly the number of trading days in one year. Notice that the MIDAS specification is not imposing a conditional variance estimator with one year of daily observations. The weighting function will determine endogenously what subset of the 252 observations is significant to capture the risk-return trade-off. The estimates of π_1 and π_2 determine the weighting function that minimizes the sum of squared residuals of Equation (3.4). Another important property of the MIDAS model is parsimony. The conditional variance estimator includes thousands of daily observations which are weighted with only two parameters π_1 and π_2. With a small number of parameters, it is more likely that we can find a relationship that is causal instead of a spurious results caused by model overfitting. The value A is used to express variance in monthly or quarterly terms. As a month has 22 trading days on average, $A = 22$ with monthly data and $A = 66$ with quarterly data.

The MIDAS model of the risk-return trade-off introduced by Ghysels et al. (2005) minimizes the sum of squared residuals of the following non-linear function, which is composed by low frequency returns and a conditional variance estimator constructed with daily returns,

$$R_{t+1} = \zeta + \beta V_t(\pi) + \epsilon_{t+1} \tag{3.4}$$

where $(\zeta, \beta, \pi_1, \pi_2)$ denote the parameters and ϵ_t is a noise process.

The MIDAS specification includes a polynomial function to reduce the dimensionality of the dataset and the number of parameters in the model. Moreover, the polynomial function is advantageous to obtain a data-driven lag selection method. A linear regression model with autoregressive regressors such as ARMA requires that the number of lag has to be set before estimation. The task of choosing the number of lags that are appropriate can be particularly complicated, especially when large amounts of high-frequency data are used. For example, the number of informative lags of the daily stock market index to forecast GDP next period is hard to determine for any type of regression model specification. The MIDAS model proposes the solution of using a flexible polynomial function in which the number of observations being averaged is determined endogenously.

The polynomial function determines the weights of the higher frequency data to affect the lower frequency specification. Even though the procedure is data-driven, we usually obtain a smooth weighting function with its shape determined by the temporal significance of the explanatory variable. In most economic and financial applications, we expect observations that are closer to time period t to be higher weighted compared to observations that are farther away from t. Graphically, the weighting polynomial function would be a decreasing function over time. Nonetheless, it is not clear if the weighting function should have a convex or concave shape. In the convex case, the observations very close to t have a much larger weight than observations that are further away, while in the concave case, observations that are farther away may have a weight close to the value of observations close to t. The shape of the weighting function is determined endogenously by the MIDAS estimation.

Estimating this model using Maximum Likelihood is not appropriate because is unclear what type of probability density function the MIDAS errors follow. To avoid making a strong distributional assumption, we estimate $(\zeta, \beta, \pi_1, \pi_2)$ using Quasi-Maximum Likelihood (QML),

$$\frac{\epsilon_{t+1}(\zeta, \beta, \pi)}{V_t(\pi)} \sim N(0,1) \tag{3.5}$$

$$Q_n(\theta) = n^{-1} \sum_{t=1}^{n} Q_t(\theta) = -\frac{1}{2n} \sum_{t=1}^{n} \left[\log(V_t(\pi) + \frac{\epsilon_t(\theta)^2}{V_t(\pi)} \right] \tag{3.6}$$

Based on the results by White (1982), the QML estimator converges in probability to the pseudo-true value θ^* [1]. Even though the model may be mispecified, the QML estimator is

[1] The estimator is consistent with respect to the pseudo-true value because the log-likelihood is misspecified, and therefore it could be inconsistent with respect to the true value θ_0.

asymptotically normal with the following variance-covariance sandwich estimator.

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} N(0, \tilde{J}^{-1}(\theta^*)\tilde{V}(\theta^*)\tilde{J}^{-1}(\theta^*)) \tag{3.7}$$

$$\tilde{J}(\theta^*) = n^{-1}\sum_{t=1}^{n} \frac{\partial Q_t(\theta^*)}{\partial \theta}\frac{\partial Q_t(\theta^*)}{\partial \theta'} \tag{3.8}$$

$$\tilde{V}(\theta^*) = n^{-1}\sum_{t=1}^{n} \frac{\partial^2 Q_t(\theta^*)}{\partial \theta \partial \theta'} \tag{3.9}$$

The QML esimator is consistent and asymptotically normal under mild regularity conditions, e.g. White (1982); Newey and McFadden (1994). Appendix C.1 presents more details on the estimation method.

3.2.1 Weak Identification

Hypothesis testing of the MIDAS model cannot be employed using the standard t-test because the parameters in the model are potentially weakly identified. If we wish to test the null hypothesis $H_0 : \beta = 0$, the parameter π is not identified under the null. Using the critical values of the standard t-statistic would be incorrect and can lead to false positives. For example, under weak identification, the standard delta method is not valid to obtain the asymptotic distribution of the parameters because the Hessian is converging to a similar matrix.

The inability to identify both parameters under the null leads a t-statistic with non-standard distributions and nuisance parameters. To construct the bootstrapped distributions under weak identification, Chapter 1 proposes a wild bootstrap method in which the bootstrap samples are generated using a multiplier. On the other hand, Chapter 2 proposes a parametric bootstrap method to reduce the number of nuisance parameters and obtain more accurate test sizes and higher power. These bootstrap procedures are not a straight-forward extension. As the estimators can have different distributions according to the value taken along the parameter space, the uniformity conditions required for the bootstrap are no longer valid. The bootstrap methods rely on generating bootstrapped distributions of the t-statistic under weak and strong identification and then combining them to construct robust critical values. The next subsections discuss the properties of the wild bootstrapped and parametric bootstrapped robust t-test.

3.2.2 Wild Bootstrapped robust t-test

The development of these bootstrap methods relies on the ability to pinpoint the cases where weak identification is present. The construction of this bootstrap relies on the asymptotic results of Andrews and Cheng (2012, 2014); Andrews and Mikusheva (2015). The authors present a unified treatment in which they characterize all possible identification categories according to the value parameters take along the parameter space. The authors present four categories of identification which depend on the speed at which the drifting sequence of true parameters is converging. We avoid the technical definitions encompassing the identification categories and refer to Andrews and Cheng (2012, 2014); Andrews and Mikusheva (2015) for an analysis in the asymptotic theory, and we refer Chapter 1 and 2 for two bootstrap methods. We combine the four categories into two groups which we refer to as strong and weak identification. This convention simplifies the construction of the robust critical values, as the four identification categories lead to only two probability distribution that should be analyzed.

The wild bootstrap method follows the bootstrap specification introduced by Wu (1986); Liu et al. (1988) when data is independent and heteroscedastic, while it follows the dependent wild bootstrap specification of Shao (2010) when data is dependent. The wild bootstrap has the benefit of using a multiplier which generates bootstrap samples in a simple way. Moreover, the bootstrap can have much better small sample performance compared to the asymptotic approximations as it can obtain an asymptotic refinement (faster higher order convergence rate in the Edgeworth expansion, e.g. Horowitz (2001)). Under weak identification, the asymptotic distribution of the estimators is asymmetric, which can be captured easily using a multiplier such as the two-point distribution by Mammen (1993). On the other hand, the selection of a wild bootstrap is important because the standard conditions of the bootstrap are not satisfied.

The wild bootstrapped robust t-test of Chapter 1 has important properties. The robust t-test does not require assuming that the identification category is known. The method is used to derive bootstrapped distributions under strong and weak identification. After computing the bootstrapped distribution for each identification category, we construct robust critical values that combine the critical values of both distributions. The results presented in Chapter 1 suggest that the bootstrapped process works well in small samples when nuisance parameters are assumed to be known. Moreover, the wild bootstrap is simpler to estimate than the asymptotic approximation counterpart, especially in multivariate settings with non-linear functional form.

The wild bootstrap method in Chapter 1 is simple to simulate when the parameters of the model are weakly identified as some closed-form expressions of the underlying processes are not required to obtain bootstrapped samples. For example, to obtain the critical values using the asymptotic approximations derived by Andrews and Cheng (2012), an estimator of the variance-covariance of the stochastic process $G(\pi)$ is required. A closed-form expression of this matrix can be complicated to obtain in models with a large number of parameters. The wild bootstrap method can be used to generate bootstrapped samples of the stochastic process $G(\pi)$ using a multiplier.

3.2.3 Parametric Bootstrap robust t-test

One of the limitations of the wild bootstrapped robust t-test is the inability to obtain the correct asymptotic size unless nuisance parameters are known. To improve the empirical sizes of the test, Chapter 2 proposes a parametric bootstrap robust t-test.

The parametric bootstrap in Chapter 2 is similar to the residual bootstrap employed in the econometric literature,Freedman (1981); Mammen (1993); Liu et al. (1988); Horowitz (2001); Davidson and Flachaire (2008). This method relies on generating bootstrap residuals imposing the parametric specification of the model, while subsequently generating bootstrapped data samples from which bootstrapped distributions can be constructed. When parameters are weakly identified, this procedure cannot be employed as residuals depend on estimators that are inconsistent. Nonetheless, the parametric bootstrap process leads to correct inference when the models have weakly identified parameters.

Under weak identification, the parametric bootstrap follows the next steps: 1) obtain residuals that do not depend on π by setting $\beta = 0$, 2) obtain bootstrapped residual samples using a wild bootstrap multiplier as in Wu (1986); Liu et al. (1988); Shao (2010), 3) fix a grid of parameters to generate bootstrapped data samples over each element in the grid, and 4) obtain bootstrapped estimators for each bootstrapped sample along the grid. It is worth mentioning two advantages of this process, 1) the bootstrapped distributions are generated by imposing the model along the grid, even if we do not have prior knowledge of the asymptotic distribution of the estimators, 2) the number of nuisance parameters in the bootstrapped distribution can be reduced because this bootstrap method does not require knowing the value of the nuisance parameter b and in some cases π_1, π_2 (see Chapter 1 and 2). The parametric bootstrap is also used to obtain the bootstrapped t-statistic distribution under strong identification and to construct robust critical values.

The parametric bootstrap leads to a t-statistic whose hypothesis tests with respect to π do not depend on nuisance parameters. The results of Ghysels et al. (2005) indicate that the statistical significance of π validates the explanatory power of the conditional variance estimator constructed with the MIDAS weighting function. In consequence the hypothesis test $H_0 : \pi = 0$ is of particular importance in this study to test the added value of the MIDAS specification.

The empirical exercise of the following section compares the p-values of the t-test by using the wild bootstrapped t-statistic in Chapter 1, the parametric bootstrap robust t-statistic of Chapter 2, and the standard t-statistic considered by Ghysels et al. (2005, 2016). The results of the empirical exercise show that the p-values of the robust t-test and the standard t-test are substantially different, suggesting that the standard t-test leads to false positives.

3.3 Empirical Analysis

Assuming a constant relative risk aversion utility function, the microeconomic and asset pricing literature suggests that $\beta > 1$. Specifically, Arrow (1970) indicate that risk aversion should be almost constant and approximately one. Farber (1978) provides empirical evidence of the coefficient being larger than one. Friend and Blume (1975) present empirical evidence that the estimate of risk aversion should be closer to two. Kydland and Prescott (1982) argue that as the parameter measures the representative consumer's willingness to substitute consumption through time, β should be between one and two to mimic the variability between consumption and investment. Mehra and Prescott (1985) argue that the coefficient of risk aversion exceeds one significantly. The results of many utility models with uncertainty depend on the coefficient of risk aversion being larger or smaller than one. For example, Bansal and Yaron (2004) use an Epstein-Zin utility function, which separates the coefficients of the intertemporal elasticity of substitution and risk aversion, to obtain a risk premium in line with empirical evidence. Their results hinge on using Epstein Zin utility function Ludvigson and Ng (2007) show that the relationship between risk and return can be positive when you summarize a large amount of economic information using a factor model. The numerical values of the coefficient of risk aversion β are of particular interest in this study as they have an economic interpretation. Moreover, the statistical significance of π_1 and π_2 is important to test because determines if the optimized weights under a MIDAS specification have a relevant explanatory power to describe the risk-return trade-off.

The data used for the estimation of the risk-return trade-off using a MIDAS specification consist of the market return portfolio and a risk-free rate. Following the approach of Ghysels et al. (2005, 2016), we use the value-weighted returns including dividends from the Center for Research and Security Prices (CRSP) as a proxy of the market portfolio. The risk-free rate is approximated using the three-month T-bill yield from CRSP. [2]

[2]Daily T-bill yields are not available before 1980, we transform the T-bill yield to daily frequency by assuming that rates stay constant within the month and compounding them.

The estimation exercise derived in this paper covers six subsample periods. The first three subsamples are equivalent to the subsamples analyzed by Ghysels et al. (2005), which include the complete subsample from $1928-2000$ and two subsamples $1928-1963$ and $1964-2000$ which are composed by roughly the same number of observations. The fourth subsample, $1928-2016$, considered by Ghysels et al. (2016), includes the last 16 years of observations. The results of Ghysels et al. (2016) suggest that the risk-return trade-off relationship holds only if the extreme market movements are eliminated from the MIDAS estimation sample. These periods, the authors argue, are characterized by investor's flight to safety instead of the long-run risk-return relationship. To test this hypothesis, the fifth and sixth subsample studied in this paper, characterize the periods $1928-2000$ and $1964-2000$ removing returns that are above or below two times their standard deviation. Specifically, we eliminate all observations such that $R_{t+1} < -2\sigma$ or $R_{t+1} > 2\sigma$, where σ denotes the unconditional standard deviation of R_{t+1}. This approach eliminates roughly 5% of the subsamples. We refer to these sample periods as flight-to-safety (FTS) subsamples and express them as $1928-2000^{FTS}$ and $1964-2000^{FTS}$.

Our choice of truncation for subsamples $1928-2000^{FTS}$ and $1964-2000^{FTS}$ differs from the truncation proposed by Ghysels et al. (2016). In their paper, they choose to truncate all monthly or quarterly returns that are below 1.5 times the unconditional standard deviation. That is, they eliminate all returns in the left tail of the distribution that satisfy the condition $R_{t+1} < -1.5\sigma$. Their approach eliminates around 5% of the number of return observations, which is close to the number of observations truncated using our truncation approach described in the previous paragraph. Nonetheless, their truncation eliminates all observations that are on the left tail, which by construction leads to a positive bias on the parameter estimates. This bias has been studied extensively in regression models with truncated data sets such as the Tobit regression model of Tobin (1958). To avoid this potential bias, we truncate the aberrant observations from the left and right tail of the distribution to analyze the risk-return trade-off around the mean the sample returns.

Table 3.1 and Table 3.2 present the summary statistics of monthly and quarterly market excess returns across the six subsample periods. Surprisingly, the mean and variance of the monthly and quarterly excess market returns change substantially across subsamples. The mean of monthly returns changes from 0.74% between $1928-1963$ to 0.48% in $1964-2000$. Moreover, the variance of monthly returns is roughly three times greater in the $1928-1963$ period than in the $1964-2000$ period. The results indicate that the market excess return volatility process increases substantially during financial crises, especially during the Great Depression and the subprime financial crisis. The negative skewness and high kurtosis of the return subsamples reflect the well-known facts in the financial literature that the unconditional distribution is asymmetric and fat-tailed.

Table 3.1: Summary statistics of monthly excess returns

Monthly	Mean ($\times 10^2$)	Variance ($\times 10^2$)	Skewness	Kurtosis
$1928-2000$	0.613	0.319	-0.125	11.256
$1928-1963$	0.747	0.466	-0.063	10.054
$1964-2000$	0.481	0.181	-0.466	5.153
$1928-2016$	0.589	0.293	-0.161	11.185
$1928-2000^{FTS}$	0.779	0.161	-0.218	2.849
$1964-2000^{FTS}$	0.592	0.148	-0.141	2.910

This table shows the mean, variance, skewness and kurtosis of monthly returns for the six subsamples analyzed. The proxy of the return from the stock market is the value weighted portfolio from the Center for Research and Security Prices (CRSP) and the risk-free rate is the three month Treasury bill from CRSP. FTS denotes Flight-to-Safety subsamples.

3.3.1 Estimation results of the wild bootstrap

We begin by discussing the estimation results of the MIDAS model using monthly excess market returns. Table 3.3 shows the estimated coefficients, their variances, and p-values of the standard t-statistic and the wild bootstrapped robust t-statistic. The p-values denote the probability of rejecting the null hypothesis of the parameter is equal to zero. As with real-world data, the identification category is unknown. We construct the wild bootstrapped robust t-test assuming weak identification to compare the differences in rejection probabilities between the robust and standard t-test.

Table 3.2: Summary statistics of quarterly excess returns

Sample	Mean ($\times 10^2$)	Variance ($\times 10^2$)	Skewness	Kurtosis
$1928-2000$	1.839	1.186	0.737	13.115
$1928-1963$	2.243	1.810	0.899	10.973
$1964-2000$	1.431	0.605	-0.729	4.900
$1928-2016$	1.767	1.086	0.639	13.017
$1928-2000^{FTS}$	2.068	0.575	-0.459	3.216
$1964-2000^{FTS}$	2.072	0.486	-0.066	2.939

This table shows the mean, variance, skewness and kurtosis of quarterly returns for the six subsamples analyzed. The proxy of the return from the stock market is the value-weighted portfolio from the Center for Research and Security Prices (CRSP) and the risk-free rate is the three month Treasury bill from CRSP. FTS denotes Flight-to-Safety subsamples.

The estimator $\hat{\beta}$, which characterizes the coefficient of risk aversion, is highly variable across subsamples, ranging from -0.7 to 3.6. Surprisingly, $\hat{\beta}$ is non-significant for any of the non-FTS subsamples according to the standard or robust t-statistic. Ghysels et al. (2016) argue that to obtain a statistically significant $\hat{\beta}$ we need to remove market crashes from the estimation data. The results of the FTS subsamples, which exclude observations above and below two standard deviations, establish that the standard t-test $\hat{\beta}$ is statistically significant at a 10% level. On the other hand, $\hat{\beta}$ is not significant for any of the FTS subsamples if the robust t-test is considered. In other words, the standard t-test erroneously concludes that the estimator $\hat{\beta}$ is significant for both FTS subsamples. We know that the results of the standard t-test are erroneous because for the null hypothesis $H_0 : \beta = 0$, the parameter π is weakly identified, consequently, the t-statistic does not have a standard normal distribution. Moreover, the standard t-statistic is growing as sample size increases because the variance-covariance matrix is singular under weak identification.

The results of the t-tests with respect to π_1 and π_2 presented in Table 3.3 are harder to interpret because they jointly characterize the shape of the weighting function. For simplicity, we analyze the statistical results of both estimators simultaneously. The estimates $\hat{\pi}_1$ and $\hat{\pi}_2$ characterize a weighting function that decreases with time for all subsamples. In other words, daily market excess returns that are closer to time period t have a higher explanatory power of conditional variance than observations that are farther away from time t. In spite of using one year of daily trading days, the weighting function shows a sharp convex shape that dies out quickly. This implies that observations that are roughly three months prior to t have a very marginal contribution in the estimation, meaning that all observations that are from one year to three months prior to time t are assigned an optimal weight close to zero for the construction of the conditional variance estimator.

In their 2005 paper, Ghysels et al. (2005) argue that the statistical significance of the t-statistic with respect to π, demonstrate the advantage of using a MIDAS specification. The authors compare their results to the model employed by French et al. (1987) which uses an estimator of conditional variance that assigns equal weight to all daily observations. Using the equally weighted variance estimator, French et al. (1987) find non-significant estimates of the risk-return trade-off. The estimates of π in the study by Ghysels et al. (2005) are highly significant. The authors argue that the explanatory power of MIDAS relies on optimally weighting daily market excess returns to generate an informative estimator of conditional variance $V_t(\pi)$. Our results suggest that the statistical significance of the t-tests of $\hat{\pi}_1$ and $\hat{\pi}_2$ found by Ghysels et al. (2005, 2016) are false positives, because there is almost no subsample period where the null hypotheses $H_0 : \hat{\pi}_1 = 0$ and $H_0 : \hat{\pi}_2 = 0$ are both rejected at a 5% level. Using the standard t-test, the MIDAS weighting function is significant at a 5% level for all non-FTS subsamples except for the period $1928-1963$. On the other hand, the robust t-test suggests that the relationship is not significant for any of the non-FTS subsamples except $1964-2000$. The results illustrate the consequences of using the standard t-test in models that have weakly identified parameters. The test leads us to incorrectly conclude that the MIDAS model has statistically significant parameters for most of the non-FTS subsamples. The evidence from our hypotheses tests suggests that daily returns and the MIDAS weighting polynomial do not provide enough information to obtain an accurate measure of conditional variance.

Figure 3.1 shows the scatter plot and regression line of the monthly excess returns MIDAS model estimation. The linear regression model has low explanatory power as the data does not seem to have a linear relationship. The R^2 is very low, sometimes even negative, for all subsamples analyzed in this study. Interestingly, the estimation of the slope estimator $\hat{\beta}$ changes substantially with the elimination of a few observations, which suggests that the model suffers from parameter instability across subsamples. Regression models that are misspecified suffer from estimates that vary considerably with outliers and across subsamples, e.g. (Rousseeuw and Leroy 2005). For example, the exclusion of financial crises changes the estimate of $\hat{\beta}$ from 0.32 to 1.39 between subsamples $1928-2000$ and $1928-2000^{FTS}$, and from 2.13 to 3.69 between subsamples $1964-2000$ and $1964-2000^{FTS}$. The non-statistical significance of the estimators $\hat{\beta}$, $\hat{\pi}_1$ and $\hat{\pi}_2$, as well as the evidence of model misspecification, suggests that the MIDAS model is not capturing the risk-return trade-off relationship appropriately.

Table 3.3 shows that the estimator $\hat{\zeta}$ is significant for most periods. In the framework of the ICAPM, the statistical significance of $\hat{\zeta}$ suggests that the model is misspecified because this coefficient captures the covariance of market returns with respect to state variables not included in the model. The results of Ludvigson and Ng (2007) suggest that incorporating economic and financial control variables reduces misspecification because the conditional mean and variance are estimated more accurately.

The results of the MIDAS estimation for quarterly excess market returns are presented in Table 3.4. The conclusions presented with respect to monthly excess returns are in line with the results obtained using quarterly excess market returns. For example, the standard t-test suggests that the estimator $\hat{\beta}$ is significant at a 5% level in subsamples $1964-2000$ and $1964-2000^{FTS}$, while the robust t-test of $\hat{\beta}$ is non-significant for all subsample periods included in this study.

In conclusion, the wild bootstrapped robust t-test indicates that the MIDAS model is insufficient to obtain a statistically significant risk-return trade-off. As the model suffers from weak identification under the null, the standard t-test erroneously leads to statistical significant estimators when estimation is employed on the FTS subsamples. On the other hand, the robust t-test rejects the statistical significance of at least one of the estimators

$\hat{\beta}, \hat{\pi}_1$ or $\hat{\pi}_2$ in all subsamples analyzed. In the next subsection, we study if these results also hold using the parametric bootstrap robust t-test.

3.3.2 Estimation results of the parametric bootstrap

One of the problems of the wild bootstrapped robust t-test presented in the previous subsection is that the bootstrapped distribution depends on nuisance parameters. The inability to identify this parameter leads to a bootstrapped t-statistic that depends on π, which has to be fixed to perform inference. The parametric bootstrap method can help in this circumstance as it is developed without bootstrapping the limit distributions directly. In this subsection, we analyze the statistical significance of the estimators in the MIDAS model using the parametric bootstrapped t-test.

Inference and testing with nuisance parameters have been a subject of study in the statistics literature for many years. Chernoff and Zacks (1964) introduce a sup-Lagrange multiplier to eliminate nuisance parameters while Davies (1977, 1987) introduce a sup-Likelihood ratio test when the nuisance parameters are present under the alternative hypothesis. Hansen (1996) introduces a transformation to eliminate nuisance parameters that are present under the null. Nuisance parameters that are not identified under the null can be either chosen at random (e.g. White (1989)) or chosen over a fine grid (e.g. Davies (1977, 1987), Hansen (1996), Andrews (1994)). Fixing nuisance parameters over a grid leads to statistical tests with correct level but incorrect size. For example, If we set the level of rejection at 5%, a test with correct test level has a rejection rate of 5% or lower. Moreover, the power of a test is reduced when the underlying distribution includes nuisance parameters.

The simulation exercise of Chapter 1 analyzes the empirical performance of the wild bootstrapped t-test and finds rejection rates that are way below 5% when a grid of nuisance parameters is employed. The empirical test sizes show to be inaccurate compared to the infeasible statistic which assumes that nuisance parameters are known. The hypothesis tests with respect to π are clearly inaccurate as the distribution of the t-test is centered around one nuisance parameter. To improve the test size and power, Chapter 2 proposes a parametric

Table 3.3: Risk-return trade-off wild bootstrapped t-test, monthly frequency.

Monthly		ζ ($\times 10^2$)	β	π_1 ($\times 10^2$)	π_2 ($\times 10^4$)	R^2
$1928 - 2000$	Coeff.	0.522	0.342	-2.971	1.045	0.002
	Std. Dev.	0.196	0.975	1.539	0.565	
	Std. p-value	0.000	0.726	0.045	0.000	
	WB p-value	0.006	0.792	0.114	0.180	
$1928 - 1963$	Coeff.	1.030	-0.775	-1.933	0.755	-0.008
	Std. Dev.	0.300	1.213	1.827	0.688	
	Std. p-value	0.000	0.523	0.270	0.000	
	WB p-value	0.000	0.622	0.292	0.282	
$1964 - 2000$	Coeff.	0.098	2.136	-4.477	1.354	-0.008
	Std. Dev.	0.274	1.692	1.929	0.776	
	Std. p-value	0.583	0.207	0.010	0.000	
	WB p-value	0.926	0.408	0.008	0.020	
$1928 - 2016$	Coeff.	0.522	0.241	-3.635	1.235	0.001
	Std. Dev.	0.173	0.805	1.428	0.528	
	Std. p-value	0.000	0.764	0.009	0.000	
	WB p-value	0.000	0.862	0.110	0.206	
$1928 - 2000^{FTS}$	Coeff.	0.475	1.392	-5.323	1.846	-0.018
	Std. Dev.	0.185	0.818	1.311	0.536	
	Std. p-value	0.000	0.089	0.000	0.000	
	WB p-value	0.002	0.232	0.016	0.038	
$1964 - 2000^{FTS}$	Coeff.	-0.060	3.691	-4.024	1.133	0.022
	Std. Dev.	0.315	1.660	1.686	0.789	
	Std. p-value	0.746	0.026	0.016	0.000	
	WB p-value	0.992	0.096	0.014	0.012	

This table presents the estimates, standard deviation and p-values of the standard and wild bootstrapped t-test of the MIDAS model using monthly frequency. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

Table 3.4: Risk-return trade-off wild bootstrapped t-test, quarterly frequency.

Quarterly		ζ $(\times 10^2)$	β	π_1 $(\times 10^2)$	π_2 $(\times 10^4)$	R^2
$1928-2000$	Coeff.	1.258	0.740	-2.232	0.823	0.009
	Std. Dev.	0.696	1.180	2.177	0.852	
	Std. p-value	0.008	0.530	0.270	0.000	
	WB p-value	0.032	0.638	0.252	0.180	
$1928-1963$	Coeff.	2.754	-0.470	-2.155	1.052	-0.014
	Std. Dev.	1.051	1.407	3.169	1.134	
	Std. p-value	0.000	0.738	0.496	0.000	
	WB p-value	0.010	0.784	0.500	0.158	
$1964-2000$	Coeff.	-0.257	3.435	-3.679	0.433	0.045
	Std. Dev.	0.735	1.707	1.883	0.729	
	Std. p-value	0.616	0.044	0.035	0.271	
	WB p-value	0.940	0.160	0.036	0.274	
$1928-2016$	Coeff.	1.314	0.539	-2.577	0.915	0.007
	Std. Dev.	0.608	0.964	2.128	0.836	
	Std. p-value	0.002	0.576	0.207	0.000	
	WB p-value	0.010	0.708	0.194	0.180	
$1928-2000^{FTS}$	Coeff.	2.023	0.066	-2.313	0.766	-0.003
	Std. Dev.	0.599	0.934	1.686	0.671	
	Std. p-value	0.000	0.943	0.129	0.000	
	WB p-value	0.000	0.960	0.106	0.506	
$1964-2000^{FTS}$	Coeff.	-0.364	5.065	0.009	-2.922	0.079
	Std. Dev.	0.846	1.788	4.023	3.142	
	Std. p-value	0.498	0.005	0.998	0.014	
	WB p-value	0.888	0.064	0.998	0.046	

This table shows the estimates of the model shown in Equation (3.5), standard deviation and p-values of the standard and wild bootstrapped t-test at a quarterly frequency. The conditional variance estimator of returns is calculated using daily returns. The variance of the coefficients is obtained using the sandwich formula of the QML estimator (White 1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row, while the p-values are not. FTS denotes Flight-to-Safety subsamples.

bootstrap that reduces the number of nuisance parameters. The bootstrap data sets are generated sequentially by imposing the null hypothesis and the regression model instead of using the limit distributions. Using this method, π_0 is the only nuisance parameter under the null because if the true value π_0 were known, the model would not suffer from weak identification. Moreover, if π is a scalar value and we wish to test $H_0 : \pi = 0$, imposing the null hypothesis leads to a statistic without nuisance parameters because, under the null, the parameters left to estimate are strongly identified.

Similar to the wild bootstrap case of the previous subsection, we compare the statistical significance of the standard t-statistic to the parametric bootstrapped robust t-statistic. Table 3.5 and Table 3.6 present the results of the parametric bootstrap for the monthly and quarterly frequency. The results of Table 3.5 suggest that the parametric bootstrap method leads to conclusions that are comparable to the wild bootstrapped robust t-test. Using the parametric bootstrapped robust t-test, $\hat{\beta}$ is non-significant for any of the subsamples except for the FTS subsamples. We find $\hat{\beta}$ is statistically significant at a 5% level in the FTS subsamples. Nonetheless, $\hat{\beta}$ is non-significant at a 1% level for any of the subsamples.

The results of the parametric bootstrapped t-test with respect to $\hat{\pi}_1$ and $\hat{\pi}_2$ are more interesting. We find that the null hypothesis $\pi_2 = 0$ is rejected for all subsamples, while $\hat{\pi}_1$ is not significant for any of them. These results suggest that the quadratic term of the weighting function is potentially informative to construct a conditional variance estimator. Using the standard t-test leads to rejection of the null hypothesis at a 5% level for all subsamples except $1928 - 1963$.

Table 3.6 presents the results of the parametric bootstrapped robust t-test using quarterly returns. The only subsamples that are significant with respect to β are $1964 - 2000$ and $1964 - 2000^{FTS}$, similar to the wild bootstrap case. There is no subsample where $\hat{\pi}_1$ and $\hat{\pi}_2$ are both significant at a 5% level. Overall, we conclude that the risk-return trade-off relationship is not captured with the MIDAS model because using the robust t-test, the parameters are non-significant. Nonetheless, the results of the parametric bootstrap suggest that the weighting function itself can be useful to construct a conditional variance estimator because π_2 is significant in all subsamples.

3.4 Alternative MIDAS specification with absolute returns

The parameter instability across subsamples found in the results of the previous section suggests that the model is misspecified. In this section, we propose an alternative MIDAS specification using the absolute value of returns to analyze the risk-return trade-off. This model does not have a microeconomic motivation like the ICAPM model, as it is not derived from a maximization problem with a representative agent. The motivation is the presence of outliers in the MIDAS specification with squared returns.

Forsberg and Ghysels (2007) argue that absolute returns are more informative to predict future increments of quadratic variation than squared returns. The empirical evidence in the paper suggests that realized absolute returns are a better predictor of volatility than realized variance. Moreover, the authors argue that absolute returns have a different persistence dynamics than squared returns and in consequence are more informative to construct forecasts.

We propose a conditional variance estimator of market excess returns constructed by a MIDAS specification with daily absolute returns.

$$R_{t+1} = \zeta + \beta V_t(\pi) + \epsilon_{t+1} \tag{3.10}$$

$$V_t(\pi) = A \sum_{d=0}^{D-1} w(d, \pi_1, \pi_2) \; |r_{t-d}| \tag{3.11}$$

$$w(d, \pi_1, \pi_2) = \frac{\exp(\pi_1 d + \pi_2 d^2)}{\sum_{k=0}^{D-1} \exp(\pi_1 k + \pi_2 k^2)} \tag{3.12}$$

The estimation results of the MIDAS model with absolute returns are presented in Appendix C.2. The results suggest that, similar to the estimation results of the squared returns, the estimators of the MIDAS model are generally non-significant when the wild bootstrapped robust t-test is implemented. Table C.1 shows the results of the estimation using monthly excess returns. The estimator $\hat{\beta}$ is non-significant for any of the subsamples. The standard t-test is non-significant for any of the subsamples except for the FTS subsamples. With respect to $\hat{\pi}_1$ and $\hat{\pi}_2$, both estimators are significant at a 1% level for samples $1928-2000$, $1928-2016$ and $1928-2000^{FTS}$. On the other hand, when you consider the robust wild bootstrapped t-test, we do not find significant results of any of the samples at a 1% level. Similar to the case of squared returns, the risk does not seem to be captured by the MIDAS model specification.

The results of the absolute return MIDAS model with quarterly excess returns are presented in Table C.2. The results are similar to the monthly returns estimation. When the wild bootstrapped robust t-test is considered, we do not find a statistical significance with respect to $\hat{\beta}$. The FTS subsample $1928-2000^{FTS}$ suggests that $\hat{\pi}_1$ and $\hat{\pi}_2$ are significant, but all other subsamples are non-significant. The R^2 does not surpass 0.06 for any of the subsamples considered.

Now we analyze the results of the parametric bootstrapped t-test using the MIDAS model with absolute returns. Table C.3 and Table C.4 suggest that none of the subsamples leads to a MIDAS specification where all parameters are significant. The estimator $\hat{\beta}$ is non-significant with respect to all subsamples except $1964-2000^{FTS}$. On the other hand, the estimator $\hat{\pi}_2$ is significant at a 10% level for all subsamples while $\hat{\pi}_1$ is not significant in any of them.

In conclusion, the results of Tables C.1 to C.4 indicate that the MIDAS specification using absolute excess returns does not improve the fit of the model compared to the specification using squared excess returns. The specification presented in this papers relies on a conditional variance estimator constructed with daily returns. The model can be improved including economic and financial variables that are informative to explain the behavior of conditional variance, e.g. Lettau and Ludvigson (2001b). Moreover, the linear specification of the ICAPM assumes that agents have Constant Relative Risk Aversion (CRRA) preferences. A utility function such as the one proposed by Epstein and Zin (2013) can potentially capture the dynamics of the risk-return trade-off more accurately because it separates the effects of the intertemporal elasticity of substitution and risk aversion.

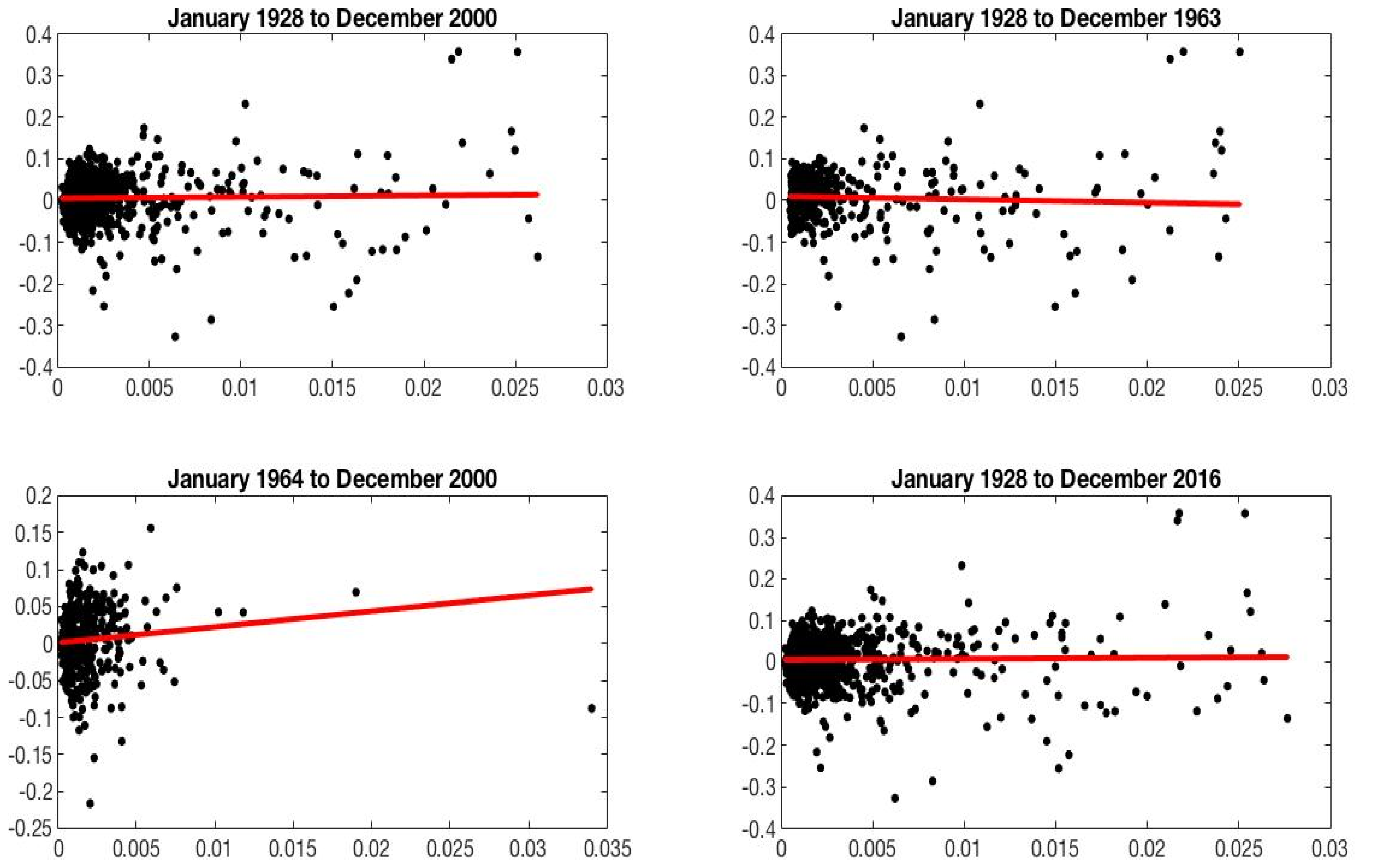

Figure 3.1: Scatter plot and regression line of MIDAS estimation using squared returns, monthly frequency. The figure plots the estimated conditional variance using MIDAS and the return at a monthly frequency for four sample periods. The line is constructed with the estimators obtained from the QML estimation of Equation (3.5) presented in Table 3.3.

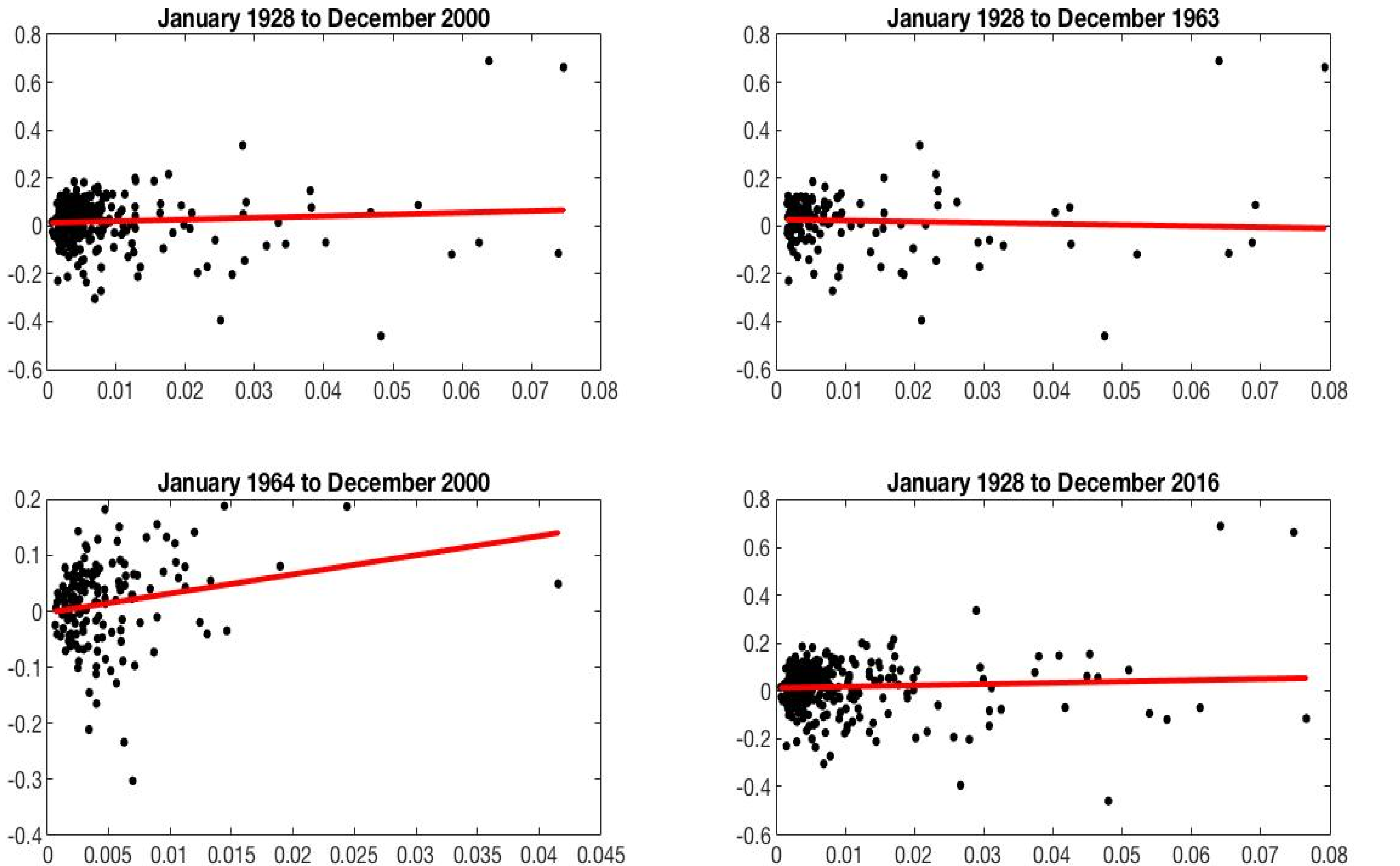

Figure 3.2: Scatter plot and regression line of MIDAS estimation using squared returns, quarterly frequency. The figure plots the estimated conditional variance using MIDAS and the return at a quarterly frequency for four sample periods. The line is constructed with the estimators obtained from the QML estimation of Equation (3.5) presented in Table 3.4.

3.5 Conclusion

Testing the significance of the parameters in a MIDAS model is complicated as the parameters cannot be separately identified under the null hypothesis. In this paper, we study the statistical significance of the MIDAS model proposed by Ghysels et al. (2005) to analyze the risk-return trade-off specification derived by Merton (1973)'s ICAPM. Using a wild bootstrapped and parametric bootstrapped t-test, the empirical results of this study suggest that the parameters of the MIDAS model are non-significant with respect to most subsamples analyzed. Moreover, the standard t-test leads to many cases of false positives. A more realistic model should consider the risk-return trade-off specification using a different stochastic discount factor as well as estimators of conditional variance that include a large number of macroeconomic and financial variables.

Table 3.5: Risk-return trade-off parametric bootstrapped t-test, monthly frequency.

Monthly		ζ ($\times 10^2$)	β	π_1 ($\times 10^2$)	π_2 ($\times 10^4$)	R^2
$1928-2000$	Coeff.	0.521	0.343	-2.999	1.055	0.002
	Std. Dev.	0.196	0.975	1.538	0.565	
	Std. p-value	0.000	0.725	0.043	0.000	
	PB p-value	0.004	0.708	0.482	0.034	
$1928-1963$	Coeff.	1.029	-0.771	-1.952	0.763	-0.008
	Std. Dev.	0.300	1.213	1.829	0.689	
	Std. p-value	0.000	0.525	0.266	0.000	
	PB p-value	0.000	0.550	0.382	0.050	
$1964-2000$	Coeff.	0.098	2.137	-4.476	1.354	-0.008
	Std. Dev.	0.274	1.692	1.929	0.776	
	Std. p-value	0.582	0.207	0.010	0.000	
	PB p-value	0.576	0.192	0.138	0.028	
$1928-2016$	Coeff.	0.524	0.228	-3.668	1.248	0.001
	Std. Dev.	0.173	0.804	1.428	0.529	
	Std. p-value	0.000	0.777	0.008	0.000	
	PB p-value	0.000	0.762	0.552	0.084	
$1928-2000^{FTS}$	Coeff.	0.475	1.393	-5.329	1.849	-0.019
	Std. Dev.	0.185	0.818	1.311	0.536	
	Std. p-value	0.000	0.089	0.000	0.000	
	PB p-value	0.004	0.092	0.162	0.000	
$1964-2000^{FTS}$	Coeff.	-0.061	3.695	-4.016	1.129	0.022
	Std. Dev.	0.315	1.660	1.686	0.789	
	Std. p-value	0.743	0.026	0.016	0.000	
	PB p-value	0.744	0.022	0.118	0.044	

This table presents the estimates, standard deviation and p-values of the standard and parametric bootstrapped t-test of the MIDAS model using monthly frequency. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

Table 3.6: Risk-return trade-off parametric bootstrapped t-test, quarterly frequency.

Quarterly		ζ ($\times 10^2$)	β	π_1 ($\times 10^2$)	π_2 ($\times 10^4$)	R^2
$1928-2000$	Coeff.	1.273	0.706	-2.279	0.842	0.008
	Std. Dev.	0.695	1.179	2.185	0.855	
	Std. p-value	0.007	0.549	0.263	0.000	
	PB p-value	0.012	0.582	0.490	0.052	
$1928-1963$	Coeff.	2.753	-0.460	-2.118	1.039	-0.014
	Std. Dev.	1.051	1.406	3.194	1.142	
	Std. p-value	0.000	0.744	0.506	0.000	
	PB p-value	0.000	0.778	0.714	0.162	
$1964-2000$	Coeff.	-0.256	3.435	-3.678	0.432	0.045
	Std. Dev.	0.735	1.707	1.883	0.730	
	Std. p-value	0.617	0.044	0.035	0.271	
	PB p-value	0.612	0.036	0.084	0.402	
$1928-2016$	Coeff.	1.318	0.529	-2.538	0.899	0.007
	Std. Dev.	0.609	0.965	2.125	0.835	
	Std. p-value	0.002	0.583	0.213	0.000	
	PB p-value	0.002	0.606	0.452	0.084	
$1928-2000^{FTS}$	Coeff.	2.023	0.064	-2.335	0.775	-0.003
	Std. Dev.	0.599	0.934	1.689	0.672	
	Std. p-value	0.000	0.945	0.126	0.000	
	PB p-value	0.000	0.932	0.786	0.584	
$1964-2000^{FTS}$	Coeff.	-0.364	5.067	0.005	-2.919	0.079
	Std. Dev.	0.846	1.788	4.022	3.141	
	Std. p-value	0.497	0.005	0.999	0.015	
	PB p-value	0.478	0.002	0.998	0.118	

This table presents the estimates, standard deviation and p-values of the standard and parametric bootstrapped t-test of the MIDAS model using quarterly frequency. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

APPENDIX A

SUPPLEMENTAL APPENDIX OF "A WILD BOOTSTRAPPED T-TEST ROBUST TO ALL IDENTIFICATION CATEGORIES"

The assumptions for the strongly identified convergence results follow the work of Newey (1991) and Andrews (1994); see also Newey and McFadden (1994). Necessary and sufficient conditions for weak convergence are convergence on finite dimensional distributions and stochastic equicontinuity Davies (1977); Pollard (1990). To show stochastic equicontinuity we assume the functions are Lipschitz, see Newey (1991).

A.1 Proofs of the main results

Proposition 1.3.1 *Suppose that Assumptions (A) to (D) hold. Under* $\{\theta_n\} \in \Theta(\theta_0)$,

(a) When $\beta_0 = 0$*, then* $\sup_{\pi\in\Pi}||\hat{\psi}_n(\pi) - \psi_n|| \xrightarrow{p} 0$*, in particular* $\hat{\psi}_n(\hat{\pi}_n) - \psi_n \xrightarrow{p} 0$.

(b) When $\beta_0 \neq 0$*, then* $||\hat{\theta}_n - \theta_n|| \xrightarrow{p} 0$.

Proof of Proposition 1.3.1

(a) Following Andrews and Cheng (2012), by Assumption A, which implies the minimum of $Q(\theta;\theta_0)$ is unique, and equicontinuity of $Q(\theta;\theta_0)$ for some $\pi \in \Pi$ (equicontinuity proven below), $\exists \epsilon > 0$ s.t. $\forall$ neighborhoods $\psi_0 \in \Psi_0$ and $\forall \pi \in \Pi$

$$\mathbb{P}(\hat{\psi}_n(\pi) \in \Psi(\pi)/\Psi_0 \ \text{ for some } \ \pi) \tag{A.1}$$

$$\leq \mathbb{P}(Q(\hat{\psi}_n(\pi), \pi; \theta_0) - Q(\psi_0, \pi; \theta_0) \geq \epsilon \ \text{ for some } \ \pi) \to 0 \tag{A.2}$$

The first inquality follows as $\beta_0 \neq 0$. The convergence result follows as $n \to \infty$ because: $\sup_{\pi}|Q(\hat{\psi}_n(\pi), \pi; \theta_0) - Q(\psi_0, \pi; \theta_0)| \xrightarrow{p} 0$ which follows as:

$$0 \leq \inf_{\pi\in\Pi}|Q(\hat{\psi}_n(\pi), \pi; \theta_0) - Q(\psi_0, \pi; \theta_0)| \tag{A.3}$$

$$\leq \sup_{\pi\in\Pi} |Q(\hat{\psi}_n(\pi), \pi; \theta_0) - Q(\psi_0, \pi; \theta_0)| \tag{A.4}$$

$$\leq \sup_{\pi\in\Pi} |Q(\hat{\psi}_n(\pi), \pi; \theta_0) - Q_n(\hat{\psi}_n(\pi), \pi)| + \sup_{\pi\in\Pi} |Q_n(\hat{\psi}_n(\pi), \pi) - Q(\psi_0, \pi; \theta_0)| \tag{A.5}$$

$$\leq \sup_{\pi\in\Pi} |Q(\hat{\psi}_n(\pi),\pi;\theta_0) - Q_n(\hat{\psi}_n(\pi),\pi)| + \sup_{\pi\in\Pi} |Q_n(\psi_0,\pi) - Q(\psi_0,\pi;\theta_0)| + o(n^{-1}) \quad \text{(A.6)}$$

$$\leq 2 \sup_{\psi\in\Psi(\pi),\pi\in\Pi} |Q_n(\psi,\pi) - Q(\psi,\pi;\theta_0)| + o(n^{-1}) \xrightarrow{p} 0 \quad \text{(A.7)}$$

The first inequality follows by assumption. The fourth inequality follows by definition of the extremum estimator objective function. And the last last line follows using the same argument as in Lemma A.2.1, that is, we show that the conditions of Theorem 2.1 Newey (1991) are satisfied. We prove: 1) a pointwise law of large numbers, 2) stochastic equicontinuity of $Q_n(\theta)$ and 3) equicontinuity of $Q(\theta;\theta_0)$.

First we show the pointwise law of large numbers. For fixed $\psi(\pi) \in \Psi(\pi)$, by McLeish et al. (1975) we obtain a pointwise law of large numbers using mixing conditions and moments of Assumption A and B. To prove Stochastic Equicontinuity of $Q_n(\psi,\pi) = n^{-1}\sum_{t=1}^{n} \epsilon_t(\psi,\pi)^2$, by Assumption B, notice that as $\epsilon_t(\psi,\pi)$ is continuous with respect to θ and Θ is compact, then $Q_n(\theta)$ is uniformly continuous, and therefore bounded $\forall n$. Also as $d_{\psi,t}(\pi)$ is Lipschitz, then $h(X_{2,t},\pi)$ and $\epsilon_t(\theta)$ are Lipschitz too. By Assumption B, we can find a random variable $C_n = O_p(1)$ s.t.

$$||Q_n(\psi,\pi) - Q_n(\tilde{\psi},\pi)|| \leq C_n||\psi(\pi) - \tilde{\psi}(\pi)|| \qquad \text{w.p.1} \quad \text{(A.8)}$$

By Markov's inequality.

$$\mathbb{P}_{\theta_n}(\sup_{\pi\in\Pi} \sup_{\psi(\pi),\tilde{\psi}(\pi)\in\Psi(\pi),||\psi(\pi)-\tilde{\psi}(\pi)||<\delta} ||Q_n(\psi,\pi) - Q_n(\tilde{\psi},\pi)|| > \eta) \quad \text{(A.9)}$$
$$\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}(\sup_{\pi\in\Pi} \sup_{\psi(\pi),\tilde{\psi}(\pi)\in\Psi(\pi),||\psi(\pi)-\tilde{\psi}(\pi)||<\delta} C_n||\psi(\pi) - \tilde{\psi}(\pi)||) \ \leq \ \frac{\delta}{\eta}\mathbb{E}_{\theta_n}(C_n)$$

By Assumption B(iii), $\mathbb{E}_{\theta_n}(C_n) = O(1)$. Let $\epsilon, \eta > 0$, consider $\delta = \epsilon\eta/\mathbb{E}_{\theta_n}(C_n)$, then

$$\mathbb{P}_{\theta_n}(\sup_{\pi\in\Pi} \sup_{\psi(\pi),\tilde{\psi}(\pi)\in\Theta,||\psi(\pi)-\tilde{\psi}(\pi)||<\delta} ||Q_n(\psi,\pi) - Q_n(\tilde{\psi},\pi)|| > \eta) \ < \ \epsilon \quad \text{(A.10)}$$

which proves stochastic equicontinuity.

To prove equicontinuity of $Q(\theta;\theta_0)$ we notice that in this case $Q(\theta;\theta_0)$ is not a sequence of non-random function, and therefore proving continuity is enough. Recall that $Q(\theta;\theta_0) = \mathbb{E}_{\theta_0}(y_t - \zeta X_{1,t} - \beta h(X_{2,t}), \pi)^2$, which is continuous by construction for each $\psi \in \Psi(\pi)$ and some $\pi \in \Pi$. Moreover, $\Psi(\pi)$ and Π are compact sets in $\mathbb{R}$, which shows equicontinuity of $Q(\theta;\theta_0)$. Notice that as $\hat{\pi}_n \in \Pi$, $\hat{\psi}_n(\hat{\pi}_n) - \psi_n \xrightarrow{p} 0$.

(b) For the semi-strong identification case, the proof is equivalent to the previous case with minor changes. As $\beta_0 \neq 0$, Equation (A.2) follows by replacing $\hat{\psi}_n(\pi)$ with $\hat{\theta}_n$, and $\Psi(\pi)$ with Θ. Equation (A.7) also follows without the supremum or infimum in place. The proofs of pointwise law of large numbers and stochastic equicontinuity follow as in (a), without fixing π. ■

Proposition 1.3.2 *Suppose that Assumptions (A) to (E) hold.*

(a) Under $\{\theta_n\} \in \Theta(\theta_0, 0, b)$ with $||b|| < \infty$, the following holds

$$\begin{pmatrix} \sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n) \\ \hat{\pi}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tau(\pi^*(\theta_0, b); \theta_0, b) \\ \pi^*(\theta_0, b) \end{pmatrix}$$

(b) Under $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, the following holds

$$\sqrt{n}B(\beta_n)(\hat{\theta}_n - \theta_n) \xrightarrow{d} N(0, J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0))$$

Proof of Proposition 1.3.2

(a) To show joint convergence we notice that $\sqrt{n}(\hat{\psi}(\pi) - \psi_n)$ and $\hat{\pi}_n$ are continuous functions of $G_{\psi,n}(\pi)$ and $H_n(\pi)$. The continuity of $\hat{\pi}_n$ follows from the mapping theorem Van Der Vaart and Wellner (1996) as

$$|\arg\min_{\pi\in\Pi} n[Q_n(\hat{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi)] - \arg\min_{\pi\in\Pi} \xi(\pi; \theta_0, b)| \tag{A.11}$$

$$= ||\arg\min_{\pi\in\Pi} \xi_n(\pi, b) - \arg\min_{\pi\in\Pi} \xi(\pi; \theta_0, b)|| \xrightarrow{p} 0 \tag{A.12}$$

in particular, $\hat{\pi}_n \xrightarrow{d} \underset{\pi\in\Pi}{\arg\min}\ \xi(\pi;\theta_0,b)$. Then $\hat{\pi}_n$ can be expressed as a continuous mapping of $G_{\psi,n}(\pi)$ and $H_n(\pi)$.

The processes $G_{\psi,n}(\pi)$ and $H_n(\pi)$ converge jointly as $H_n(\pi)$ converges uniformly to a non-random limit by Lemmas A.2.1 to A.2.3. Then, $\{\sqrt{n}(\hat{\psi}(\pi)-\psi_n),\pi\} \Rightarrow \{\tau(\pi;\theta_0,b),\pi^*(b)\}$ follows by Lemma A.2.5, Van der Vaart (1994) mapping theorem, joint convergence follows as $\pi^* \in int(\Pi)$ which is a compact set. That is,

$$(\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n)-\psi_n),\hat{\pi}_n) \xrightarrow{d} (\tau(\pi^*(\theta_0,b);\theta_0,b),\pi^*(\theta_0,b)) \tag{A.13}$$

(b) For the data generating process proposed, we have the following terms,

$$\begin{aligned}
d_{\theta,t} &= (X_{1,t},h(X_{2,t},\pi),h_\pi(X_{2,t},\pi))' && \text{(A.14)}\\
\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta)Q_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n}[B(\beta)d_{\theta,t}(\pi)d_{\theta,t}(\pi)'B(\beta)-\epsilon_t(\theta)D_t(\theta)]\\
D_t(\theta) &= \begin{pmatrix} 0 & 0 & 0\\ 0 & 0 & h_\pi(X_{2,t},\pi)\\ 0 & h_\pi(X_{2,t},\pi) & h_{\pi,\pi}(X_{2,t},\pi)\beta \end{pmatrix}\\
\frac{\partial}{\partial\theta}Q_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\theta)B(\beta)d_{\theta,t}(\pi)
\end{aligned}$$

And $B(\beta)$ is the selection matrix as in Equation (1.55). Using a first order Taylor expansion

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n-\theta_n) =& B^{-1}(\beta_n)[n^{-1}\sum_{t=1}^{n}[d_{\theta,t}(\theta)d_{\theta,t}(\theta)' - \underbrace{B^{-1}(\beta_n)\epsilon_t(\theta_n)D_t(\beta_n)B^{-1}(\beta_n)}_{(I)}]]^{-1}\\
&\times B^{-1}(\beta_n)[n^{-1/2}\sum_{t=1}^{n}\epsilon_t(\theta_n)B(\beta_n)d_{\theta,t}(\theta_n)]
\end{aligned} \tag{A.15}$$

Term (I) converges in probability to zero as

$$n^{-1}\sum_{t=1}^{n}B^{-1}(\beta_n)\epsilon_t(\theta_n)D_t(\beta_n)B^{-1}(\beta_n) = n^{-1/2}\sum_{t=1}^{n}\epsilon_t(\theta_n)\tilde{D}_t(\beta_n) = Op(1)o_p(1) = o_p(1) \tag{A.16}$$

$$\text{where} \quad \tilde{D}_t(\beta_n) = \begin{pmatrix} 0 & 0 & h_\pi(X_{2,t},\pi)/\sqrt{n}||\beta_n|| \\ 0 & 0 & 0 \\ h_\pi(X_{2,t},\pi)/\sqrt{n}||\beta_n|| & 0 & \beta h_{\pi,\pi}(X_{2,t},\pi)/\sqrt{n}||\beta_n|| \end{pmatrix}$$

as $\tilde{D}_t(\beta_n) \xrightarrow{p} 0_{d_\theta \times d_\theta}$ Equation (A.16) converges in distribution (and therefore in probability) to zero. By Equation (A.15)

$$\sqrt{n}B(\beta_n)(\hat{\theta}_n - \theta_n) = [n^{-1}\sum_{t=1}^{n}[d_{\theta,t}(\theta_n)d_{\theta,t}(\theta_n) + o_p(1)]^{-1}[-1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\theta_n)d_{\theta,t}(\theta_n)] + o_p(1) \tag{A.17}$$

$$\xrightarrow{d} N(0, J(\theta_0)V(\theta_0)J(\theta_0))$$

the first term converges pointwise to $J(\theta_0)$ by McLeish et al. (1975) (Theorem 2.10) law of large numbers as in Proof of Proposition 1.3.1. The second term converges to a standard normal by Wooldridge and White (1988) pointwise central limit theorem. Notice that

$$\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n)d_{\theta,t}(\theta_n)) = 0 \tag{A.18}$$

$$\mathbb{V}_{\theta_n}(\epsilon_t(\theta_n)d_{\theta,t}(\theta_n)) = \mathbb{E}_{\theta_n}(\sigma_t^2(\theta_n)d_{\theta,t}(\theta_n)d_{\theta,t}(\theta_n)') \tag{A.19}$$

$$Cov(\epsilon_t(\theta_n)d_{\theta,t}(\theta_n), \epsilon_{t-j}(\theta_n)d_{\theta,t-j}(\theta_n)) = 0 \tag{A.20}$$

all results follow from the Law of iterated expectations as ϵ_t is a martingale difference. To obtain Wooldridge and White (1988) pointwise central limit theorem, we have that $\sup_t \mathbb{E}_{\theta_n}||\epsilon_t^2 d_{\theta,t}(\theta_n)d_{\theta,t}(\theta_n)'||^{1+\iota} < C < \infty$ by Assumption B(i). By Assumption B(ii) and as they are α-mixing of size $-r/(r-2)$ for $r > 2$. The mixing and moment conditions hold, the central limit theorem follows.

$$1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\theta_n)d_{\theta,t}(\theta_n) \xrightarrow{d} N(0, V(\theta_0)) \tag{A.21}$$

where $V(\theta_0) = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mathbb{E}_{\theta_n}(\sigma_t^2(\theta_n)d_{\theta,t}(\theta_n)d_{\theta,t}(\theta_n)')$. ■

Proposition 1.3.3 *Suppose assumptions (A) to (H) hold*

(a) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$ *and* $\dim(r_\pi(\theta)) = 0$, $T_n \xrightarrow{d} T^\psi(\pi^*(\theta_0, b); \theta_0, b)$

(b) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$ *and* $\dim(r_\pi(\theta)) = 1$, $T_n \xrightarrow{d} T^\pi(\pi^*(\theta_0, b); \theta_0, b)$

(c) Under $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, $T_n \xrightarrow{d} N(0, 1)$

Proof of Proposition 1.3.3

(a) For the case of $d^*_\pi = 0$ we have $[r_\psi(\hat{\theta}_n), 0]$ then

$$r_\psi(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n B^{-1}(\hat{\beta}_n)r_\psi(\hat{\theta}_n)' = r_\psi(\hat{\theta}_n)\Sigma_n r_\psi(\hat{\theta}_n) \tag{A.22}$$

Now taking a mean value expansion around $\hat{\pi}_n$, $r(\psi_n, \hat{\pi}) - r(\psi_n, \pi_n) = r_\pi(\psi_n, \tilde{\pi})(\hat{\pi}_n - \pi_n) + o_p(1) = o_p(1)$ as $r_\pi(\theta) = 0$, where $\tilde{\pi}_n$ is a mean value between $\hat{\pi}_n$ and π_n. This implies $r(\psi_m, \hat{\pi}_n) - r(\psi_n, \pi_n) = o_p(1)$. Taking a mean value expansion of $\hat{\psi}_n$

$$\begin{aligned} r(\hat{\theta}_n) - r(\theta_n) &= r(\hat{\psi}_n, \hat{\pi}_n) - r(\psi_n, \hat{\pi}_n) + r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n) && \text{(A.23)} \\ &= r_\psi(\tilde{\psi}_n, \hat{\pi}_n)(\hat{\psi}_n(\hat{\pi}_n) - \psi_n) + o_p(1) && \text{(A.24)} \end{aligned}$$

here $\tilde{\psi}_n$ is a mean value between $\hat{\psi}_n$ and ψ_n. Taking the numerator and denominator and using the expressions from above.

$$\begin{aligned} T_n &= \frac{r_\psi(\tilde{\psi}_n, \hat{\pi}_n)\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n)}{[r_\psi(\hat{\theta})\hat{\Sigma}_n r_\psi(\hat{\theta})']^{1/2}} + o_p(1) && \text{(A.25)} \\ &= \frac{r_\psi(\psi_n, \hat{\pi}_n)\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n)}{[r_\psi(\psi_n, \hat{\pi}_n)\hat{\Sigma}_n r_\psi(\psi_n, \hat{\pi}_n)']^{1/2}} + o_p(1) \\ &= T_{\psi,n}(\hat{\pi}_n) + o_p(1) \end{aligned}$$

This follows as $\hat{\psi}_n$ is uniformly consistent over $\pi \in \Pi$ and Assumption F, because $\tau(\pi)$ and $\hat{\pi}_n$ can be written as continuous functions of $G_{\psi,n}(\pi)$. By Proposition 1.3.2, we have $T_{\psi,n}(\hat{\pi}_n) + o_p(1) \xrightarrow{d} T_\psi(\pi^*(\theta_0, b); \theta_0, b)$ which proves (a). which follows as $\tau_n(\pi)$ and $\hat{\pi}_n$ can be written as continuous functions of $G_{\psi,n}$.

(b) First we notice that

$$
\begin{aligned}
r_\theta(\hat{\theta}_n)B^{-1}(\hat{\beta}_n) &= [r_\psi(\hat{\theta}_n), ||\beta_n||^{-1}r_\pi(\hat{\theta}_n)] \\
&= ||\beta_n||^{-1}[r_\psi(\hat{\theta}_n)\hat{\beta}_n, r_\pi(\hat{\theta}_n)] \\
&= ||\beta_n||^{-1}([0, r_\pi(\hat{\theta}_n)] + o_p(1)) \qquad \text{(A.26)}
\end{aligned}
$$

Now take a mean value expansion with respect to ψ for fixed $\hat{\pi}_n$, $r(\hat{\theta}_n) = r(\psi_n, \hat{\pi}_n) + r_\psi(\tilde{\psi}_n, \hat{\pi}_n)(\hat{\psi}_n(\hat{\pi}_n) - \psi_n)$. Letting $\tilde{\psi}_n$ denote the mean value:

$$
\sqrt{n}||\hat{\beta}_n||(r(\hat{\theta}_n) - r(\theta_n)) \qquad \text{(A.27)}
$$

$$
= \sqrt{n}||\hat{\beta}_n||(r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n)) + ||\hat{\beta}_n||r_\psi(\tilde{\psi}_n, \hat{\pi}_n)\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n) \qquad \text{(A.28)}
$$

$$
= \sqrt{n}||\hat{\beta}_n||[r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n)] + o_p(1) \qquad \text{(A.29)}
$$

which follow by uniform consistency of $\hat{\psi}_n$ (and therefore of $\tilde{\psi}_n$) and the fact that $\hat{\beta}_n = o_p(1)$, $\sqrt{n}(\hat{\psi}_n - \psi_n) = O_p(1)$ and $r_\psi(\tilde{\psi}_n, \hat{\pi}_n) = O_p(1)$. From the expression from above we obtain.

$$
\begin{aligned}
T_n &= \frac{||\sqrt{n}\hat{\beta}_n||(r(\hat{\theta}_n) - r(\theta_n))}{[r_\pi(\hat{\theta}_n)\hat{\Sigma}_n r_\pi(\hat{\theta}_n)]^{1/2}} + o_p(1) \qquad \text{(A.30)} \\
&= \frac{||\sqrt{n}\hat{\beta}_n||(r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n))}{[r_\pi(\psi_0, \hat{\pi}_n)\hat{\Sigma}_n r_\pi(\psi_0, \hat{\pi}_n)]^{1/2}} + o_p(1) \\
&= T_{\pi,n}(\hat{\pi}_n) + o_p(1)
\end{aligned}
$$

By the joint convergence result in Proposition 1.3.2, $T_{\pi,n}(\hat{\pi}_n) \xrightarrow{d} T_\pi(\pi^*(\theta_0, b); b, \theta_0)$

(c) By Proposition 1.3.2 and the delta method

$$
\sqrt{n}(r(\hat{\theta}_n)) - r(\theta_n)) \xrightarrow{d} r_\theta(\theta_0)B^{-1}(\hat{\beta}_0)\Sigma(\theta_0)B^{-1}(\hat{\beta}_0)r_\theta(\theta_0) \qquad \text{(A.31)}
$$

Rearranging terms and the fact that parameters are consistent under semi-strong identification Proposition 1.3.1 and Lemma A.2.7.

$$\frac{\sqrt{n}(r(\hat{\theta}_n)) - r(\theta_n))}{(r_\theta(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n B^{-1}(\hat{\beta}_n)r_\theta(\hat{\theta}_n))^{1/2}} \xrightarrow{d} N(0,1) \tag{A.32}$$

which proves Proposition 1.3.3. ■

Theorem 1.4.1 *Suppose that Assumptions (A) to (E) hold. Under weak identification, let* $\hat{\pi}_n^m$ *be constructed using Equation* (1.39) *and* $\hat{\tau}_n^m(\hat{\pi}_n^m; \pi_0, b)$ *be constructed using Equation* (1.40)*, while under strong identification let* $\hat{\tau}_{\theta,n}^m(\hat{\theta}_n)$ *be constructed using Equation* (1.42)*. Denote* $c_{n,1-\alpha}^{m,a}$ *and* $c_{1-\alpha}^a$ *with* $a = \pi, \tau, \tau_\theta$ *be the critical values of the bootstrapped and asymptotic distributions, respectively. Letting* $M_n \to \infty$ *as* $n \to \infty$ *then*

(a) Under $\{\theta_n\} \in \Theta(\theta_0, 0, b)$ *with* $||b|| < \infty$, $|c_{n,1-\alpha}^{m,\tau} - c_{1-\alpha}^{\tau}| \xrightarrow{p} 0$ *and* $|c_{n,1-\alpha}^{m,\pi} - c_{1-\alpha}^{\pi}| \xrightarrow{p} 0$.

(b) Under $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, $|c_{n,1-\alpha}^{m,\tau_\theta} - c_{1-\alpha}^{\tau_\theta}| \xrightarrow{p} 0$.

Proof of Theorem 1.4.1

We prove weak convergence of each individual term and then prove joint convergence using the Cramer Wold Device.

First we show $\hat{\pi}_n^m \xrightarrow{d} \pi^*(b)$. By Lemma A.2.1, Lemma A.2.2 and Lemma A.2.9 it follows that $\{\xi_n^m : \pi \in \Pi\} \Rightarrow \{\xi : \pi \in \Pi\}$ where

$$\xi_n^m(\pi, b) = -\frac{1}{2}\{G_{\psi,n}^m(\pi) + K_n(\pi)b\}' H_n^{-1}(\pi)\{G_{\psi,n}^m(\pi) + K_n(\pi)b\}' \tag{A.33}$$

which follows by the continuous mapping theorem as ξ_n^m is a continuous function of $G_{\psi,n}^m, H_n, K_n$ and the last two have non-random uniform limits. By mapping theorem (Van der Vaart (1994))

$$\arg\min_{\pi \in \Pi} \xi_n^m(\pi, b) \xrightarrow{d} \arg\min_{\pi \in \Pi} \xi_\psi(\pi, b) \tag{A.34}$$

which proves the first claim.

Now we show $\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n^m) - \psi_n) \xrightarrow{d} \tau(\pi^*; \theta_0, b)$. Considering the equivalent expression from Lemma A.2.5,

$$\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) = -H_n^{-1}(\pi)(G_{\psi,n}^m(\pi) + K_n(\psi_{0,n})\sqrt{n}\beta_n) + (\psi_n - \psi_{0,n}) + o_{p,\pi}(1) \Rightarrow \tau(\pi; \theta_0, b) \tag{A.35}$$

The argument is the same as Step 1, the function τ is a continuous function of $G_{\psi,n}^m, H_n, K_n$ with the last two having nonrandom limits. Finally, as $\hat{\pi}_n^m$ converges in distribution to $\pi^*(\theta_0, b)$, we conclude $\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n^m) - \psi_n) \xrightarrow{d} \tau(\pi^*; \theta_0, b)$.

Joint convergence follows form the Cramer Wold Device. Let $\lambda_1, \lambda_2 \in \mathbb{R}$.

$$\lambda_1\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n^m) - \psi_n) + \lambda_2\hat{\pi}_n^m = \lambda_1\tau_n(\hat{\pi}_n^m) + \lambda_2\hat{\pi}_n^m \xrightarrow{d} \lambda_1\tau(\pi^*(\theta_0, b)) + \lambda_2\pi^*(\theta_0, b) \tag{A.36}$$

by the Continuous Mapping Theorem.

Now we show consistency of the critical values. Recall the definitions. Define $c_{1-\alpha}^{\tau} = \inf\{z \in \mathbb{R}^{d_\psi} : \mathbb{P}_{\theta_0}(\tau(\pi^*(\theta_0, b); \theta_0, b) \leq z) \geq 1 - \alpha\}$ and $c_{1-\alpha}^{\pi} = \inf\{z \in \mathbb{R}^{d_\pi} : \mathbb{P}_{\theta_0}(\pi^*(\theta_0, b) \leq z) \geq 1 - \alpha\}$ be the $1 - \alpha$ asymptotic critical values of $\tau(\pi^*(\theta_0, b); \theta_0, b)$ and $\pi^*(\theta_0, b)$ respectively. To obtain the critical values of the bootstrapped statistics, let M_n denote the number of bootstrapped samples. For the sequence of bootstrapped distributions $\{\hat{\tau}_n^m\}_{m=1}^{M_n} = \{\tau_n^m(\hat{\pi}_n^m; \pi_0, b)\}_{m=1}^{M_n}$, define the order statistics $\hat{\tau}_n^{[1]} \leq \hat{\tau}_n^{[2]} \leq ... \leq \hat{\tau}_n^{[M_n]}$. The approximated $1 - \alpha$ critical value of $\{\hat{\tau}_n^m\}_{m=1}^{M_n}$ is defined by $c_{n,1-\alpha}^{m,\tau} = \hat{\tau}_n^{[(1-\alpha)M_n]}$. Similarly for $\{\hat{\pi}_n^m\}_{m=1}^{M_n}$, we define $c_{n,1-\alpha}^{m,\pi} = \hat{\pi}_n^{[(1-\alpha)M_n]}$. Under strong identification, let $c_{1-\alpha}^{\tau_\theta}$ be the $1 - \alpha$ critical value of the distribution $N(0, J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0))$. Also let $\{\hat{\tau}_{\theta,n}^m\}_{m=1}^{M_n} = \{\tau_{\theta,n}^m(\hat{\theta}_n)\}_{m=1}^{M_n}$ be the bootstrapped samples with order statistics $\hat{\tau}_n^{[1]} \leq \hat{\tau}_n^{[2]} \leq ... \leq \hat{\tau}_n^{[M_n]}$, and $1 - \alpha$ critical values $c_{n,1-\alpha}^{m,\tau_\theta} = \hat{\tau}_{\theta,n}^{[(1-\alpha)M_n]}$.

To prove consistency of the critical values, we condition with respect to the sample W_t. Under H_0, the bootstrapped draws $\{\tau_n^m\}_{m=1}^{M_n}$ and $\{\pi_n^m\}_{m=1}^{M_n}$ are independent and identically distributed. Moreover, each of them converges weakly to the asymptotic distributions $\tau(\pi^*(\theta_0, b); \theta_0, b)$ and $\pi^*(\theta_0, b)$ respectively by Equation (A.34) and (A.35). Hence by the

Glivenko Cantelli Theorem.

$$\sup_{z\in\mathbb{R}^{d_\tau}} |\mathbb{P}_{\theta_0}(\hat{\tau}_n^m \leq z|W_t) - \mathbb{P}_{\theta_0}(\tau(\pi^*(\theta_0,b);\theta_0,b) \leq z)| \xrightarrow{p} 0 \tag{A.37}$$

$$\sup_{z\in\mathbb{R}^{d_\pi}} |\mathbb{P}_{\theta_0}(\hat{\pi}_n^m \leq z|W_t) - \mathbb{P}_{\theta_0}(\pi^*(\theta_0,b) \leq z)| \xrightarrow{p} 0 \tag{A.38}$$

as $M_n \to \infty$. By the continuous mapping theorem, we can express $c_{n,1-\alpha}^{m,\tau} = c_{n,1-\alpha}^{[1]} + o_p(1) = c_{1-\alpha}^{\tau} + o_p(1)$, where $c_{n,1-\alpha}^{[1]}$ denotes the $1-\alpha$ critical value of $\tau_n^{[1]}$. It follows that $|c_{n,1-\alpha}^{m,\tau} - c_{1-\alpha}^{\tau}| \xrightarrow{p} 0$ with $M_n \to \infty$ as $n \to \infty$. By the same argument $|c_{n,1-\alpha}^{m,\pi} - c_{1-\alpha}^{\pi}| \xrightarrow{p} 0$ with $M_n \to \infty$ as $n \to \infty$.

(b) Notice that, $T_{\theta,n}^m = \tau_{\theta,n}^m(\theta)$. By Lemma A.2.10 and the Delta method.

$$\sqrt{n}B(\hat{\beta}_n)(r(\hat{\theta}_n^m)) - r(\theta_n)) = r_\theta(\hat{\theta}_n)\tau_{\theta,n}^m(\hat{\theta}_n) \xrightarrow{d} N(0, r_\theta(\theta_0)' J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0)r_\theta(\theta_0)) \tag{A.39}$$

Under $\theta_n \in \Theta(\theta_0,\infty,\omega_0)$, we have $\hat{\theta}_n \xrightarrow{p} \theta_0$. Also by Lemma A.2.7 the variance-covariance estimator converges in probability to the true variance under the proper normalization.

$$\frac{\sqrt{n}(r(\hat{\theta}_n)) - r(\theta_n))}{(r_\theta(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)\hat{\Sigma}_n B^{-1}(\hat{\beta}_n) r_\theta(\hat{\theta}_n))^{1/2}} \xrightarrow{d} N(0,1) \tag{A.40}$$

which shows convergence in distribution. Consistency of the critical values follows from the Glivenko Cantelli theorem, as the bootstrapped samples are i.i.d. draws following the same argument as with Equation (A.37) and (A.38). ■

Theorem 1.4.2 *Suppose assumptions (A) to (H) hold. Let $T_{\psi,n}^m$, $T_{\pi,n}^m$ be the bootstrapped t-statistics of Equation* (1.44) *to* (1.46). *Denote $c_{n,1-\alpha}^{m,a}$ and $c_{1-\alpha}^{a}$ with $a = T^\psi, T^\pi, T^\theta, N$ the $1-\alpha$ critical values of the bootstrapped and asymptotic distributions, respectively. Let $M_n \to \infty$ as $n \to \infty$*

(a) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ with $||b|| < \infty$ and $\dim(r_\pi(\theta)) = 0$, $|c_{n,1-\alpha}^{m,T^\psi} - c_{1-\alpha}^{T^\psi}| \xrightarrow{p} 0$

(b) Under $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ with $||b|| < \infty$ and $\dim(r_\pi(\theta)) = 1$, $|c_{n,1-\alpha}^{m,T^\pi} - c_{1-\alpha}^{T^\pi}| \xrightarrow{p} 0$

(c) *Under* $\{\theta_n\} \in \Theta(\theta_0, \infty, \omega_0)$, $|c^{m,T^\theta}_{n,1-\alpha} - c^N_{1-\alpha}| \xrightarrow{p} 0$

Proof of Theorem 1.4.2

The proofs follow similar arguments to Proposition 1.3.3, using different supporting lemmas from Appendix A.2.

(a) Follows immediately from Proposition 1.3.3, Theorem 1.4.1, specifically in Equation (A.24), we replace $\hat{\psi}_n$ by with $\hat{\psi}^m_n$ and notice that the denominator converges to the same (non-random) probability limit and using Lemmas A.2.1, A.2.2, A.2.8 and A.2.9. (b) Same argument as in (i) but now use Equation (A.29) and use results Lemmas A.2.1, A.2.2, A.2.8 and A.2.9 (c) The result of Proposition 1.3.3c) is enough as we have convergence a standard Normal by Lemmas A.2.7 and A.2.10. The quantile function is a simple transformation to obtain consistent critical values. ■

Theorem 1.4.3 *Let* $M_n \to \infty$ *as* $n \to \infty$. *Under the null hypothesis* $H_0 : r(\theta) = q$, *the* LF *and* ICS_1 *critical values of the t-test have the correct asymptotic size with probability approaching one,*

(a) $AsySz^{LF,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c^{LF,m}_{n,1-\alpha}(r(\theta))) = 1-\alpha$

(b) $AsySz^{ICS_1,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c^{ICS_1,m}_{n,1-\alpha}(r(\theta))) = 1-\alpha$

(c) *If* H_0 *is false, that is* $r(\theta_n) \neq q$, *then* $T_n(r(\theta)) \xrightarrow{p} \infty$

Proof of Theorem 1.4.3

(a) This proof follows the lines of Hill(2017) and Lemma 2.1 of Andrews and Cheng (2012). We work with the absolute value of the t-statistic to focus on one sided critical values. The proof for the t-statistic with two sided critical values is analogous. Define the sample quantile function $c^{m,a}_n(p,u) \equiv \inf\{c \geq 0 : \mathbb{P}_m((T^{m,a}_n(p))^2 \leq c) \geq u\}$, $c^a(p,u) \equiv \inf\{c \geq 0 : \mathbb{P}_{\theta_0}((T^a(p))^2 \leq c) \geq u\}$ and $c^a_n(p,u) \equiv \inf\{c \geq 0 : \mathbb{P}_{\theta_n}((T^a_n(p))^2 \leq c) \geq u\}$ where $a = \psi, \pi$ and $p \in \mathcal{P}$. By Proposition 1.3.3, under weak identification we have that $c_n(p,u) \to c(p,u), \forall p \in \mathcal{P}$ and $u \in [0,1]$. Conditioning with respect to W_t, we obtain that the only source of randomness in the critical values comes from stochastic processes $G^m_{\psi,n}$ and G_ψ.

By Theorem 1.4.2 (c), using the probability measure conditional on W_t, we obtain $\sup_{c\geq 0}|\mathbb{P}_m((T_n^{m,a}(p))^2 \leq c) - \mathbb{P}_{\theta_n}((T^a(p))^2 \leq c)| \xrightarrow{p} 0$ and consequently $\sup_{u\in[0,1]}|c_n^{m,a}(p,u|W_t) - c_n^a(p,u)| \xrightarrow{p} 0$, $\forall p \in \mathcal{P}$, under semi-strong identification when $n \to \infty$ and $M_n \to \infty$, as the bootstrapped draws are independent. Similarly, we have that $\sup_{u\in[0,1]}|c_n^{m,a}(p,u|W_t) - c^a(p,u)| \xrightarrow{p} 0, \forall p \in \mathcal{P}$ as $n \to \infty$ which follows from Theorem 1.4.2 (c). We combine the three results to obtain $|c_n^{m,a}(p,u|W_t) - c^a(p,u)| \xrightarrow{p} 0, \forall p \in \mathcal{P}$. The critical values of the $|N(0,1)|$ under semi-strong identification are the asymptotic ones, therefore they are equal to $c^N_{1-\alpha}$. By the continuous mapping theorem of the max function, we obtain $|c^{LF,m}_{n,1-\alpha}(p) - c^{LF}_{1-\alpha}(p)| \xrightarrow{p} 0$ $\forall p \in \mathcal{P}$.

We have shown the bootstrapped critical values are valid for each identification category, strong, semi-strong and weak identification under $H_0 : r(\theta) = q$. It is left to show that the asymptotic size is equal to α when using the asymptotic critical values as we shown $|c^{LF^a,m}_{n,1-\alpha}(p) - c^{LF^a}_{1-\alpha}(p)| \xrightarrow{p} 0$, $\forall p \in \mathcal{P}$, where $a = \psi, \pi$. For the reminder of the proof we use the asymptotic Least Favorable critical values $c^{LF,m}_{n,1-\alpha}$ and follow the argument of Andrews and Cheng (2012). The Least Favorable asymptotic critical values take the form $c^{LF^a,m}_{n,1-\alpha} = \max\{\sup\limits_{p\in\mathcal{P}} c^{m,T^a}_{n,1-\alpha}(p), c^m_{n,\theta,1-\alpha}\}$. The asymptotic size with respect to the asymptotic critical values is

$$AsySz^{LF^a,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta:r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c^{LF^a,m}_{n,1-\alpha}(r(\theta))) \tag{A.41}$$

$$= \min\{\mathbb{P}_{\theta_0}(T^a(r(\theta)) \leq c^{LF^a}_{1-\alpha}), \mathbb{P}_{\theta_0}(|N(0,1)| \leq c^{LF^a}_{1-\alpha})\} + o_p(1) \tag{A.42}$$

which follows form Lemma 2.1 of Andrews and Cheng (2012), and T^a denotes the asymptotic distribution of either T^ψ or T^π. The second argument in Equation (A.42) follows because under strong identification the t-statistic is pivotal. The asymptotic size of the Equation (A.42) is greater or equal to $1-\alpha$ because $\mathbb{P}_{\theta_0}(|N(0,1)| \leq c^{LF^a}_{1-\alpha}(p)) \geq \mathbb{P}_{\theta_0}(|N(0,1)| \leq c^{|N(0,1)|}_{1-\alpha}) =$

$1-\alpha$, where the second critical value denotes the critical value of the $|N(0,1)|$ distribution. Similarly, for the first critical value of Equation (A.42), $\mathbb{P}_{\theta_0}(T^a(r(\theta)) \leq c_{1-\alpha}^{LF^a}(p)) \geq \mathbb{P}_{\theta_0}(T^a(r(\theta)) \leq c_{1-\alpha}^{T^a}(p)) = 1-\alpha$, $\forall p \in \mathcal{P}$ where the second critical value denotes the critical value of $T^a(p)$ for each fixed $p \in \mathcal{P}$.

On the other hand, the critical values are less than or equal to $1-\alpha$ because if $c_{1-\alpha}^{LF^a} = c_{1-\alpha}^{|N(0,1)|}$ then $\mathbb{P}_{\theta_0}(|N(0,1)| \leq c_{1-\alpha}^{LF^a}) = 1-\alpha$ and if $c_{1-\alpha}^{LF^a} > c_{1-\alpha}^{|N(0,1)|}$ then $\mathbb{P}_{\theta_0}(T^\theta(p_{max}) \leq c_{1-\alpha}^{LF}) = \mathbb{P}_{\theta_0}(T^\theta(p_{max}) \leq c_{1-\alpha}^{T^\theta}(p_{max}))$ by Assumption H. Then $AsySz^{LF,m} = 1-\alpha$ w.p.a.1.

(b) The proof is equivalent to proof of Theorem 1.4.3 arguing that under weak identification, $\mathbb{P}(A_n \leq \kappa_n) \xrightarrow{p} 1$ and under any other case $\mathbb{P}(A_n \leq \kappa_n) \xrightarrow{p} 0$ as $M_n \to \infty$ with n as $n \to \infty$. This implies that in either case, for large enough m and n, $c^{ICS_1}(p,b) = c^{LF}(p,b) + o_p(1)$, $\forall p \in \mathcal{P}$.

(c) First consider T_n, only for restrictions of ψ, that is with $\dim(r_\pi) = 0$. If H_0 is false $r(\theta_n) \neq q_n$ and by Equation (A.23) of Proposition 1.3.3 replacing it with $\hat{\theta}_n = (\hat{\zeta}_n', \hat{\beta}_n', \hat{\pi}_n')'$

$$r(\hat{\theta}_n) - r(\theta_n) = r(\hat{\theta}_n) - r(\theta_n) + r(\theta_n) - q_n = CI_k + o_p(1) \tag{A.43}$$

for some $C \neq 0$ and identity matrix I_k and large enough n as $||\hat{\psi}_n - \psi_n|| \xrightarrow{p} 0$ (Theorem 1.4.1). Then we have

$$T_n = \frac{\sqrt{n}C}{r_\psi(\hat{\theta}_n)\Sigma_n r_\psi(\hat{\theta}_n)} + o_p(1) \to \infty \text{ as } n \to \infty \tag{A.44}$$

as the denominator converges in probability to $r_\psi(\theta_0)\Sigma(\theta_0)r_\psi(\theta_0)$ which is finite and non-random.

Similarly consider T_n with $\dim(r_\pi) = 1$,

$$r(\hat{\theta}_n) - r(\hat{\theta}_n) = r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n) + r(\psi_n, \pi_n) - q_n = CI_k + O_p(1) \tag{A.45}$$

Using the arguments of the proof in Proposition 1.3.3, specifically by Equation (A.30) and

Equation (A.27) - Equation (A.29)

$$T_n = \frac{||\sqrt{n}\hat{\beta}_n||(r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n))}{[r_\pi(\psi_0, \hat{\pi}_n)\Sigma_n r_\pi(\psi_0, \hat{\pi}_n)]^{1/2}} + o_p(1) \tag{A.46}$$

$$= \frac{\sqrt{n}C}{r_\psi(\hat{\theta}_n)\Sigma_n r_\psi(\hat{\theta}_n)} + o_p(1) \xrightarrow{p} \infty \text{ as } n \to \infty \tag{A.47}$$

as the denominator converges in probability to $r_\pi(\theta_0)\Sigma(\theta_0)r_\pi(\theta_0)$ which is finite and non-random. ■

A.2 Supporting Lemmas

The following lemmas are necessary for the resutls in the previous section of the Appendix.

Lemma A.2.1 *Under* $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$*,* $\sup_{\pi\in\Pi}||H_n(\pi) - H(\pi;\theta_n)|| \xrightarrow{p} 0$

Proof. Recall that $||\cdot||$ denotes the l_1 norm. To prove the Uniform Law of Large Numbers (uniform law of large numbers) we use Theorem 2.1 of Newey (1991). We need four conditions to be satisfied: i) Π compact, ii) Pointwise convergence $|H_n(\pi) - H(\pi;\theta_n)| \xrightarrow{p} 0$, for fixed $\pi \in \Pi$, iii) Stochastic Equicontinuity of $H_n(\pi)$, and iv) Equicontinuity of $H(\pi, \theta_n)$.

i) Is satisfied by Assumption D.

ii) To prove pointwise Law of Large Numbers, let $\pi \in \Pi$. We use McLeish et al. (1975) (Theorem 2.10) law of large numbers. By Assumption B(i) we have that $H_n(\theta)$ is mixing size $-r/(2r-1)$ for $r \geq 1$, as functions of mixing are mixing too. Moreover we show that for some $\delta > 0$

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}_{\theta_n}(||H_n(\pi) - H(\pi;\theta_n)||)}{k^{1+\delta}} \tag{A.48}$$

which follows as

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}_{\theta_n}(||H_n(\pi) - H(\pi;\theta_n)||)}{k^{1+\delta}} \leq \sum_{k=1}^{\infty} \frac{\mathbb{E}_{\theta_n}(||H_n(\pi)|| + \mathbb{E}_{\theta_n}||H(\pi;\theta_n)||)}{k^{1+\delta}} \leq C_1 \sum_{k=1}^{\infty} \frac{1}{k^{1+\delta}} < C_2 \tag{A.49}$$

for some constants C_1, C_2, as each $d_{\psi,t}(\pi)$ are uniformly L_p bounded for $p = 4 + \iota$ by Assumption B(ii).

iii) To prove Stochastic Equicontinuity of H_n, notice that as $d_{\psi,t}(\pi)$ is continuous and Π is compact, then $H_n(\pi)$ is uniformly continuous and by Assumption B it is bounded $\forall n$. Also as $d_{\psi,t}(\pi)$ is Lipschitz then by Assumption B

$$||H_n(\pi) - H_n(\tilde{\pi})|| = C_n||\pi - \tilde{\pi}|| \ w.p.1 \tag{A.50}$$

for some Lipschitz constant $C_n = O_p(1)$ from Assumption B. By Markov's inequality.

$$\begin{aligned} &\mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta}||H_n(\pi) - H_n(\tilde{\pi})|| > \eta) \\ &\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} C_n||\pi - \tilde{\pi}||) \ \leq \ \frac{1}{\eta}\delta\mathbb{E}_{\theta_n}(C_n) \end{aligned} \tag{A.51}$$

By Assumption B(iii), $\mathbb{E}_{\theta_n}(C_n) = O(1)$. Let $\epsilon, \eta > 0$, consider $\delta = \epsilon\eta/\mathbb{E}_{\theta_n}(C_n)$, then

$$\mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta}||H_n(\pi) - H_n(\tilde{\pi})|| > \eta) \ < \ \epsilon \tag{A.52}$$

which proves stochastic equicontinuity.

iv) To prove equicontinuity of $H(\pi, \theta_n)$, follow the same lines of iii), let $\epsilon, \eta > 0$, then

$$\begin{aligned} &\mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta}||H(\pi;\theta_n) - H(\tilde{\pi};\theta_n)|| > \eta) \\ &\leq \mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta}\mathbb{E}_{\theta_n}||H_n(\pi;\theta_n) - H_n(\tilde{\pi};\theta_n)| > \eta) \\ &\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} C_n||\pi - \tilde{\pi}||) \ \leq \ \frac{1}{\eta}\delta\mathbb{E}_{\theta_n}(C_n) \ < \ \epsilon \end{aligned} \tag{A.53}$$

for $\delta = \epsilon\eta/\mathbb{E}_{\theta_n}(C_n)$ which is finite as C_n is $O_p(1)$. Where the first inequality follows by Jensen's. This shows equicontinuity of $H(\pi;\theta_n)$. Furthermore, we have $\underline{eig}(H(\pi;\theta_n) > 0$ and $\overline{eig}(H(\pi;\theta_n)) < \infty$ by Assumption C. ■

Lemma A.2.2 *Under* $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$, $\sup_{\pi\in\Pi}||K_n(\pi) - K(\pi;\theta_n)|| \xrightarrow{p} 0$

Proof. The proof is equivalent to the proof for Lemma A.2.1 following steps i) - iv). Proof of ii) follows as the $K_n(\pi)$ are mixing and L_p bounded for $p = 4 + \iota$. For iii), stochastic equicontinuity of K_n follows as,

$$\mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} ||K_n(\pi) - K_n(\tilde{\pi})|| > \eta) \tag{A.54}$$
$$\leq \mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} n^{-1}\sum_{t=1}^{n}(||\sup_t h(X_{2t},\pi_n)||\ ||d_{\psi,t}(\pi) - d_{\psi,t}(\tilde{\pi})||) > \eta)$$
$$\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} C_s C_n||\pi - \tilde{\pi}||) \ \leq\ \frac{1}{\eta}\delta\mathbb{E}_{\theta_n}(C_s C_n) \leq\ \frac{1}{\eta}\delta(\mathbb{E}_{\theta_n}(C_s^2))^{1/2}(\mathbb{E}_{\theta_n}(C_n^2))^{1/2}$$

where $C_s = ||\sup_t h(X_{2t},\pi_n)|| = O_p(1)$ and $||n^{-1}\sum_{t=1}^{n} d_{\psi,t}(\pi) - n^{-1}\sum_{t=1}^{n} d_{\psi,t}(\tilde{\pi})|| \leq C_n||\pi - \tilde{\pi}||$ which by Assumption B is $O_p(1)$. The last inequality follows by Holder's inequality and C_n, C_s are bounded and ϵ_t and $d_{\psi,t}(\pi)$ are L_p bounded $p = 4 + \iota$. Now let $\epsilon, \eta > 0$ and let $\delta = \epsilon\eta/(\mathbb{E}_{\theta_n}(C_s^2))^{1/2}(\mathbb{E}_{\theta_n}(C_n^2))^{1/2}$ which shows stochastic equicontinuity. The same argument follows for the equicontinuity of $K(\pi;\theta_n)$ as

$$\mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} ||K(\pi;\theta_n) - K(\tilde{\pi};\theta_n)|| > \eta) \tag{A.55}$$
$$\leq \mathbb{P}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} \mathbb{E}_{\theta_n}||K_n(\pi) - K_n(\tilde{\pi})| > \eta)\ \blacksquare$$

Lemma A.2.3 $\{G_{\psi,n}(\pi) : \pi \in \Pi\} \Rightarrow \{G_{\psi}(\pi) : \pi \in \Pi\}$

Proof. As the parameter space Θ is compact, weak convergence requires pointwise convergence for each $\pi \in \Pi$ and Stochastic Equicontinuity (Dudley (1978) and Pollard (1990)). For pointwise convergence notice that

$$G_{\psi,n}(\pi) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\epsilon_t(\psi_{0,n})d_{\psi,t}(\pi) - \mathbb{E}_{\theta_n}(\epsilon_t(\psi_{0,n})d_{\psi,t}(\pi)] \tag{A.56}$$
$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\epsilon_t(\theta_n)d_{\psi,t}(\pi) + \beta_n' h(X_{2,t},\pi_n) - \mathbb{E}_{\theta_n}(\epsilon_t(\theta_n)d_{\psi,t}(\pi) + \beta_n' h(X_{2,t},\pi))] \tag{A.57}$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\theta_n) d_{\psi,t}(\pi) + b\ n^{-1} \sum_{t=1}^{n} [h(X_{2,t}, \pi_n) d_{\psi,t}(\pi)) - \mathbb{E}_{\theta_n}(h(X_{2,t}, \pi_n) d_{\psi,t}(\pi))] \tag{A.58}$$

By the same arguments as in Lemma A.2.2 the second terms are $o_{p,\pi}(1)$. That is, the second term satisfies $\sup_{\pi \in \Pi} ||K_n(\pi) - K(\pi;\theta_n)|| \xrightarrow{p} 0$

Consider $G_{\psi,n}(\pi) = \frac{1}{\sqrt{(n)}} \sum_{t=1}^{n} \epsilon_t(\theta_n) d_{\psi,t}(\pi) + o_{p,\pi}(1)$. For the pointwise central limit theorem, we use Wooldridge and White (1988) pointwise central limit theorem (1975) law of large numbers for dependent heterogeneously distributed random variables. Notice that

$$\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n) d_{\psi,t}(\pi)) = 0 \tag{A.59}$$

$$\mathbb{V}_{\theta_n}(\epsilon_t(\theta_n) d_{\psi,t}(\pi)) = \mathbb{E}_{\theta_n}(\sigma_t^2(\theta_n) d_{\psi,t}(\pi) d_{\psi,t}(\pi)') \tag{A.60}$$

$$Cov(\epsilon_t(\theta_n) d_{\psi,t}(\pi), \epsilon_{t-j}(\theta_n) d_{\psi,t-j}(\pi)) = 0 \tag{A.61}$$

all results follow using the Law of iterated expectations as ϵ_t is a martingale difference. We have that $\sup_t \mathbb{E}_{\theta_n} |\epsilon_t^2 d_{\psi,t}(\pi) d_{\psi,t}(\pi)'|^{1+\iota} < C < \infty$ by Assumption B(i). By Assumption B(ii) and as they are α-mixing of size $-r/(r-2)$ for $r > 2$. The moment and mixing conditions hold and therefore,

$$G_{\psi,n}(\pi) \xrightarrow{d} N(0, Avar(G_{\psi,n}(\pi))) \tag{A.62}$$

where the asymptotic variance for fixed $\pi \in \Pi$ equals,

$$\Omega(\pi) = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \mathbb{E}_{\theta_n}(\sigma_t^2(\theta_n) d_{\psi,t}(\pi) d_{\psi,t}(\pi)') \tag{A.63}$$

Stochastic equicontinuity follows from the Lipschitz argument used in Lemmas A.2.1 and A.2.2

$$\mathbb{P}_{\theta_n}(\sup_{\pi, \tilde{\pi} \in \Pi, ||\pi - \tilde{\pi}|| < \delta} ||G_{\psi,n}(\pi) - G_{\psi,n}(\tilde{\pi})|| > \eta) \tag{A.64}$$

$$\leq \mathbb{P}_{\theta_n}(\sup_{\pi, \tilde{\pi} \in \Pi, ||\pi - \tilde{\pi}|| < \delta} n^{-1} \sum_{t=1}^{n} (|| \sup_t \epsilon_t || \ ||d_{\psi,t}(\pi) - d_{\psi,t}(\tilde{\pi})||) > \eta)$$

$$\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}(\sup_{\pi,\tilde{\pi}\in\Pi,||\pi-\tilde{\pi}||<\delta} C_s C_n||\pi-\tilde{\pi}||) \ \leq \ \frac{1}{\eta}\delta\mathbb{E}_{\theta_n}(C_s C_n) \leq \ \frac{1}{\eta}\delta(\mathbb{E}_{\theta_n}(C_s^2))^{1/2}(\mathbb{E}_{\theta_n}(C_n^2))^{1/2}$$

The last inequality follows by Holder's inequality and as C_n, C_s are bounded and ϵ_t and $d_{\psi,t}(\pi)$ are L_p bounded $p = 4 + \iota$, specifically $C_s = ||\sup_t \epsilon_t(\theta_n)|| = O_p(1)$ and the Lipschitz constant C_n is $O_p(1)$ from Assumption B with $||n^{-1}\sum_{t=1}^n d_{\psi,t}(\pi) - n^{-1}\sum_{t=1}^n d_{\psi,t}(\tilde{\pi})|| \leq C_n||\pi - \tilde{\pi}||$. Now let $\epsilon, \eta > 0$ and let $\delta = \epsilon\eta/(\mathbb{E}_{\theta_n}(C_s^2))^{1/2}(\mathbb{E}_{\theta_n}(C_n^2))^{1/2}$ which shows stochastic equicontinuity. ■

Lemma A.2.4 *Under Assumption A to Assumption D*

a) For some non-stochastic function $Q(\theta;\theta_0)$, $\sup_{\theta\in\Theta}|Q_n(\theta) - Q(\theta,\theta_n)| \xrightarrow{p} 0$

b) When $\beta = 0$ *for every neighborhood* Ψ_0 *of* ψ_0, $\inf_{\pi\in\Pi}(\inf_{\psi\in\Psi(\pi)/\Psi_0} Q(\psi,\pi;\theta_0) - Q(\psi_0,\pi;\theta_0)) > 0 \ \forall \ \theta_0 = (\psi_0, \pi_0)$

Proof. First notice that the objective function doesn't depend on π when $\beta = 0$.

To prove a) we follow Newey (1991) (Theorem 2.1) and prove pointwise law of large numbers and SE for $Qn(\theta)$. Notice that for fixed $\theta \in \Theta$ by McLeish et al. (1975) (Theorem 2.10) law of large numbers, as the $\{y_t, X_t\}$ are mixing, then any non-random function is mixing of the same size $-r/(2r-1)$ with $r \geq 1$ (Assumption B). Also the condition

$$\sum_{k=1}^{\infty}\frac{\mathbb{E}_{\theta_n}(||Q_n(\pi) - Q(\pi;\theta_n)||)}{k^{1+\delta}} \leq \sum_{k=1}^{\infty}\frac{\mathbb{E}_{\theta_n}(||Q_n(\pi)|| + \mathbb{E}_{\theta_n}||Q(\pi;\theta_n)||)}{k^{1+\delta}} \leq C_1\sum_{k=1}^{\infty}\frac{1}{k^{1+\delta}} < C_2 \tag{A.65}$$

which follows using $C_1 = \sup_t \mathbb{E}_{\theta_n}||n^{-1}\sum_{t=1}^n \epsilon_t^2(\theta)||$, by Assumption A. This shows that for fixed $\theta \in \Theta$

$$|Q_n(\theta) - Q(\theta,\theta_n)| \xrightarrow{p} 0 \tag{A.66}$$

To prove Stochastic Equicontinuity of Q_n, we use Markov's Inequality and the Mean Value

Theorem.

$$\begin{aligned}
&\mathbb{P}_{\theta_n}(\sup_{\theta,\tilde{\theta}\in\Theta,||\theta-\tilde{\theta}||<\delta}||Q_n(\pi)-Q_n(\tilde{\pi})||>\eta) \qquad\qquad (A.67)\\
&\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}((\sup_{\theta,\tilde{\theta}\in\Theta,||\theta-\tilde{\theta}||<\delta}||\theta-\tilde{\theta}||)\ (2n^{-1}\sum_{t=1}^{n}|\sup_{\theta\in\Theta}\epsilon_t^2(\theta)|))\ \leq\ \frac{1}{\eta}\delta\mathbb{E}_{\theta_n}(C_n)
\end{aligned}$$

for $C_n = 2n^{-1}\sum_{t=1}^{n}|\sup_{\theta\in\Theta}\epsilon_t^2(\theta)|$, which is bounded by Assumption A. Let $\epsilon,\eta>0$, the for $\delta=\epsilon\eta/\mathbb{E}_{\theta_n}(Cn)$ satisfies stochastic equicontinuity.

The same argument applies to prove Equicontinuity of $Q(\theta;\theta_n)$ we can use the same $\delta=\epsilon\eta/\mathbb{E}_{\theta_n}(Cn)$, as by Jensen's Inequality

$$\begin{aligned}
&\mathbb{P}_{\theta_n}(\sup_{\theta,\tilde{\theta}\in\Theta,||\theta-\tilde{\theta}||<\delta}||Q(\theta;\theta_n)-Q(\tilde{\theta};\theta_n)||>\eta) \qquad\qquad (A.68)\\
&\leq \mathbb{P}_{\theta_n}(\sup_{\theta,\tilde{\theta}\in\Theta,||\theta-\tilde{\theta}||<\delta}\mathbb{E}_{\theta_n}||Q_n(\theta)-Q_n(\tilde{\theta})|>\eta)
\end{aligned}$$

b) To prove the identification condition let $\beta=0$. By definition of concentrated estimator and as all Q_n are uniformly continuous (Θ is a compact set), converging uniformly to $Q(\psi,\pi;\theta_0)$, the limit function is uniformly continuous. The extreme value theorem guarantees that a minimum exists, and by Assumption A(iv) it is unique. ■

Lemma A.2.5 *Under* $\{\theta_n\in\Theta(\theta_0,0,b)\}$ *with* $||b||<\infty$, $\sqrt{n}(\hat{\psi}_n(\pi)-\psi_n)\Rightarrow -H^{-1}(\pi;\theta_0)(G_\psi(\pi;\theta_0)+K(\pi;\theta_0)b)-(b,0_{d_\zeta})\equiv\tau(\pi;\theta_0,b)$

Proof. Let $\rho_t(\theta)=\epsilon_t^2(\theta)$. Taking derivatives of the objective function with respect to the strongly identified parameters and using the Mean Value Theorem.

$$\begin{aligned}
0=\frac{\partial}{\partial\psi}Q_n(\hat{\psi}_n(\pi),\pi) &= \frac{\partial}{\partial\psi}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)+\frac{\partial}{\partial\psi\partial\psi'}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\bar{\psi}_{0,n},\pi)(\hat{\psi}_n(\pi)-\psi_{0,n})+o_{p,\pi}(1)\\
&= \frac{\partial}{\partial\psi}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)+\frac{\partial}{\partial\psi\partial\psi'}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)(\hat{\psi}_n(\pi)-\psi_n+\psi_n-\psi_{0,n})+o_{p,\pi}(1)
\end{aligned}$$

$$\Leftrightarrow \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) = -H_n(\psi_{0,n}, \pi)^{-1}(\frac{\partial}{\partial \psi}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\rho_t(\psi_{0,n}, \pi)) - \sqrt{n}(\psi_n - \psi_{0,n}) + o_{p,\pi}(1) \tag{A.69}$$

for some mean value $\bar{\theta}$. Notice that $\psi_{0,n}$ is an intermediate value between ψ_0 and ψ_n. By Equation (A.69),

$$\begin{aligned}\frac{\partial}{\partial \psi}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n}, \pi) &= \frac{\partial}{\partial \psi}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n}, \pi) - \mathbb{E}_{\theta_n}\frac{\partial}{\partial \psi}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n}, \pi) + \mathbb{E}_{\theta_n}\frac{\partial}{\partial \psi}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n}, \pi) \\ &= n^{-1/2}G_{\psi,n}(\psi_{0,n}, \pi) + \underbrace{\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\theta_n}\frac{\partial}{\partial \psi}\rho_t(\psi_{0,n}, \pi)}_{(I)} \end{aligned} \tag{A.70}$$

Further from (I) using the mean value theorem with respect to the parameter β solely.

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\theta_n}\frac{\partial}{\partial \psi}\rho_t(\psi_{0,n}, \pi) \\ &= \underbrace{\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\theta_{0,n}}\frac{\partial}{\partial \psi}\rho_t(\psi_{0,n}, \pi)}_{(II)} + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial \beta}\mathbb{E}_{\bar{\theta}_n}\frac{\partial}{\partial \psi}\rho_t(\psi_{0,n}, \pi)\beta_n = K_n(\psi_{0,n}, \pi; \bar{\theta_n})\beta_n \end{aligned} \tag{A.71}$$

Where the last equality follows as (II) equals zero by the Law of Iterated Expectations and Assumption A(iv), specifically,

$$n^{-1}\sum_{t=1}^{n}\mathbb{E}_{\theta_{0,n}}\frac{\partial}{\partial \psi}\rho_t(\psi_{0,n}, \pi) = n^{-1}\sum_{t=1}^{n}\mathbb{E}_{\theta_{0,n}}(\mathbb{E}_{\theta_{0,n}}(\epsilon_t(\psi_{0,n}, \pi)|W_t)d_{\psi,t}(\pi)) = 0 \tag{A.72}$$

As $\bar{\theta}_n$ is a mean value between θ_n and $\theta_{0,n}$, and consistency of $\hat{\psi}_n$ we have by Lemma A.2.1

$$\sup_{\pi \in \Pi}||K_n(\psi_{0,n}, \pi; \bar{\theta}_n) - K(\pi; \theta_n)|| \xrightarrow{p} 0 \tag{A.73}$$

uniformly over $\pi \in \Pi$. Therefore $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial \psi}\rho_t(\psi_{0,n}, \pi) = G_{\psi,n}(\psi_{0,n}, \pi) + K_n(\psi_{0,n}, \pi; \bar{\theta_n})\beta_n$. By Lemma A.2.3 $G_{\psi,n}(\psi_{0,n}, \pi) \Rightarrow G_\psi(\pi)$, where $G_\psi(\pi)$ is zero mean and covariance kernel

$$\Omega(\pi_1, \pi_2; \theta_0) = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mathbb{E}_{\theta_n}(\frac{\partial}{\partial \psi}\rho_t(W_t, \psi_n, \pi_1)\frac{\partial}{\partial \psi}\rho_t(W_i, \psi_n, \pi_2)'). \tag{A.74}$$

Also as $\sqrt{n}\beta_n \to b$, we obtain.

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\psi}\rho_t(\psi_{0,n},\pi) \Rightarrow G_\psi(\pi) + K(\psi_0,\pi;\theta_0)b \tag{A.75}$$

Finally, by $\sqrt{n}(\psi_n - \psi_{0,n}) = \sqrt{n}(\beta_n', 0')' \to (b, 0_{d_\zeta})$ Lemma A.2.1, the continuous mapping theorem and $\sup_{\pi\in\Pi}||H_n^{-1}(\pi) - H^{-1}(\pi;\theta_n)|| \xrightarrow{p} 0$, we obtain the desired result,

$$\begin{aligned}\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) &= -H_n^{-1}(\psi_{0,n},\pi)(G_{\psi,n}(\psi_{0,n},\pi) + K_n(\psi_{0,n},\pi;\bar{\theta}_n)\beta_n) - \sqrt{n}(\psi_n - \psi_{0,n}) + o_{p,\pi}(1)\\ &\Rightarrow -H^{-1}(\pi;\theta_0)(G_\psi(\pi;\theta_0) + K(\pi;\theta_0)b) - (b, 0_{d_\zeta}) \equiv \tau(\pi;\theta_0,b) \quad \blacksquare\end{aligned} \tag{A.76}$$

Lemma A.2.6 *Under* $\{\theta_n \in \Theta(\theta_0, 0, b)\}$ *with* $||b|| < \infty$,

$$n[Q_n(\hat{\psi}_n(\pi),\pi) - Q_n(\psi_{0,n},\pi)] \Rightarrow -\frac{1}{2}\tau(\pi;\theta_0,b)H(\pi;\theta_0)\tau(\pi;\theta_0,b) \equiv \xi(\pi;\theta_0,b) \tag{A.77}$$

Proof. Take a Taylor expansion of the objective function with respect to the point of non-identification $\psi_{0,n} = (\beta_0, \zeta_n)$ and by consistency of the strongly identified parameters.

$$\begin{aligned}Q_n(\hat{\psi}_n(\pi),\pi) - Q_n(\psi_{0,n},\pi) &= \frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi)(\hat{\psi}_n(\pi) - \psi_{0,n})\\ &+ \frac{1}{2}(\hat{\psi}_n(\pi) - \psi_{0,n})(\frac{\partial^2}{\partial\psi\partial\psi}Q_n(\psi_{0,n},\pi))(\hat{\psi}_n(\pi) - \psi_{0,n}) + o_{p,\pi}(1)\\ n[Q_n(\hat{\psi}_n(\pi),\pi) - Q_n(\psi_{0,n},\pi)] &= \frac{\partial}{\partial\psi}\sqrt{n}Q_n(\psi_{0,n},\pi)\sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n})\\ &+ \frac{1}{2}\sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n})(\frac{\partial^2}{\partial\psi\partial\psi}Q_n(\psi_{0,n},\pi))\sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n})) + o_{p,\pi}(1)\end{aligned} \tag{A.78}$$

Where $\sqrt{n}Q_n(\psi_{0,n},\pi) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)$. By Equation (A.69) from the previous lemma

$$\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) + \sqrt{n}(\psi_n - \psi_{0,n}) = -[\frac{\partial^2}{\partial\psi\partial\psi'}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)]^{-1}[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\psi}\rho_t(\psi_{0,n},\pi)] + o_{p,\pi}(1)$$

$$\text{i.e.} \quad \sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n}) = -[\frac{\partial^2}{\partial\psi\partial\psi'}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)]^{-1}[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\psi}\rho_t(\psi_{0,n},\pi)] \equiv Z_n(\pi) \tag{A.79}$$

By Equation (A.69) we also get

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\psi}\rho_t(\psi_{0,n},\pi) = -[n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\psi\partial\psi'}\rho_t(\psi_{0,n},\pi)]\sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n}) \tag{A.80}$$

combining Equations (A.78) to (A.80)

$$\begin{aligned} & n[Q_n(\hat{\psi}_n(\pi),\pi) - Q_n(\psi_{0,n},\pi)] \\ & = -Z_n(\pi)[n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\psi\partial\psi'}\rho_t(\psi_{0,n},\pi)]Z_n(\pi) + \frac{1}{2}Z_n(\pi)[\frac{\partial^2}{\partial\psi\partial\psi'}\frac{1}{n}\sum_{t=1}^{n}\rho_t(\psi_{0,n},\pi)]Z_n(\pi) \\ & = -\frac{1}{2}Z_n(\pi)H_n(\psi_{0,n},\pi)Z_n(\pi) \end{aligned} \tag{A.81}$$

By Lemmas A.2.1 to A.2.3 $Z_n(\pi) \Rightarrow -H^{-1}(\pi;\theta_0)(G_\psi(\pi;\theta_0) + K(\pi;\theta_0)b)$ as H_n and K_n have non-random limits. Using the continuous mapping theorem we obtain the desired result

$$\begin{aligned} & n[Q_n(\hat{\psi}_n(\pi),\pi) - Q_n(\psi_{0,n},\pi)] \\ & \Rightarrow -\frac{1}{2}(G_\psi(\pi;\theta_0) + K(\pi;\theta_0)b)'H^{-1}(\pi;\theta_0)(G_\psi(\pi;\theta_0) + K(\pi;\theta_0)b) \equiv \xi(\pi;\theta_0,b) \quad \blacksquare \end{aligned} \tag{A.82}$$

Lemma A.2.7 *Under* $\theta_n \in \Theta(\theta_0,\infty,\omega_0)$

(a) $\hat{J}_n = J_n(\hat{\theta}_n) = B^{-1}(\hat{\beta}_n)\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\hat{\theta}_n)B^{-1}(\hat{\beta}_n) - J(\theta_0) \xrightarrow{p} 0$ *where* $J(\theta_0)$ *is non-singular and symmetric.*

(b) $n^{1/2}B^{-1}(\hat{\beta}_n)\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n) \xrightarrow{d} N(0,V(\theta_0))$ *for some symmetric and positive definite matrix* $V(\theta_0)$*, and* $||\hat{V}_n - V(\theta_0)|| \xrightarrow{p} 0$

Proof. (a) Under semi-strong identification we need pointwise law of large numbers and central limit theorem which follow as the parameters are consistent. For fixed $\theta \in \theta$ we use McLeish et al. (1975) (Theorem 2.10) law of large numbers, as in Equations (A.48) and (A.49). By Markov's inequality, $\sup_t \mathbb{E}_{\theta_n}||d_{\theta,t}(\theta_n)d_{\theta,t}(\theta_n)'||^{1+\iota} < C < \infty$, as $d_{\theta,t}(\theta)$ is uniformly L_p bounded for $p = 4+\iota$ by Assumption B(ii), see also Equation (A.16). The law of large numbers conditions hold.

b) For the pointwise central limit theorem, we prove the conditions of Wooldridge and White (1988) central limit theorem (1975) for dependent heterogeneously distributed random variables are satisfied. We have that $\sup_t \mathbb{E}_{\theta_n}||\epsilon_t^2 d_{\theta,t}(\theta_n)d_{\theta,t}(\theta_n)'||^{1+\iota} < C < \infty$, by Assumption B(ii) and as they are α-mixing of size $-r/(r-2)$ for $r > 2$ and by Assumption B(i) they are uniformly bounded. The central limit theorem conditions hold. ■

Lemma A.2.8 *Under $\theta_n \in \Theta(\theta_0, 0, b)$ with $||b|| < \infty$, the matrices $\hat{J}_n = J_n(\hat{\theta}_n)$ and $\hat{V}_n = V_n(\hat{\theta}_n)$ satisfy a uniform law of large numbers with non-singular limits.*

$$\sup_{\theta\in\Theta} ||J_n(\theta) - J(\theta;\theta_n)|| \xrightarrow{p} 0 \tag{A.83}$$

$$\sup_{\theta\in\Theta} ||V_n(\theta) - V(\theta;\theta_n)|| \xrightarrow{p} 0 \tag{A.84}$$

Proof. We prove that conditions i) - iv) of Newey (1991) (Theorem 2.1) hold. i) Compactness is given by Assumption D. ii) The law of large numbers for pointwise $\theta \in \Theta$ is proven in Lemma A.2.7. We are left to show Stochastic Equicontinuity.

iii) To prove Stochastic Equicontinuity of J_n and V_n, we use a similar approach to Lemmas A.2.1 and A.2.2. Specifically, let $\epsilon, \eta > 0$, for J_n consider $\delta = \epsilon\eta/\mathbb{E}_{\theta_n}(C_n)$, with $C_n = \sup_t ||d_{\theta,t}(\theta)d_{\theta,t}(\tilde{\theta})'||$. The proof for V_n follows the same approach $\delta = \epsilon\eta/2\mathbb{E}_{\theta_n}(C_n)$, with $C_n = (\sup_t ||\epsilon_t^2(\theta)||_2)^{1/2}(||d_{\theta,t}(\theta)d_{\theta,t}(\tilde{\theta})'||_2)^{1/2}$ for V_n. The constants C_n are $O_p(1)$ by Assumption B and are obtained using Markov's and Holders inequalities.

iv) To prove equicontinuity of $V(\theta,\theta_n)$ and $J(\theta,\theta_n)$, follow the same lines of iii). Consider $J(\theta;\theta_n)$

$$\begin{aligned} &\mathbb{P}_{\theta_n}\Big(\sup_{\theta,\tilde{\theta}\in\Theta,||\theta-\tilde{\theta}||<\delta} ||J(\theta;\theta_n) - J(\tilde{\theta};\theta_n)|| > \eta\Big) \qquad \text{(A.85)} \\ &\le \mathbb{P}_{\theta_n}\Big(\sup_{\theta,\tilde{\theta}\in\Theta,||\theta-\tilde{\theta}||<\delta} \mathbb{E}_{\theta_n}||J_n(\theta;\theta_n) - J_n(\tilde{\theta};\theta_n)| > \eta\Big) \end{aligned}$$

The inequality follows by Jensen's. This shows equicontinuity of $J(\pi;\theta_n)$. Furthermore, we have $\underline{eig}(J(\pi;\theta_n) > 0$ and $\overline{eig}(J(\pi;\theta_n)) < \infty$ by Assumption G. To show equicontinuity of $V(\theta;\theta_n)$, just replace J_n in the inequality of Equation (A.85) and use the same δ. ■

Lemma A.2.9 *Let $M_n \to \infty$ as $n \to \infty$, under $\theta_n \in \Theta(\theta_0, 0, b)$ with $||b|| < \infty$, $\{G^m_{\psi,n}(\pi) : \pi \in \Pi\} \Rightarrow \{G_\psi(\pi) : \pi \in \Pi\}$*

Proof. We follow the proof of Lemma A.2.3, we show pointwise convergence in distribution and stochastic equicontinuity.

To show pointwise convergence in distribution, we show the conditions of Wooldridge and White (1988) central limit theorem hold. The moments of $G^m_{\psi,n}(\pi)$ are equivalent to equations (A.59) - (A.61). We also have that as z^m_t is i.i.d. then it follows that $G^m_{\psi,n}(\pi)$ is mixing of size $-r/(r-1)$ for $r > 1$, also we have

$$\mathbb{E}_{\theta_n} \sup_t ||(z^m_t)^2 \epsilon^2_t(\theta_n) d_{\psi,t}(\pi) d_{\psi,t}(\pi)'||^{1+\iota} \tag{A.86}$$

$$\leq \mathbb{E}_{\theta_n} \sup_t |(z^m_t)^2|^{1+\iota}\ \mathbb{E}_{\theta_n} \sup_t ||\epsilon^2_t(\theta_n) d_{\psi,t}(\pi) d_{\psi,t}(\pi)'||^{1+\iota} < C < \infty, \quad \forall \pi \in \Pi, \tag{A.87}$$

which follows by Assumption A and B. The central limit theorem follows for fixed $\pi \in \Pi$, $G^m_{\psi,n}(\pi) \xrightarrow{d} G_\psi(\pi)$ follows by Lemma A.2.3. By law of iterated expectations, the asymptotic variance kernel is equal to

$$\Omega(\pi, \tilde{\pi}; \theta_0) = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathbb{E}_{\theta_n}(\epsilon^2_t(\theta_n) d_{\psi,t}(\pi) d_{\psi,t}(\tilde{\pi})') \tag{A.88}$$

To show Stochastic Equicontinuity we let $\epsilon, \eta > 0$ and let $\delta = \epsilon\eta/(\mathbb{E}_{\theta_n}(C^2_s))^{1/2}(\mathbb{E}_{\theta_n}(C^2_n))^{1/2}$ with $C_s = ||\sup_t z^m_t \epsilon_t(\theta_n)|| = O_p(1)$ and the constant $C_n = O_p(1)$ is the Lipschitz constant from Assumption B, $||n^{-1}\sum_{t=1}^n d_{\psi,t}(\pi) - n^{-1}\sum_{t=1}^n d_{\psi,t}(\tilde{\pi})|| \leq C_n||\pi - \tilde{\pi}||$. This shows stochastic equicontinuity of $G^m_{\psi,n}(\pi)$. ■

Lemma A.2.10 *Consider the bootstrapped distribution $\tau^m_{\theta,n}(\hat{\theta}_n)$. Let $M_n \to \infty$ as $n \to \infty$, under $\theta_n \in \Theta(\theta_0, \infty, \omega_0)$, $\tau^m_{\theta,n}(\hat{\theta}_n) \xrightarrow{d} N(0, J^{-1}(\theta_0)V(\theta_0)J^{-1}(\theta_0))$*

Proof. From Equation (1.42)

$$\begin{aligned}\tau^m_{\theta,n}(\theta) &= [B^{-1}(\hat{\beta}_n)\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\hat{\theta}_n)B^{-1}(\hat{\beta}_n)]^{-1}B^{-1}(\hat{\beta}_n)\sqrt{n}\frac{\partial}{\partial\theta}Q^m_n(\hat{\theta}_n) \\ &= [B^{-1}(\hat{\beta}_n)B(\hat{\beta}_n)J_n^{-1}(\hat{\theta}_n)B(\hat{\beta}_n)B^{-1}(\hat{\beta}_n)]B^{-1}(\hat{\beta}_n)B(\hat{\beta}_n)G^m_{\theta,n}(\hat{\theta}_n) \\ &= J_n^{-1}(\hat{\theta}_n)G^m_{\theta,n}(\hat{\theta}_n) = O_p(1)\end{aligned} \tag{A.89}$$

By Lemma A.2.8, we only need to show $G^m_{\theta,n}(\hat{\theta}_n) \xrightarrow{d} N(0, V(\theta_0))$. As $\hat{\theta}_n$ is consistent, the Central Limit Theorem is pointwise, we use Wooldridge and White (1988). Mixing conditions follow from Assumption B(i). By Assumption A and B the inequality follows

$$\sup_t \mathbb{E}_{\theta_n}|(z^m_t)^2\epsilon^2_t d_{\psi,t}(\pi)d_{\psi,t}(\pi)'|^{1+\iota} = \sup_t \mathbb{E}_{\theta_n}|\epsilon^2_t d_{\psi,t}(\pi)d_{\psi,t}(\pi)'|^{1+\iota} < C < \infty \tag{A.90}$$

as ϵ_t is a martingale difference and z^m_t is an i.i.d. process. The conditions of the central limit theorem hold and therefore

$$G^m_{\theta,n}(\hat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^n z^m_t\epsilon_t(\hat{\theta}_n)d_{\theta,t}(\hat{\theta}_n) + o_p(1) \xrightarrow{d} N(0, V(\theta_0)) \tag{A.91}$$

The matrix $J_n \xrightarrow{p} J(\theta_0)$ which is non-random, and $J(\theta_0), V(\theta_0)$ matrices are positive semidefinite by Assumption G. We have the desired result using the product limit rule. ■

A.3 Vector β

When β is a vector, the derivation of the t-statistic requires slight changes in the assumptions of the variance-covariance matrix (see Supplemental Appendix A of Andrews and Cheng (2012)). The proofs are essentially equivalent, we only need to redefine some of the processes used. Let $\omega = \beta/||\beta||$ if $\beta \neq 0$ and $\omega = 1_{d_\beta}/||1_{d_\beta}||$ where 1_{d_β} denotes a vector of ones. Let $\theta^+ = (||\beta||, \omega, \zeta', \pi')' \in \Theta^+$, where $\Theta^+ = \{\theta^+ : \theta^+ = (||\beta||, \beta'/||\beta||, \zeta', \pi')', \theta \in \Theta\}$.

We define analogously the matrices with respect to θ^+, $J(\theta^+;\theta_n)$ and $V(\theta^+;\theta_n)$ and let

$$\Sigma(\theta^+;\theta_n) = J^{-1}(\theta^+;\theta_n)V(\theta^+;\theta_n)J^{-1}(\theta^+;\theta_n) \tag{A.92}$$

$$\Sigma(\pi, \omega, \theta_n) = \Sigma(||\beta_0||, \omega, \zeta_0, \pi; \theta_n) \tag{A.93}$$

For the vector β we must modify Assumption F to apply for matrices $J(\theta^+; \theta_n)$ and $V(\theta^+; \theta_n)$. Moreover, Lemma A.2.8 applies for this matrices, that is

$$\sup_{\theta \in \Theta} ||J_n(\theta) - J(\theta^+; \theta_n)|| \xrightarrow{p} 0 \tag{A.94}$$

$$\sup_{\theta \in \Theta} ||V_n(\theta) - V(\theta^+; \theta_n)|| \xrightarrow{p} 0 \tag{A.95}$$

The difference now is that we focus on the convergence of ω_n instead of β_n. Combining these results along with Assumption E is equivalent to the Assumptions V1 of Vector β from Supplemental Appendix A of Andrews and Cheng (2012)). Proposition 1.3.3 applies with analogous proof defining

$$\bar{\Sigma}(\pi; \theta_0, b) = \Sigma(\pi, \omega^*(\pi; \theta_n, b); \theta_n) \tag{A.96}$$

$$\omega^*(\pi; \theta_n, b) = \tau_\beta(\pi; \theta_n, b)/||\tau_\beta(\pi; \theta_n, b)|| \tag{A.97}$$

We define $\bar{\Sigma}_{\beta,\beta}$ and $\bar{\Sigma}_{\pi,\pi}$ analogously to the β scalar case as the upper and lower matrix of Σ.

A.4 Simulation Details

In this section, we provide with the details of the simulation performed to obtain the Tables of the Size and power of the t-test. The sample sizes used are $n = 100, 250$ and 500. The number of simulations are 1000. We assume four distributions for the error component, $N(0,1)$, $t(4)$, GARCH with $\omega = 0.1, \alpha = 0.3, \beta = 0.6$ and GARCH with $\omega = 0.1, \alpha = 0.6, \beta = 0.3$. The number of bootstrapped samples is 500.

The method for optimization used is MATLAB fmincon function providing the gradient. For the initial estimation of the parameters, we use 100 uniformly distributed initial values for estimation, which are considered enough as the problem has low dimensionality and the functions are smooth. Following Andrews and Cheng (2012) we use $\kappa_n = \log(n)^{1/2}$ which is equivalent to the BIC.

The robust t-statistic are obtained the using the t-statistic Equation (1.28), while the standard t-statistic is the usual statistic used in econometric literature that assumes strong identification. The critical value of the t-statistic are obtain simulating the asymptotic distribution in Proposition 1.3.3 and Theorem 1.4.2 for each pair of nuisance parameters (b, π_0). The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$. Section 1.3 presents the asymptotic approximation statistic of Andrews and Cheng (2012) while Section 1.4 derives the distribution of the bootstrapped t-statistic. The stochastic process $G_{\psi,n}$ is obtained simulating each sample path individually using the mvnrnd command in MATLAB using the estimated variance-covariance matrix.

$$\hat{\Omega}_{\psi,n}(\pi, \tilde{\pi}) = n^{-1} \sum_{t=1}^{n} \epsilon_t(\hat{\theta}_{0,n}) d_{\psi,t}(\pi) d_{\psi,t}(\tilde{\pi}) \qquad \text{under heteroscedasticity} \tag{A.98}$$

$$\hat{\Omega}_{\psi,n}(\pi, \tilde{\pi}) = \hat{\sigma}^2(\hat{\theta}_{0,n})\ n^{-1} \sum_{t=1}^{n} d_{\psi,t}(\pi) d_{\psi,t}(\tilde{\pi}) \qquad \text{under homoscedasticity} \tag{A.99}$$

As the derivative with respect to ζ does not depend on π, the $\hat{\Omega}_{\psi,n}(\pi, \tilde{\pi})$ is constructed using derivatives with respect to β solely. The bootstrapped asymptotic distribution is constructed using the stochastic processes depicted in Equation (1.57). For each of the 1000 simulations, critical values are obtained using 500 draws from $G_{\psi,n}$ and 500 bootstrapped draws from $G^m_{\psi,n}$. The stochastic process T^ψ and T^π follow using the processes $G_{\psi,n}$ or $G^m_{\psi,n}$, probability limits in Lemmas A.2.1 and A.2.2 and processes ξ_n and τ_n.

The t-statistics T_n and T^s_n are compared to the critical values, rejection rates are obtained for all simulations that surpass these values for each simulation. The critical values obtained for each simulation are computed for the grid of nuisance parameters, with the robust critical values derived using Equation (1.47) and (1.49).

A.5 Supplemental Tables and Figures

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 100$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.014	0.056	0.012	0.023	0.115	0.116	0.115	0.116	0.115	0.097	0.121	0.000	0.005	0.122	0.126	0.122	0.126	0.122
5%	0.009	0.028	0.008	0.012	0.065	0.067	0.065	0.067	0.065	0.050	0.059	0.000	0.003	0.067	0.065	0.067	0.065	0.067
1%	0.007	0.011	0.007	0.008	0.018	0.021	0.018	0.021	0.018	0.007	0.012	0.000	0.000	0.015	0.016	0.015	0.016	0.015
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.064	0.111	0.057	0.088	0.303	0.308	0.303	0.308	0.303	0.301	0.331	0.002	0.041	0.345	0.341	0.345	0.341	0.345
5%	0.039	0.069	0.028	0.056	0.203	0.202	0.203	0.202	0.203	0.205	0.229	0.001	0.019	0.238	0.241	0.238	0.241	0.238
1%	0.010	0.037	0.008	0.025	0.075	0.096	0.075	0.096	0.075	0.083	0.095	0.000	0.002	0.104	0.108	0.104	0.108	0.104
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.591	0.826	0.544	0.804	0.950	0.955	0.950	0.955	0.950	0.910	0.914	0.250	0.612	0.920	0.917	0.920	0.917	0.920
5%	0.459	0.757	0.383	0.701	0.912	0.906	0.912	0.906	0.912	0.866	0.873	0.105	0.465	0.880	0.879	0.880	0.879	0.880
1%	0.246	0.561	0.158	0.465	0.769	0.778	0.769	0.778	0.769	0.733	0.746	0.011	0.208	0.764	0.762	0.764	0.762	0.764
Weakly identified $n = 100$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.111	0.113	0.057	0.051	0.148	0.143	0.141	0.134	0.277	0.068	0.067	0.001	0.001	0.098	0.088	0.054	0.055	0.077
5%	0.055	0.061	0.024	0.023	0.085	0.075	0.077	0.071	0.174	0.024	0.031	0.001	0.000	0.047	0.041	0.031	0.037	0.033
1%	0.010	0.012	0.002	0.004	0.027	0.025	0.024	0.025	0.040	0.009	0.009	0.000	0.000	0.012	0.011	0.012	0.011	0.012
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.282	0.277	0.269	0.267	0.284	0.277	0.272	0.268	0.379	0.386	0.391	0.127	0.127	0.465	0.451	0.164	0.160	0.401
5%	0.246	0.244	0.225	0.223	0.247	0.244	0.226	0.224	0.324	0.304	0.307	0.086	0.078	0.389	0.373	0.135	0.126	0.317
1%	0.150	0.162	0.122	0.121	0.150	0.162	0.122	0.121	0.255	0.184	0.194	0.022	0.029	0.271	0.260	0.107	0.105	0.194
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.874	0.877	0.855	0.859	0.928	0.932	0.916	0.918	0.961	0.806	0.806	0.612	0.600	0.842	0.834	0.617	0.608	0.812
5%	0.824	0.832	0.794	0.804	0.871	0.870	0.847	0.848	0.935	0.764	0.765	0.531	0.524	0.809	0.804	0.540	0.535	0.770
1%	0.703	0.703	0.636	0.642	0.724	0.725	0.662	0.670	0.837	0.667	0.680	0.411	0.387	0.735	0.740	0.426	0.404	0.680
Non-identified $n = 100$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.117	0.110	0.032	0.040	0.130	0.111	0.083	0.082	0.245	0.055	0.059	0.001	0.000	0.122	0.097	0.031	0.031	0.059
5%	0.051	0.057	0.015	0.022	0.089	0.083	0.082	0.082	0.131	0.019	0.022	0.000	0.000	0.057	0.042	0.021	0.021	0.023
1%	0.008	0.017	0.002	0.004	0.036	0.037	0.036	0.037	0.036	0.005	0.005	0.000	0.000	0.008	0.009	0.005	0.009	0.005
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.431	0.435	0.312	0.302	0.439	0.439	0.320	0.309	0.488	0.359	0.360	0.090	0.083	0.450	0.444	0.109	0.102	0.371
5%	0.347	0.352	0.226	0.223	0.354	0.358	0.230	0.228	0.452	0.270	0.272	0.051	0.043	0.379	0.368	0.077	0.068	0.283
1%	0.193	0.203	0.069	0.080	0.206	0.206	0.075	0.080	0.289	0.139	0.149	0.009	0.012	0.246	0.245	0.053	0.051	0.146
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.930	0.934	0.868	0.863	0.942	0.945	0.886	0.879	0.969	0.790	0.793	0.571	0.562	0.846	0.840	0.573	0.565	0.798
5%	0.884	0.895	0.817	0.822	0.892	0.900	0.826	0.833	0.946	0.736	0.738	0.491	0.479	0.797	0.797	0.496	0.481	0.741
1%	0.793	0.812	0.693	0.708	0.803	0.819	0.698	0.717	0.852	0.636	0.640	0.354	0.350	0.718	0.717	0.360	0.358	0.644

Table A.1: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with Normal(0,1) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	$LF\ \beta$				$ICS_1\ \beta$					$LF\ \pi$				$ICS_1\ \pi$				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 100$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.025	0.059	0.020	0.036	0.101	0.107	0.101	0.107	0.101	0.089	0.113	0.004	0.012	0.117	0.120	0.117	0.120	0.117
5%	0.015	0.042	0.013	0.027	0.064	0.069	0.064	0.069	0.064	0.050	0.062	0.001	0.007	0.067	0.072	0.067	0.072	0.067
1%	0.010	0.024	0.008	0.015	0.031	0.034	0.031	0.034	0.031	0.016	0.026	0.000	0.003	0.025	0.030	0.025	0.030	0.025
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.055	0.103	0.047	0.078	0.289	0.289	0.289	0.289	0.289	0.286	0.303	0.005	0.042	0.325	0.318	0.325	0.318	0.325
5%	0.033	0.059	0.028	0.049	0.176	0.185	0.176	0.185	0.176	0.183	0.220	0.001	0.019	0.227	0.233	0.227	0.233	0.227
1%	0.022	0.032	0.015	0.023	0.064	0.078	0.064	0.078	0.064	0.071	0.096	0.000	0.006	0.093	0.107	0.093	0.107	0.093
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.592	0.830	0.528	0.799	0.939	0.933	0.939	0.933	0.939	0.920	0.928	0.253	0.587	0.940	0.941	0.940	0.941	0.940
5%	0.442	0.746	0.351	0.701	0.893	0.893	0.893	0.893	0.893	0.866	0.888	0.128	0.450	0.898	0.901	0.898	0.901	0.898
1%	0.207	0.550	0.140	0.464	0.759	0.769	0.759	0.769	0.759	0.718	0.750	0.022	0.205	0.757	0.767	0.757	0.767	0.757
Weakly identified $n = 100$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.078	0.089	0.038	0.043	0.123	0.118	0.112	0.109	0.240	0.045	0.051	0.002	0.001	0.095	0.084	0.040	0.039	0.062
5%	0.031	0.050	0.016	0.020	0.061	0.066	0.055	0.061	0.129	0.021	0.023	0.000	0.000	0.043	0.044	0.023	0.026	0.029
1%	0.007	0.012	0.002	0.006	0.019	0.021	0.019	0.021	0.032	0.004	0.007	0.000	0.000	0.009	0.012	0.008	0.011	0.008
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.263	0.256	0.251	0.247	0.270	0.261	0.257	0.252	0.349	0.375	0.367	0.107	0.096	0.463	0.435	0.137	0.118	0.388
5%	0.232	0.219	0.209	0.199	0.233	0.221	0.210	0.201	0.303	0.278	0.281	0.064	0.066	0.396	0.373	0.106	0.093	0.293
1%	0.141	0.131	0.097	0.101	0.143	0.131	0.099	0.101	0.228	0.157	0.168	0.016	0.026	0.261	0.246	0.076	0.074	0.171
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.861	0.847	0.841	0.828	0.900	0.886	0.884	0.872	0.954	0.789	0.784	0.579	0.558	0.843	0.817	0.589	0.566	0.792
5%	0.802	0.793	0.765	0.767	0.838	0.824	0.802	0.799	0.918	0.741	0.744	0.511	0.491	0.801	0.789	0.521	0.502	0.749
1%	0.676	0.666	0.614	0.617	0.700	0.684	0.640	0.639	0.812	0.638	0.636	0.393	0.379	0.727	0.697	0.406	0.392	0.651
Non-identified $n = 100$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.107	0.109	0.042	0.047	0.132	0.111	0.092	0.087	0.231	0.052	0.060	0.001	0.002	0.122	0.105	0.037	0.036	0.067
5%	0.040	0.059	0.014	0.026	0.097	0.087	0.088	0.086	0.129	0.024	0.032	0.000	0.001	0.065	0.051	0.029	0.028	0.040
1%	0.008	0.022	0.000	0.005	0.039	0.042	0.039	0.042	0.039	0.006	0.009	0.000	0.000	0.014	0.010	0.007	0.010	0.010
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.422	0.403	0.328	0.308	0.433	0.409	0.335	0.315	0.491	0.350	0.354	0.092	0.093	0.453	0.420	0.120	0.109	0.368
5%	0.345	0.327	0.239	0.209	0.362	0.328	0.242	0.210	0.428	0.263	0.276	0.042	0.050	0.378	0.357	0.084	0.078	0.283
1%	0.171	0.188	0.071	0.070	0.209	0.193	0.090	0.072	0.265	0.146	0.160	0.008	0.013	0.259	0.242	0.057	0.058	0.165
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.904	0.920	0.859	0.869	0.918	0.931	0.877	0.884	0.952	0.790	0.791	0.557	0.543	0.855	0.830	0.562	0.548	0.796
5%	0.870	0.889	0.811	0.815	0.880	0.892	0.822	0.823	0.924	0.728	0.735	0.487	0.446	0.801	0.788	0.494	0.451	0.736
1%	0.776	0.791	0.697	0.696	0.796	0.802	0.711	0.709	0.849	0.622	0.642	0.347	0.310	0.743	0.713	0.361	0.321	0.629

Table A.2: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with $t(4)$ errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 100$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
$H_{0,1} : \beta = \beta_n$										$H_{0,1} : \pi = \pi_0$								
10%	0.103	0.084	0.069	0.055	0.156	0.163	0.156	0.163	0.156	0.138	0.128	0.007	0.007	0.153	0.152	0.153	0.152	0.153
5%	0.066	0.049	0.041	0.036	0.093	0.112	0.093	0.112	0.093	0.088	0.087	0.004	0.004	0.095	0.105	0.095	0.105	0.095
1%	0.021	0.021	0.014	0.017	0.032	0.074	0.032	0.073	0.033	0.033	0.033	0.001	0.001	0.033	0.044	0.033	0.044	0.033
$H_{0,2} : \beta = \beta_n + \sigma_\beta$										$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.273	0.291	0.250	0.273	0.384	0.353	0.383	0.353	0.385	0.409	0.409	0.032	0.026	0.418	0.419	0.418	0.419	0.418
5%	0.203	0.228	0.181	0.207	0.284	0.259	0.284	0.258	0.285	0.313	0.312	0.012	0.010	0.321	0.329	0.321	0.329	0.321
1%	0.103	0.107	0.077	0.098	0.149	0.127	0.149	0.127	0.149	0.159	0.177	0.001	0.001	0.169	0.202	0.169	0.202	0.169
$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$										$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.900	0.908	0.885	0.905	0.933	0.918	0.933	0.918	0.933	0.915	0.916	0.552	0.536	0.917	0.918	0.916	0.917	0.916
5%	0.852	0.856	0.834	0.848	0.902	0.866	0.902	0.866	0.902	0.877	0.879	0.424	0.390	0.880	0.888	0.879	0.887	0.879
1%	0.738	0.732	0.691	0.714	0.796	0.749	0.795	0.748	0.798	0.789	0.800	0.197	0.178	0.798	0.808	0.798	0.808	0.798
Weakly identified $n = 100$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
$H_{0,1} : \beta = \beta_n$										$H_{0,1} : \pi = \pi_0$								
10%	0.175	0.162	0.112	0.093	0.193	0.208	0.183	0.190	0.313	0.089	0.086	0.005	0.004	0.120	0.118	0.075	0.075	0.119
5%	0.115	0.099	0.067	0.060	0.129	0.152	0.117	0.143	0.224	0.048	0.043	0.002	0.001	0.063	0.069	0.044	0.053	0.065
1%	0.045	0.041	0.018	0.015	0.055	0.091	0.054	0.090	0.087	0.012	0.010	0.000	0.000	0.019	0.029	0.019	0.027	0.026
$H_{0,2} : \beta = \beta_n + \sigma_\beta$										$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.335	0.334	0.313	0.310	0.346	0.349	0.333	0.331	0.465	0.437	0.438	0.143	0.136	0.485	0.489	0.201	0.201	0.461
5%	0.293	0.288	0.267	0.271	0.301	0.304	0.279	0.290	0.395	0.349	0.349	0.089	0.085	0.417	0.421	0.161	0.161	0.384
1%	0.214	0.217	0.178	0.182	0.216	0.226	0.181	0.191	0.298	0.222	0.225	0.025	0.031	0.286	0.287	0.122	0.125	0.267
$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$										$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.930	0.937	0.920	0.931	0.947	0.947	0.942	0.943	0.967	0.813	0.808	0.621	0.621	0.835	0.834	0.628	0.628	0.830
5%	0.904	0.912	0.874	0.896	0.924	0.926	0.906	0.915	0.950	0.778	0.777	0.552	0.557	0.801	0.804	0.563	0.568	0.795
1%	0.800	0.824	0.763	0.784	0.832	0.846	0.807	0.820	0.897	0.681	0.692	0.429	0.436	0.726	0.723	0.450	0.456	0.712
Non-identified $n = 100$ ($\beta_n = 0$, $\pi_0 = 0$)																		
$H_{0,1} : \beta = \beta_n$										$H_{0,1} : \pi = \pi_0$								
10%	0.193	0.173	0.091	0.078	0.195	0.178	0.152	0.152	0.339	0.077	0.083	0.000	0.000	0.126	0.126	0.059	0.062	0.103
5%	0.124	0.104	0.051	0.046	0.153	0.153	0.152	0.152	0.210	0.042	0.037	0.000	0.000	0.067	0.068	0.043	0.043	0.061
1%	0.052	0.038	0.015	0.014	0.077	0.131	0.077	0.131	0.077	0.006	0.006	0.000	0.000	0.009	0.014	0.008	0.014	0.014
$H_{0,2} : \beta = \beta_n + \sigma_\beta$										$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.466	0.467	0.385	0.376	0.471	0.477	0.393	0.389	0.528	0.406	0.398	0.108	0.107	0.479	0.466	0.144	0.143	0.425
5%	0.412	0.411	0.319	0.305	0.414	0.418	0.323	0.314	0.476	0.314	0.319	0.061	0.063	0.400	0.398	0.112	0.116	0.340
1%	0.292	0.298	0.172	0.180	0.293	0.307	0.175	0.191	0.356	0.196	0.208	0.017	0.018	0.285	0.289	0.093	0.097	0.224
$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$										$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.938	0.943	0.910	0.924	0.942	0.948	0.924	0.933	0.960	0.813	0.809	0.593	0.590	0.846	0.845	0.606	0.603	0.828
5%	0.924	0.928	0.887	0.901	0.928	0.929	0.900	0.905	0.940	0.758	0.764	0.517	0.514	0.802	0.802	0.529	0.529	0.781
1%	0.873	0.877	0.788	0.821	0.880	0.890	0.808	0.827	0.901	0.658	0.677	0.366	0.380	0.723	0.741	0.384	0.400	0.683

Table A.3: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	Strongly identified $n=100$ ($\beta_n=0.9$, $\pi_0=0$)																	
	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
	$H_{0,1}:\beta=\beta_n$									$H_{0,1}:\pi=\pi_0$								
10%	0.105	0.087	0.076	0.063	0.154	0.180	0.154	0.180	0.155	0.140	0.138	0.007	0.007	0.153	0.154	0.153	0.154	0.153
5%	0.069	0.054	0.045	0.045	0.090	0.129	0.090	0.129	0.090	0.096	0.085	0.005	0.005	0.110	0.108	0.110	0.108	0.110
1%	0.027	0.021	0.019	0.020	0.036	0.078	0.036	0.078	0.036	0.039	0.038	0.002	0.001	0.045	0.065	0.045	0.065	0.045
	$H_{0,2}:\beta=\beta_n+\sigma_\beta$									$H_{0,2}:\pi=\pi_0+\sigma_\pi$								
10%	0.342	0.358	0.315	0.346	0.448	0.412	0.448	0.412	0.452	0.457	0.447	0.058	0.039	0.463	0.459	0.463	0.459	0.463
5%	0.271	0.287	0.248	0.273	0.355	0.317	0.355	0.317	0.357	0.368	0.362	0.026	0.019	0.379	0.380	0.379	0.380	0.379
1%	0.177	0.168	0.128	0.151	0.216	0.193	0.216	0.193	0.216	0.203	0.217	0.004	0.003	0.215	0.244	0.215	0.244	0.215
	$H_{0,3}:\beta=\beta_n+3\sigma_\beta$									$H_{0,3}:\pi=\pi_0+3\sigma_\pi$								
10%	0.906	0.912	0.894	0.905	0.935	0.922	0.935	0.922	0.935	0.914	0.914	0.614	0.595	0.920	0.918	0.919	0.917	0.919
5%	0.873	0.879	0.860	0.879	0.902	0.885	0.902	0.885	0.903	0.891	0.889	0.507	0.490	0.896	0.898	0.895	0.897	0.895
1%	0.798	0.801	0.769	0.789	0.837	0.812	0.837	0.812	0.839	0.811	0.825	0.304	0.269	0.822	0.834	0.822	0.834	0.822
	Weakly identified $n=100$ ($\beta_n=0.9/\sqrt{n}$, $\pi_0=0$)																	
	$H_{0,1}:\beta=\beta_n$									$H_{0,1}:\pi=\pi_0$								
10%	0.177	0.157	0.107	0.099	0.204	0.214	0.195	0.205	0.317	0.095	0.087	0.004	0.003	0.129	0.126	0.081	0.084	0.147
5%	0.120	0.101	0.068	0.066	0.131	0.171	0.119	0.161	0.216	0.042	0.041	0.003	0.002	0.069	0.074	0.056	0.058	0.076
1%	0.052	0.044	0.021	0.024	0.065	0.103	0.056	0.101	0.094	0.011	0.011	0.000	0.000	0.022	0.030	0.022	0.029	0.029
	$H_{0,2}:\beta=\beta_n+\sigma_\beta$									$H_{0,2}:\pi=\pi_0+\sigma_\pi$								
10%	0.380	0.386	0.366	0.357	0.403	0.415	0.393	0.390	0.541	0.435	0.444	0.134	0.123	0.505	0.505	0.200	0.196	0.472
5%	0.331	0.340	0.309	0.313	0.347	0.362	0.328	0.337	0.468	0.355	0.359	0.077	0.080	0.426	0.422	0.167	0.167	0.402
1%	0.252	0.255	0.209	0.225	0.260	0.265	0.219	0.238	0.350	0.224	0.231	0.028	0.030	0.283	0.294	0.131	0.137	0.272
	$H_{0,3}:\beta=\beta_n+3\sigma_\beta$									$H_{0,3}:\pi=\pi_0+3\sigma_\pi$								
10%	0.937	0.945	0.929	0.939	0.941	0.950	0.934	0.946	0.957	0.829	0.830	0.646	0.640	0.847	0.848	0.657	0.651	0.850
5%	0.914	0.925	0.905	0.922	0.926	0.935	0.919	0.932	0.949	0.795	0.798	0.572	0.566	0.820	0.820	0.585	0.579	0.815
1%	0.871	0.888	0.846	0.869	0.888	0.906	0.870	0.881	0.910	0.707	0.720	0.422	0.439	0.743	0.751	0.450	0.468	0.748
	Non-identified $n=100$ ($\beta_n=0$, $\pi_0=0$)																	
	$H_{0,1}:\beta=\beta_n$									$H_{0,1}:\pi=\pi_0$								
10%	0.198	0.180	0.117	0.091	0.199	0.188	0.170	0.167	0.359	0.082	0.082	0.005	0.004	0.128	0.119	0.067	0.066	0.127
5%	0.141	0.113	0.068	0.052	0.172	0.167	0.166	0.167	0.212	0.044	0.051	0.001	0.001	0.070	0.066	0.046	0.048	0.076
1%	0.050	0.053	0.020	0.016	0.092	0.154	0.092	0.154	0.092	0.008	0.010	0.000	0.000	0.015	0.021	0.012	0.017	0.021
	$H_{0,2}:\beta=\beta_n+\sigma_\beta$									$H_{0,2}:\pi=\pi_0+\sigma_\pi$								
10%	0.510	0.510	0.439	0.423	0.516	0.523	0.445	0.438	0.575	0.398	0.395	0.096	0.090	0.463	0.468	0.144	0.139	0.427
5%	0.463	0.462	0.357	0.357	0.465	0.473	0.362	0.369	0.519	0.311	0.317	0.050	0.055	0.389	0.388	0.115	0.115	0.351
1%	0.351	0.361	0.229	0.244	0.351	0.373	0.232	0.253	0.410	0.175	0.193	0.020	0.016	0.258	0.271	0.096	0.099	0.213
	$H_{0,3}:\beta=\beta_n+3\sigma_\beta$									$H_{0,3}:\pi=\pi_0+3\sigma_\pi$								
10%	0.944	0.950	0.929	0.936	0.950	0.958	0.941	0.946	0.961	0.806	0.806	0.594	0.586	0.852	0.851	0.607	0.603	0.826
5%	0.934	0.937	0.907	0.916	0.939	0.947	0.918	0.927	0.949	0.758	0.761	0.505	0.506	0.806	0.808	0.519	0.520	0.783
1%	0.888	0.904	0.828	0.863	0.895	0.918	0.842	0.875	0.914	0.669	0.673	0.340	0.362	0.724	0.722	0.360	0.383	0.692

Table A.4: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with GARCH(0.1,0.6,0.3) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n=\log(n)^{1/2}$. Sample size $n=250$, simulations $S=1000$, bootstrapped samples $M_n=500$. The grid contains 100 combinations of $b\in[-1,1]$ and $\pi\in[-2,2]$, with true values $b=\sqrt{n}\beta_n$ and $\pi_0=0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 250$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.008	0.039	0.006	0.021	0.105	0.105	0.105	0.105	0.105	0.067	0.094	0.000	0.001	0.100	0.100	0.100	0.100	0.100
5%	0.006	0.023	0.004	0.010	0.051	0.054	0.051	0.054	0.051	0.035	0.052	0.000	0.000	0.052	0.054	0.052	0.054	0.052
1%	0.002	0.006	0.002	0.002	0.015	0.014	0.015	0.014	0.015	0.006	0.008	0.000	0.000	0.010	0.010	0.010	0.010	0.010
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.036	0.098	0.029	0.081	0.258	0.255	0.258	0.255	0.258	0.220	0.272	0.001	0.028	0.273	0.275	0.273	0.275	0.273
5%	0.016	0.061	0.009	0.047	0.163	0.166	0.163	0.166	0.163	0.143	0.183	0.000	0.013	0.186	0.190	0.186	0.190	0.186
1%	0.006	0.018	0.004	0.011	0.064	0.076	0.064	0.076	0.064	0.042	0.066	0.000	0.000	0.060	0.076	0.060	0.076	0.060
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.451	0.798	0.390	0.771	0.944	0.942	0.944	0.942	0.944	0.888	0.919	0.090	0.523	0.918	0.922	0.918	0.922	0.918
5%	0.305	0.702	0.233	0.661	0.895	0.890	0.895	0.890	0.895	0.832	0.871	0.029	0.334	0.863	0.875	0.863	0.875	0.863
1%	0.126	0.490	0.073	0.395	0.709	0.723	0.709	0.723	0.709	0.643	0.700	0.001	0.110	0.714	0.706	0.714	0.706	0.714
Weakly identified $n = 250$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.079	0.089	0.043	0.047	0.106	0.100	0.091	0.087	0.239	0.053	0.063	0.001	0.002	0.100	0.102	0.031	0.031	0.074
5%	0.032	0.038	0.017	0.017	0.054	0.052	0.046	0.046	0.131	0.018	0.022	0.000	0.000	0.044	0.051	0.021	0.022	0.029
1%	0.007	0.008	0.003	0.003	0.017	0.015	0.016	0.014	0.028	0.002	0.007	0.000	0.000	0.007	0.011	0.004	0.010	0.005
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.302	0.299	0.290	0.286	0.304	0.300	0.292	0.288	0.384	0.369	0.360	0.090	0.081	0.464	0.431	0.107	0.095	0.372
5%	0.261	0.262	0.241	0.230	0.263	0.263	0.243	0.231	0.344	0.278	0.286	0.052	0.043	0.384	0.360	0.073	0.059	0.288
1%	0.157	0.153	0.110	0.104	0.159	0.153	0.113	0.104	0.260	0.141	0.146	0.010	0.014	0.248	0.228	0.040	0.039	0.152
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.872	0.876	0.853	0.861	0.896	0.897	0.879	0.884	0.970	0.785	0.789	0.611	0.583	0.818	0.817	0.614	0.588	0.793
5%	0.810	0.816	0.781	0.786	0.822	0.828	0.793	0.798	0.933	0.739	0.746	0.533	0.499	0.791	0.785	0.537	0.504	0.744
1%	0.672	0.666	0.622	0.613	0.678	0.674	0.627	0.620	0.825	0.653	0.655	0.383	0.378	0.726	0.709	0.390	0.383	0.660
Non-identified $n = 250$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.102	0.098	0.039	0.043	0.111	0.099	0.058	0.048	0.215	0.046	0.056	0.002	0.001	0.117	0.100	0.027	0.026	0.054
5%	0.053	0.055	0.012	0.017	0.070	0.057	0.049	0.048	0.113	0.021	0.029	0.001	0.001	0.053	0.051	0.017	0.020	0.028
1%	0.006	0.017	0.000	0.002	0.035	0.034	0.032	0.034	0.032	0.003	0.006	0.000	0.000	0.012	0.010	0.007	0.008	0.007
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.440	0.439	0.337	0.302	0.448	0.442	0.343	0.305	0.500	0.362	0.359	0.081	0.085	0.460	0.433	0.094	0.097	0.365
5%	0.364	0.339	0.215	0.206	0.373	0.340	0.219	0.206	0.457	0.260	0.270	0.044	0.046	0.405	0.373	0.061	0.059	0.268
1%	0.165	0.165	0.073	0.068	0.184	0.170	0.079	0.068	0.265	0.140	0.156	0.010	0.013	0.248	0.244	0.032	0.030	0.149
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.920	0.928	0.871	0.872	0.926	0.934	0.878	0.878	0.969	0.784	0.783	0.579	0.547	0.829	0.832	0.582	0.550	0.788
5%	0.886	0.885	0.807	0.807	0.892	0.889	0.813	0.812	0.936	0.727	0.735	0.488	0.475	0.797	0.783	0.491	0.478	0.731
1%	0.775	0.780	0.675	0.675	0.786	0.785	0.681	0.678	0.844	0.626	0.634	0.348	0.337	0.722	0.706	0.352	0.338	0.632

Table A.5: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with $t(4)$ errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 250$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
$H_{0,1} : \beta = \beta_n$										$H_{0,1} : \pi = \pi_0$								
10%	0.106	0.089	0.073	0.066	0.151	0.170	0.151	0.170	0.151	0.130	0.133	0.003	0.001	0.139	0.150	0.139	0.150	0.140
5%	0.061	0.059	0.048	0.040	0.088	0.121	0.088	0.121	0.088	0.076	0.085	0.001	0.000	0.083	0.101	0.083	0.101	0.084
1%	0.025	0.028	0.021	0.021	0.031	0.069	0.031	0.069	0.031	0.024	0.033	0.000	0.000	0.030	0.052	0.030	0.052	0.030
$H_{0,2} : \beta = \beta_n + \sigma_\beta$										$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.338	0.351	0.312	0.335	0.453	0.433	0.453	0.433	0.454	0.447	0.444	0.033	0.028	0.455	0.457	0.454	0.456	0.455
5%	0.252	0.270	0.221	0.248	0.347	0.319	0.347	0.319	0.348	0.355	0.354	0.015	0.011	0.361	0.373	0.360	0.373	0.361
1%	0.149	0.147	0.115	0.129	0.186	0.169	0.186	0.169	0.186	0.168	0.184	0.001	0.002	0.180	0.210	0.180	0.210	0.181
$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$										$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.910	0.917	0.904	0.915	0.934	0.919	0.934	0.919	0.934	0.919	0.914	0.672	0.668	0.926	0.921	0.925	0.920	0.926
5%	0.889	0.893	0.875	0.890	0.912	0.895	0.911	0.895	0.912	0.894	0.889	0.572	0.554	0.899	0.897	0.898	0.896	0.899
1%	0.834	0.840	0.800	0.834	0.861	0.848	0.861	0.848	0.862	0.834	0.834	0.328	0.307	0.839	0.853	0.839	0.853	0.840
Weakly identified $n = 250$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
$H_{0,1} : \beta = \beta_n$										$H_{0,1} : \pi = \pi_0$								
10%	0.136	0.117	0.082	0.073	0.153	0.141	0.138	0.133	0.281	0.080	0.074	0.004	0.002	0.113	0.103	0.050	0.049	0.104
5%	0.092	0.075	0.046	0.045	0.105	0.107	0.087	0.098	0.184	0.044	0.043	0.000	0.000	0.064	0.065	0.039	0.041	0.067
1%	0.030	0.024	0.016	0.017	0.037	0.056	0.036	0.056	0.058	0.014	0.012	0.000	0.000	0.022	0.024	0.020	0.024	0.023
$H_{0,2} : \beta = \beta_n + \sigma_\beta$										$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.401	0.399	0.379	0.377	0.405	0.411	0.386	0.389	0.553	0.393	0.394	0.099	0.101	0.458	0.454	0.121	0.123	0.416
5%	0.350	0.351	0.327	0.330	0.352	0.356	0.330	0.336	0.478	0.318	0.323	0.058	0.060	0.384	0.385	0.086	0.089	0.347
1%	0.264	0.262	0.226	0.233	0.265	0.268	0.228	0.238	0.357	0.187	0.193	0.024	0.027	0.246	0.240	0.069	0.071	0.206
$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$										$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.930	0.936	0.920	0.928	0.934	0.939	0.925	0.932	0.957	0.810	0.808	0.591	0.589	0.822	0.828	0.600	0.598	0.825
5%	0.909	0.918	0.900	0.912	0.911	0.923	0.902	0.918	0.939	0.765	0.760	0.520	0.526	0.792	0.786	0.530	0.535	0.783
1%	0.863	0.882	0.833	0.863	0.869	0.892	0.841	0.872	0.908	0.670	0.681	0.382	0.400	0.708	0.715	0.395	0.414	0.700
Non-identified $n = 250$ ($\beta_n = 0$, $\pi_0 = 0$)																		
$H_{0,1} : \beta = \beta_n$										$H_{0,1} : \pi = \pi_0$								
10%	0.164	0.139	0.083	0.071	0.164	0.140	0.103	0.098	0.302	0.074	0.073	0.001	0.000	0.098	0.096	0.040	0.038	0.103
5%	0.105	0.082	0.038	0.036	0.109	0.100	0.095	0.096	0.182	0.029	0.034	0.000	0.000	0.045	0.054	0.026	0.031	0.054
1%	0.030	0.024	0.004	0.009	0.051	0.091	0.051	0.091	0.051	0.004	0.004	0.000	0.000	0.009	0.014	0.006	0.011	0.011
$H_{0,2} : \beta = \beta_n + \sigma_\beta$										$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.522	0.515	0.460	0.463	0.528	0.526	0.466	0.474	0.583	0.368	0.370	0.081	0.079	0.442	0.437	0.114	0.112	0.392
5%	0.477	0.473	0.401	0.405	0.479	0.485	0.403	0.416	0.531	0.296	0.297	0.041	0.036	0.365	0.369	0.076	0.076	0.321
1%	0.368	0.386	0.250	0.262	0.370	0.395	0.250	0.269	0.432	0.157	0.173	0.009	0.012	0.227	0.238	0.052	0.055	0.188
$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$										$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.956	0.958	0.942	0.947	0.957	0.962	0.947	0.954	0.967	0.810	0.805	0.569	0.570	0.838	0.841	0.577	0.579	0.819
5%	0.938	0.946	0.916	0.933	0.942	0.955	0.922	0.938	0.959	0.765	0.769	0.489	0.501	0.805	0.805	0.501	0.512	0.778
1%	0.898	0.913	0.858	0.884	0.901	0.926	0.865	0.891	0.916	0.649	0.668	0.349	0.361	0.718	0.730	0.364	0.378	0.665

Table A.6: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with GARCH(0.1,0.6,0.3) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 500$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.004	0.038	0.002	0.012	0.092	0.089	0.092	0.089	0.092	0.080	0.099	0.000	0.003	0.097	0.099	0.097	0.099	0.097
5%	0.002	0.019	0.000	0.005	0.044	0.050	0.044	0.050	0.044	0.036	0.048	0.000	0.000	0.047	0.049	0.047	0.049	0.047
1%	0.000	0.003	0.000	0.001	0.009	0.011	0.009	0.011	0.009	0.005	0.014	0.000	0.000	0.010	0.014	0.010	0.014	0.010
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.018	0.065	0.011	0.053	0.244	0.246	0.244	0.246	0.244	0.215	0.245	0.000	0.021	0.248	0.245	0.248	0.245	0.248
5%	0.007	0.037	0.006	0.026	0.152	0.159	0.152	0.159	0.152	0.123	0.144	0.000	0.007	0.135	0.145	0.135	0.145	0.135
1%	0.001	0.011	0.000	0.008	0.045	0.048	0.045	0.048	0.045	0.045	0.056	0.000	0.000	0.052	0.060	0.052	0.060	0.052
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.415	0.749	0.358	0.706	0.904	0.908	0.904	0.908	0.904	0.880	0.892	0.060	0.463	0.897	0.894	0.897	0.894	0.897
5%	0.256	0.638	0.165	0.581	0.851	0.850	0.851	0.850	0.851	0.816	0.827	0.011	0.291	0.830	0.829	0.830	0.829	0.830
1%	0.065	0.398	0.034	0.311	0.659	0.659	0.659	0.659	0.659	0.600	0.637	0.000	0.073	0.621	0.644	0.621	0.644	0.621
Weakly identified $n = 500$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.100	0.105	0.054	0.056	0.110	0.110	0.099	0.100	0.262	0.067	0.067	0.000	0.001	0.103	0.107	0.027	0.027	0.068
5%	0.063	0.054	0.027	0.023	0.069	0.068	0.057	0.060	0.153	0.024	0.027	0.000	0.000	0.050	0.055	0.017	0.020	0.024
1%	0.012	0.012	0.004	0.003	0.015	0.015	0.014	0.014	0.047	0.003	0.003	0.000	0.000	0.006	0.005	0.003	0.004	0.003
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.280	0.278	0.267	0.267	0.281	0.279	0.269	0.270	0.364	0.362	0.361	0.106	0.108	0.427	0.424	0.115	0.114	0.363
5%	0.241	0.247	0.220	0.226	0.242	0.247	0.221	0.227	0.315	0.282	0.285	0.071	0.072	0.355	0.353	0.082	0.081	0.284
1%	0.154	0.150	0.115	0.114	0.154	0.151	0.115	0.114	0.254	0.156	0.160	0.018	0.017	0.233	0.225	0.045	0.043	0.156
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.876	0.876	0.852	0.857	0.888	0.886	0.864	0.869	0.962	0.813	0.813	0.566	0.570	0.842	0.840	0.568	0.573	0.813
5%	0.815	0.819	0.794	0.787	0.824	0.827	0.803	0.796	0.938	0.763	0.759	0.503	0.496	0.810	0.812	0.506	0.499	0.763
1%	0.684	0.672	0.614	0.619	0.687	0.675	0.620	0.623	0.840	0.653	0.659	0.375	0.374	0.727	0.726	0.380	0.379	0.657
Non-identified $n = 500$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.092	0.094	0.026	0.031	0.092	0.094	0.031	0.032	0.240	0.039	0.042	0.000	0.000	0.097	0.087	0.012	0.013	0.039
5%	0.044	0.044	0.013	0.012	0.047	0.044	0.028	0.028	0.113	0.013	0.015	0.000	0.000	0.037	0.042	0.009	0.009	0.013
1%	0.013	0.010	0.002	0.002	0.024	0.024	0.024	0.024	0.024	0.004	0.004	0.000	0.000	0.009	0.010	0.004	0.004	0.004
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.425	0.435	0.297	0.290	0.429	0.438	0.302	0.294	0.496	0.366	0.370	0.088	0.090	0.436	0.438	0.092	0.094	0.369
5%	0.342	0.350	0.199	0.197	0.343	0.351	0.200	0.198	0.451	0.292	0.293	0.041	0.044	0.376	0.377	0.047	0.050	0.292
1%	0.171	0.175	0.062	0.061	0.174	0.180	0.062	0.061	0.260	0.148	0.159	0.012	0.007	0.258	0.261	0.018	0.016	0.148
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.927	0.928	0.860	0.854	0.927	0.928	0.860	0.855	0.968	0.803	0.804	0.562	0.562	0.851	0.845	0.564	0.564	0.803
5%	0.883	0.886	0.792	0.791	0.883	0.886	0.794	0.792	0.937	0.752	0.754	0.487	0.480	0.814	0.811	0.490	0.483	0.753
1%	0.777	0.768	0.658	0.650	0.779	0.774	0.660	0.652	0.837	0.640	0.643	0.349	0.338	0.738	0.731	0.354	0.343	0.640

Table A.7: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with Normal(0,1) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 500$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1}: \beta = \beta_n$									$H_{0,1}: \pi = \pi_0$								
10%	0.001	0.046	0.000	0.017	0.099	0.100	0.099	0.100	0.099	0.082	0.102	0.000	0.001	0.112	0.103	0.112	0.103	0.112
5%	0.000	0.023	0.000	0.008	0.052	0.052	0.052	0.052	0.052	0.040	0.060	0.000	0.000	0.059	0.063	0.059	0.063	0.059
1%	0.000	0.004	0.000	0.000	0.009	0.011	0.009	0.011	0.009	0.005	0.012	0.000	0.000	0.009	0.012	0.009	0.012	0.009
	$H_{0,2}: \beta = \beta_n + \sigma_\beta$									$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.009	0.075	0.003	0.058	0.243	0.238	0.243	0.238	0.243	0.239	0.285	0.000	0.031	0.286	0.291	0.286	0.291	0.286
5%	0.002	0.044	0.001	0.031	0.153	0.156	0.153	0.156	0.153	0.158	0.202	0.000	0.007	0.204	0.205	0.204	0.205	0.204
1%	0.000	0.013	0.000	0.006	0.043	0.051	0.043	0.051	0.043	0.061	0.085	0.000	0.000	0.083	0.089	0.083	0.089	0.083
	$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$									$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.339	0.754	0.284	0.727	0.908	0.904	0.908	0.904	0.908	0.888	0.916	0.088	0.528	0.919	0.918	0.919	0.918	0.919
5%	0.199	0.657	0.143	0.596	0.862	0.849	0.862	0.849	0.862	0.828	0.864	0.036	0.361	0.870	0.866	0.870	0.866	0.870
1%	0.057	0.416	0.027	0.322	0.657	0.665	0.657	0.665	0.657	0.630	0.690	0.000	0.129	0.709	0.698	0.709	0.698	0.709
Weakly identified $n = 500$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1}: \beta = \beta_n$									$H_{0,1}: \pi = \pi_0$								
10%	0.083	0.085	0.048	0.046	0.097	0.092	0.080	0.078	0.248	0.046	0.054	0.001	0.002	0.108	0.098	0.022	0.021	0.057
5%	0.033	0.040	0.020	0.025	0.048	0.051	0.042	0.046	0.133	0.021	0.027	0.000	0.001	0.049	0.052	0.018	0.019	0.027
1%	0.009	0.012	0.004	0.005	0.013	0.012	0.011	0.010	0.031	0.004	0.005	0.000	0.000	0.005	0.008	0.004	0.006	0.004
	$H_{0,2}: \beta = \beta_n + \sigma_\beta$									$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.297	0.293	0.291	0.284	0.299	0.293	0.295	0.288	0.355	0.360	0.364	0.094	0.099	0.439	0.419	0.100	0.104	0.365
5%	0.267	0.254	0.241	0.234	0.269	0.255	0.244	0.237	0.325	0.288	0.287	0.050	0.050	0.370	0.365	0.062	0.058	0.293
1%	0.162	0.164	0.131	0.111	0.163	0.165	0.131	0.112	0.266	0.150	0.161	0.014	0.015	0.253	0.242	0.033	0.031	0.155
	$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$									$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.876	0.869	0.859	0.838	0.886	0.879	0.870	0.849	0.970	0.790	0.788	0.589	0.574	0.823	0.824	0.591	0.575	0.794
5%	0.813	0.792	0.777	0.767	0.818	0.798	0.783	0.773	0.945	0.747	0.749	0.521	0.512	0.798	0.791	0.523	0.514	0.752
1%	0.662	0.652	0.602	0.581	0.665	0.654	0.604	0.584	0.812	0.642	0.633	0.398	0.380	0.723	0.703	0.403	0.385	0.647
Non-identified $n = 500$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1}: \beta = \beta_n$									$H_{0,1}: \pi = \pi_0$								
10%	0.103	0.101	0.035	0.036	0.106	0.101	0.042	0.036	0.222	0.043	0.044	0.000	0.000	0.101	0.095	0.014	0.013	0.048
5%	0.050	0.055	0.013	0.019	0.060	0.055	0.034	0.033	0.117	0.014	0.018	0.000	0.000	0.050	0.045	0.009	0.010	0.018
1%	0.009	0.010	0.000	0.001	0.026	0.028	0.026	0.028	0.026	0.001	0.002	0.000	0.000	0.005	0.006	0.000	0.002	0.001
	$H_{0,2}: \beta = \beta_n + \sigma_\beta$									$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.447	0.441	0.336	0.304	0.456	0.449	0.344	0.312	0.510	0.348	0.350	0.080	0.083	0.452	0.447	0.089	0.088	0.353
5%	0.372	0.354	0.216	0.199	0.375	0.357	0.217	0.202	0.461	0.253	0.250	0.043	0.042	0.383	0.355	0.052	0.048	0.260
1%	0.176	0.172	0.078	0.068	0.192	0.178	0.081	0.069	0.257	0.136	0.138	0.010	0.013	0.241	0.226	0.024	0.024	0.140
	$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$									$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.921	0.931	0.876	0.872	0.923	0.934	0.877	0.875	0.964	0.799	0.801	0.566	0.557	0.843	0.838	0.568	0.559	0.802
5%	0.888	0.890	0.814	0.809	0.889	0.891	0.817	0.813	0.941	0.748	0.749	0.484	0.463	0.807	0.804	0.486	0.465	0.751
1%	0.773	0.769	0.669	0.666	0.784	0.774	0.670	0.668	0.839	0.634	0.633	0.345	0.324	0.737	0.724	0.348	0.325	0.635

Table A.8: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with $t(4)$ errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	$LF\ \beta$				$ICS_1\ \beta$					$LF\ \pi$				$ICS_1\ \pi$				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 500$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.065	0.059	0.051	0.038	0.109	0.121	0.109	0.121	0.109	0.099	0.097	0.000	0.000	0.102	0.104	0.102	0.104	0.102
5%	0.040	0.032	0.026	0.023	0.053	0.070	0.053	0.070	0.053	0.053	0.051	0.000	0.000	0.054	0.058	0.054	0.058	0.054
1%	0.014	0.010	0.007	0.005	0.018	0.030	0.018	0.030	0.018	0.007	0.010	0.000	0.000	0.008	0.013	0.008	0.013	0.008
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.200	0.200	0.174	0.182	0.338	0.302	0.338	0.302	0.338	0.266	0.270	0.008	0.002	0.268	0.276	0.268	0.276	0.268
5%	0.139	0.140	0.110	0.122	0.212	0.190	0.212	0.190	0.212	0.196	0.203	0.001	0.000	0.197	0.210	0.197	0.210	0.197
1%	0.058	0.054	0.044	0.046	0.081	0.074	0.081	0.074	0.081	0.075	0.083	0.000	0.000	0.076	0.090	0.076	0.090	0.076
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.903	0.907	0.888	0.903	0.932	0.920	0.932	0.920	0.932	0.890	0.892	0.416	0.372	0.893	0.894	0.893	0.894	0.893
5%	0.860	0.869	0.828	0.856	0.907	0.885	0.907	0.885	0.907	0.851	0.852	0.261	0.210	0.852	0.855	0.852	0.855	0.852
1%	0.710	0.714	0.654	0.685	0.769	0.733	0.769	0.733	0.769	0.713	0.720	0.060	0.054	0.715	0.726	0.715	0.726	0.715
Weakly identified $n = 500$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.111	0.102	0.065	0.058	0.116	0.112	0.098	0.096	0.285	0.062	0.059	0.002	0.004	0.094	0.095	0.029	0.032	0.068
5%	0.060	0.057	0.032	0.030	0.072	0.076	0.062	0.068	0.172	0.029	0.028	0.001	0.001	0.041	0.047	0.025	0.023	0.032
1%	0.017	0.013	0.008	0.005	0.026	0.039	0.026	0.038	0.045	0.005	0.006	0.000	0.000	0.007	0.010	0.005	0.007	0.009
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.342	0.339	0.329	0.329	0.343	0.344	0.330	0.334	0.448	0.384	0.389	0.100	0.098	0.448	0.450	0.116	0.115	0.397
5%	0.301	0.299	0.275	0.276	0.301	0.300	0.275	0.277	0.395	0.301	0.305	0.060	0.062	0.386	0.386	0.078	0.079	0.314
1%	0.207	0.199	0.166	0.168	0.207	0.199	0.166	0.168	0.302	0.167	0.182	0.018	0.019	0.242	0.254	0.043	0.047	0.181
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.919	0.924	0.903	0.911	0.932	0.932	0.920	0.923	0.960	0.809	0.808	0.586	0.577	0.845	0.843	0.592	0.583	0.812
5%	0.881	0.888	0.862	0.874	0.896	0.896	0.880	0.885	0.941	0.748	0.747	0.508	0.506	0.791	0.792	0.515	0.513	0.758
1%	0.781	0.801	0.751	0.767	0.791	0.811	0.762	0.777	0.881	0.640	0.651	0.388	0.403	0.695	0.704	0.395	0.411	0.654
Non-identified $n = 500$ ($\beta_n = 0$, $\pi_0 = 0$)																		
	$H_{0,1} : \beta = \beta_n$									$H_{0,1} : \pi = \pi_0$								
10%	0.133	0.114	0.060	0.048	0.133	0.114	0.060	0.050	0.271	0.064	0.068	0.000	0.000	0.109	0.105	0.021	0.022	0.073
5%	0.066	0.064	0.023	0.022	0.067	0.065	0.040	0.041	0.143	0.029	0.027	0.000	0.000	0.056	0.059	0.018	0.019	0.037
1%	0.017	0.020	0.005	0.004	0.033	0.036	0.033	0.036	0.033	0.005	0.006	0.000	0.000	0.007	0.009	0.007	0.006	0.009
	$H_{0,2} : \beta = \beta_n + \sigma_\beta$									$H_{0,2} : \pi = \pi_0 + \sigma_\pi$								
10%	0.467	0.456	0.383	0.378	0.470	0.463	0.388	0.385	0.532	0.374	0.377	0.108	0.105	0.447	0.447	0.115	0.111	0.383
5%	0.404	0.411	0.287	0.284	0.407	0.416	0.290	0.289	0.476	0.294	0.300	0.053	0.047	0.381	0.384	0.062	0.056	0.302
1%	0.254	0.245	0.125	0.123	0.254	0.247	0.126	0.123	0.336	0.174	0.185	0.010	0.012	0.265	0.264	0.028	0.028	0.181
	$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$									$H_{0,3} : \pi = \pi_0 + 3\sigma_\pi$								
10%	0.945	0.948	0.918	0.926	0.949	0.951	0.925	0.932	0.964	0.790	0.795	0.559	0.558	0.832	0.828	0.561	0.560	0.796
5%	0.932	0.932	0.878	0.889	0.934	0.934	0.884	0.893	0.947	0.741	0.743	0.484	0.490	0.787	0.787	0.487	0.492	0.745
1%	0.858	0.869	0.781	0.793	0.860	0.879	0.787	0.798	0.901	0.645	0.653	0.340	0.352	0.698	0.701	0.344	0.356	0.658

Table A.9: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

	LF β				ICS_1 β					LF π				ICS_1 π				
	No nuis		Nuis		No nuis		Nuis			No nuis		Nuis		No nuis		Nuis		
	AC	WB	AC	WB	AC	WB	AC	WB	Stand	AC	WB	AC	WB	AC	WB	AC	WB	Stand
Strongly identified $n = 500$ ($\beta_n = 0.9$, $\pi_0 = 0$)																		
$H_{0,1}: \beta = \beta_n$										$H_{0,1}: \pi = \pi_0$								
10%	0.089	0.070	0.056	0.046	0.126	0.135	0.126	0.135	0.127	0.102	0.103	0.001	0.001	0.111	0.127	0.111	0.127	0.111
5%	0.048	0.042	0.034	0.029	0.071	0.095	0.071	0.095	0.071	0.057	0.059	0.001	0.001	0.063	0.074	0.063	0.074	0.063
1%	0.013	0.012	0.007	0.008	0.020	0.045	0.020	0.045	0.020	0.009	0.017	0.001	0.001	0.012	0.025	0.012	0.025	0.012
$H_{0,2}: \beta = \beta_n + \sigma_\beta$										$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.310	0.322	0.281	0.301	0.443	0.420	0.443	0.420	0.444	0.394	0.384	0.024	0.017	0.404	0.396	0.404	0.396	0.404
5%	0.235	0.243	0.200	0.218	0.343	0.311	0.342	0.311	0.343	0.284	0.291	0.006	0.003	0.292	0.304	0.292	0.304	0.292
1%	0.111	0.107	0.085	0.094	0.159	0.144	0.159	0.144	0.159	0.133	0.154	0.002	0.001	0.141	0.170	0.141	0.170	0.141
$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$										$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.907	0.908	0.899	0.905	0.928	0.916	0.928	0.916	0.929	0.899	0.896	0.644	0.632	0.902	0.899	0.902	0.899	0.903
5%	0.884	0.885	0.865	0.883	0.905	0.889	0.905	0.889	0.906	0.872	0.870	0.524	0.504	0.875	0.874	0.875	0.874	0.875
1%	0.815	0.816	0.776	0.808	0.842	0.819	0.842	0.819	0.842	0.793	0.797	0.285	0.272	0.800	0.810	0.800	0.810	0.800
Weakly identified $n = 500$ ($\beta_n = 0.9/\sqrt{n}$, $\pi_0 = 0$)																		
$H_{0,1}: \beta = \beta_n$										$H_{0,1}: \pi = \pi_0$								
10%	0.134	0.123	0.087	0.075	0.144	0.135	0.123	0.114	0.311	0.068	0.064	0.003	0.002	0.106	0.104	0.028	0.029	0.092
5%	0.078	0.065	0.038	0.038	0.086	0.086	0.069	0.076	0.190	0.037	0.035	0.000	0.000	0.058	0.052	0.024	0.023	0.055
1%	0.025	0.021	0.011	0.010	0.032	0.044	0.030	0.043	0.051	0.007	0.010	0.000	0.000	0.013	0.018	0.009	0.013	0.019
$H_{0,2}: \beta = \beta_n + \sigma_\beta$										$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.457	0.457	0.442	0.438	0.460	0.463	0.446	0.445	0.584	0.419	0.413	0.112	0.112	0.470	0.467	0.130	0.132	0.433
5%	0.405	0.401	0.369	0.376	0.406	0.408	0.371	0.384	0.524	0.320	0.327	0.059	0.062	0.384	0.394	0.081	0.085	0.345
1%	0.310	0.317	0.267	0.289	0.310	0.323	0.267	0.295	0.404	0.191	0.196	0.020	0.021	0.255	0.265	0.049	0.050	0.209
$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$										$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.931	0.939	0.927	0.936	0.934	0.943	0.930	0.940	0.956	0.825	0.822	0.586	0.587	0.840	0.848	0.589	0.590	0.836
5%	0.916	0.926	0.909	0.922	0.919	0.932	0.913	0.927	0.939	0.771	0.770	0.524	0.529	0.807	0.802	0.528	0.532	0.780
1%	0.887	0.899	0.860	0.888	0.890	0.903	0.868	0.892	0.908	0.672	0.674	0.385	0.416	0.727	0.723	0.392	0.422	0.695
Non-identified $n = 500$ ($\beta_n = 0$, $\pi_0 = 0$)																		
$H_{0,1}: \beta = \beta_n$										$H_{0,1}: \pi = \pi_0$								
10%	0.149	0.134	0.077	0.071	0.149	0.135	0.077	0.071	0.294	0.084	0.081	0.003	0.003	0.121	0.120	0.023	0.024	0.109
5%	0.080	0.078	0.041	0.037	0.080	0.080	0.057	0.054	0.172	0.032	0.031	0.000	0.000	0.058	0.062	0.018	0.017	0.058
1%	0.023	0.030	0.011	0.012	0.042	0.054	0.042	0.054	0.042	0.006	0.007	0.000	0.000	0.013	0.017	0.007	0.009	0.012
$H_{0,2}: \beta = \beta_n + \sigma_\beta$										$H_{0,2}: \pi = \pi_0 + \sigma_\pi$								
10%	0.568	0.562	0.504	0.507	0.570	0.567	0.506	0.512	0.627	0.377	0.372	0.101	0.099	0.446	0.442	0.110	0.111	0.390
5%	0.517	0.514	0.443	0.446	0.519	0.519	0.445	0.450	0.571	0.294	0.301	0.054	0.060	0.366	0.372	0.069	0.076	0.306
1%	0.404	0.416	0.300	0.318	0.404	0.421	0.300	0.322	0.466	0.174	0.186	0.011	0.011	0.248	0.252	0.039	0.041	0.187
$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$										$H_{0,3}: \pi = \pi_0 + 3\sigma_\pi$								
10%	0.957	0.958	0.952	0.956	0.959	0.963	0.954	0.959	0.966	0.796	0.794	0.570	0.567	0.828	0.828	0.572	0.569	0.807
5%	0.950	0.952	0.939	0.947	0.950	0.958	0.939	0.950	0.958	0.740	0.746	0.500	0.513	0.780	0.789	0.503	0.515	0.754
1%	0.920	0.934	0.885	0.910	0.921	0.944	0.889	0.912	0.939	0.646	0.661	0.352	0.372	0.698	0.706	0.357	0.374	0.666

Table A.10: Size and power of t-statistic under strong, weak and non-identification for the Smoothing Exponential model with GARCH(0.1,0.6,0.3) errors. Numerical values are rejection probabilities at the given level. AC, WB and Stand denote Andrews and Cheng, wild bootstrapped and standard t-statistic rejection probabilities, respectively. Nuis and No Nuis denotes known and unknown nuisance parameters, repectively. LF denotes Least Favorable and ICS_1 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, simulations $S = 1000$, bootstrapped samples $M_n = 500$. The grid contains 100 combinations of $b \in [-1, 1]$ and $\pi \in [-2, 2]$, with true values $b = \sqrt{n}\beta_n$ and $\pi_0 = 0$.

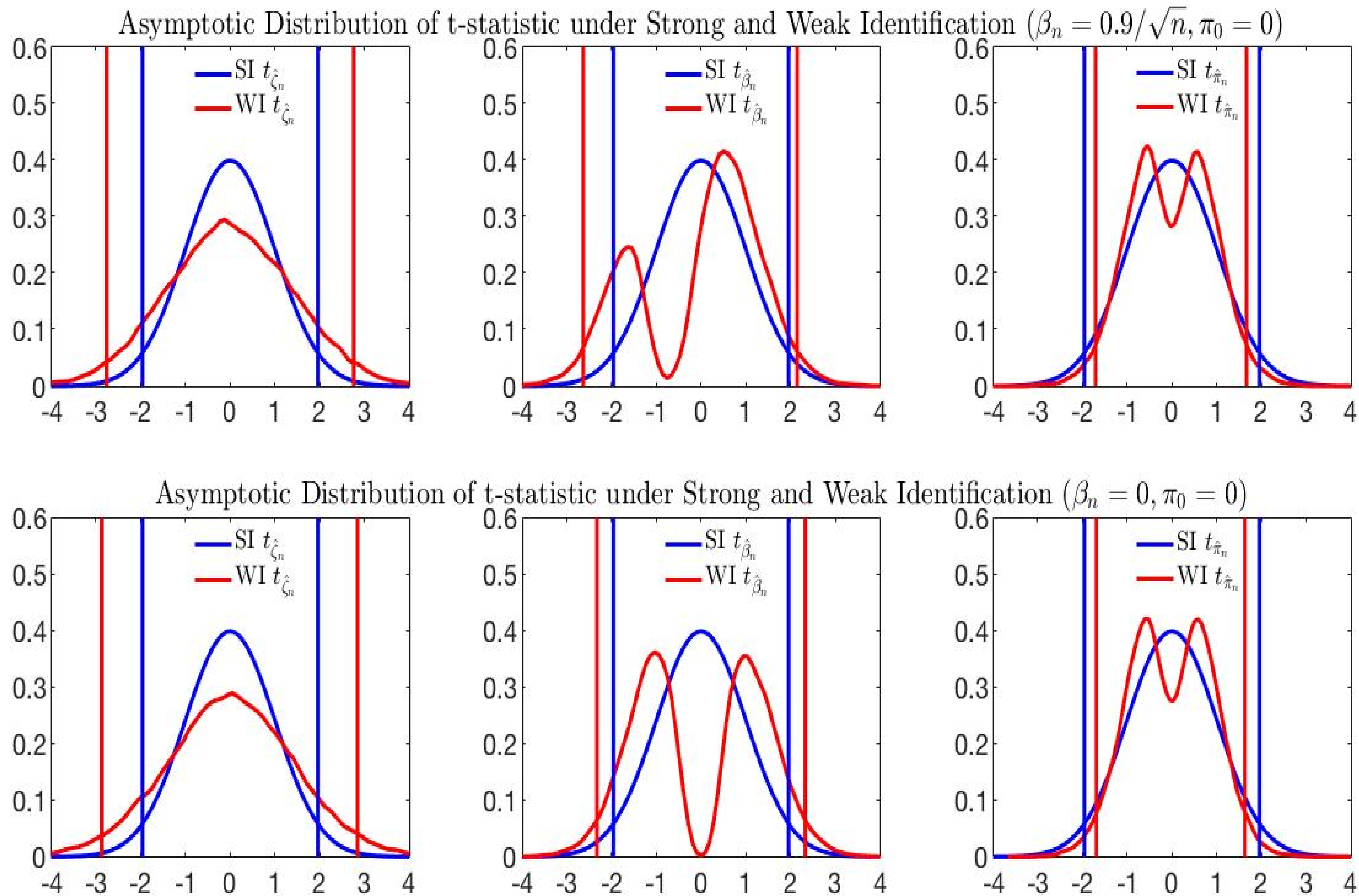

Figure A.1: Asymptotic distribution of the t-statistic for the Smoothing Exponential model with $t(4)$ errors. The blue line denotes the asymptotic distribution of the t-statistic under strong identification (standard normal), while the red line denotes the asymptotic distribution under weak identification (see Proposition 1.3.3). Vertical lines illustrate the 95% rejection critical values for each distribution. Each column illustrates the distribution for $\hat{\zeta}_n$, $\hat{\beta}_n$ and $\hat{\pi}_n$ respectively, while the rows illustrate the change in the parameter b. Sample size $n = 500$, simulations $S = 1000$, bootstrapped samples $M_n = 500$.

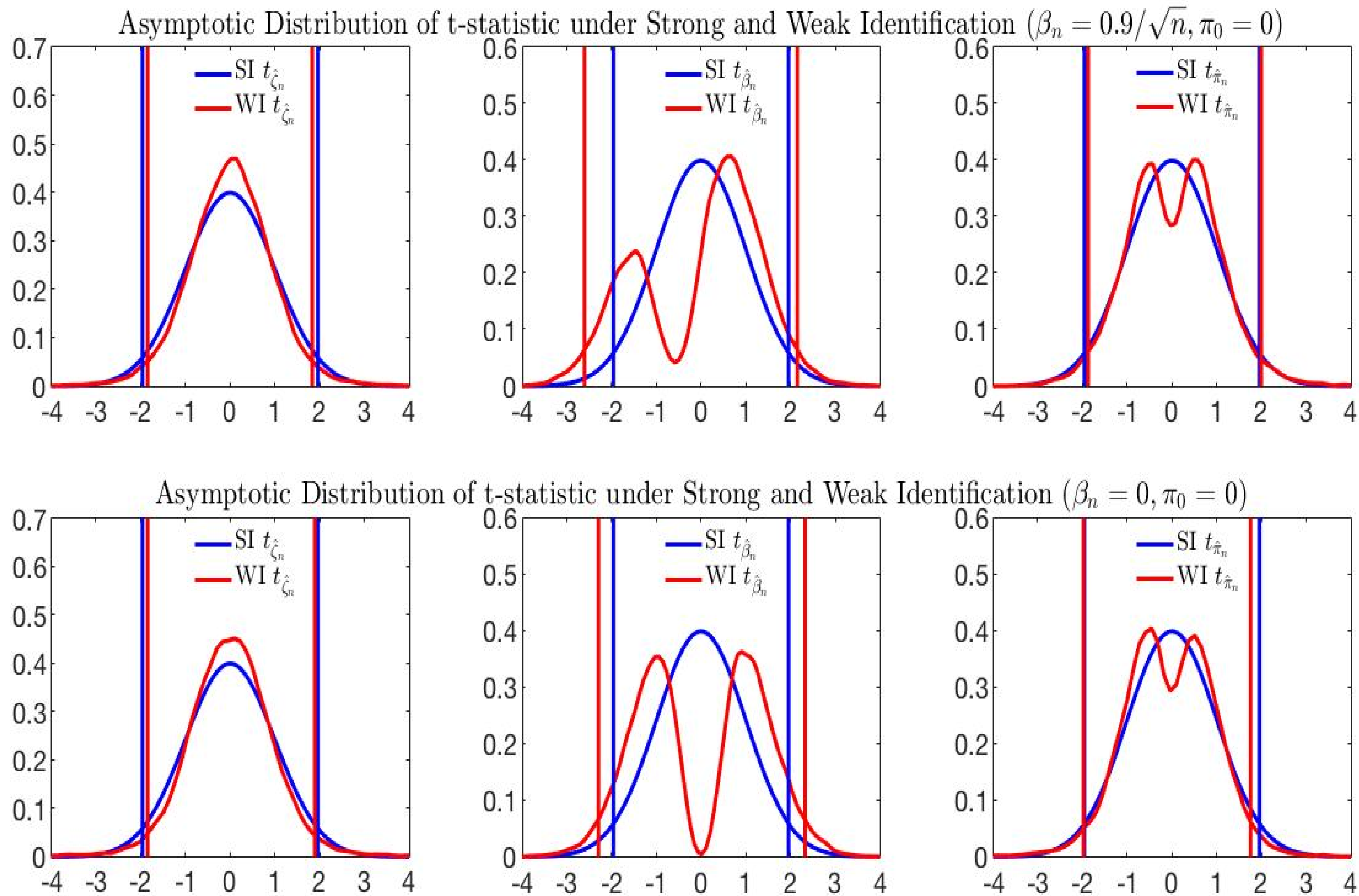

Figure A.2: Asymptotic distribution of the t-statistic for the Smoothing Exponential model with GARCH(0.1,0.6,0.3) errors. The blue line denotes the asymptotic distribution of the t-statistic under strong identification (standard normal), while the red line denotes the asymptotic distribution under weak identification (see Proposition 1.3.3). Vertical lines illustrate the 95% rejection critical values for each distribution. Each column illustrates the distribution for $\hat{\zeta}_n$, $\hat{\beta}_n$ and $\hat{\pi}_n$ respectively, while the rows illustrate the change in the parameter b. Sample size $n = 500$, simulations $S = 1000$, bootstrapped samples $M_n = 500$.

APPENDIX B

SUPPLEMENTAL APPENDIX OF "PARAMETRIC WILD BOOTSTRAP INFERENCE WITH WEAKLY IDENTIFIED PARAMETERS"

To show Proposition 2.4.1, we first need to show that the steps of the parametric bootstrap under strong and weak identification are valid. We prove the bootstrap steps using the following lemmas. Let $X_n(\pi) = o_{p,\pi}(1)$ be defined as $\sup_{\pi\in\Pi} ||X_n(\pi)|| = o_p(1)$, where $||\cdot||$ denotes the Euclidean norm and $|\cdot|$ denotes the absolute value. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes indexed by $\pi \in \Pi$ for some set Π. See Chapter 1 for more details. We begin with the parametric bootstrap under strong identification.

B.1 Proofs of the main results

Lemma B.1.1 *Step 1 SI leads to consistent residuals, that is*

(a) $||\epsilon_t(\hat{\theta}_n) - \epsilon_t(\theta_n)|| \xrightarrow{p} 0$

Proof of Lemma B.1.1 Let $\epsilon_t(\hat{\theta}_n)$ be the sample residuals,

$$||\epsilon_t(\hat{\theta}_n) - \epsilon_t(\theta_n)|| = ||y_t - \hat{\beta}_n' h(X_{2,t}, \hat{\pi}_n) - \hat{\zeta}_n' X_{1,t} - y_t - \beta_n' h(X_{2,t}, \pi_n) - \zeta_n' X_{1,t}]|| \tag{B.1}$$

$$= ||(\zeta_n - \hat{\zeta}_n)' X_{1,t} + \hat{\beta}_n' h(X_{2,t}, \hat{\pi}_n) - \beta_n' h(X_{2,t}, \pi_n)|| \tag{B.2}$$

By Theorem 3.2 of Andrews and Cheng (2012) or Chapter 1 in our context, we know that $\sqrt{n}(\hat{\zeta}_n - \zeta_n) = O_p(1)$ and $\sqrt{n}||\hat{\beta}_n||(\hat{\pi}_n - \pi_n) = O_p(1)$. Notice that as $h(\cdot)$ is a non-random bounded function, and by the mean value theorem. $h(X_{2,t}, \hat{\pi}_n) = h(X_{2,t}, \pi_n) + \frac{\partial}{\partial \pi} h(X_{2,t}, \bar{\pi})(\hat{\pi}_n - \pi_n)$. Then, from Equation (B.2)

$$||\epsilon_t(\hat{\theta}_n) - \epsilon_t(\theta_n)|| \tag{B.3}$$

$$= ||(\zeta_n - \hat{\zeta}_n)' X_{1,t} + \hat{\beta}_n' (h(X_{2,t}, \pi_n) + \frac{\partial}{\partial \pi} h(X_{2,t}, \bar{\pi})(\hat{\pi}_n - \pi_n)) - \beta_n' h(X_{2,t}, \pi_n)|| \tag{B.4}$$

$$= ||\underbrace{(\zeta_n - \hat{\zeta}_n)'}_{O_p(n^{-1/2})} X_{1,t} + \underbrace{(\hat{\beta}_n - \beta_n)'}_{O_p(n^{-1/2})} \underbrace{h(X_{2,t}, \pi_n)}_{O_p(1)} + \underbrace{\hat{\beta}_n' \frac{\partial}{\partial \pi} h(X_{2,t}, \bar{\pi})}_{O_p(n^{-1/2})} \underbrace{(\hat{\pi}_n - \pi_n)}_{O_p(1)}|| = o_p(1) \blacksquare \tag{B.5}$$

This proves Step 1 SI. In Step 2 SI, we argue that the Wild bootstrap is a valid procedure to simulate random draws. For a formal proof we refer to Liu et al. (1988) for heteroscedastic independent data, and Shao (2010) for dependent data. Now we consider Step 3 SI.

Lemma B.1.2 *The bootstrapped samples $y_t^{m,s}$ have the same distribution as the original sample.*

$$\sup_{z\in\mathbb{R}^{d_\theta}}\left|\mathbb{P}_m\begin{pmatrix} n^{-1/2}\sum_{t=1}^n y_t^{m,s}\leq z_1 \\ n^{-1/2}\sum_{t=1}^n X_t^{m,s}\leq z_2\end{pmatrix} - \mathbb{P}_{\theta_n}\begin{pmatrix} n^{-1/2}\sum_{t=1}^n y_t(\theta_n)\leq z_1 \\ n^{-1/2}\sum_{t=1}^n X_t(\theta_n)\leq z_2\end{pmatrix}\right| \xrightarrow{p} 0 \tag{B.6}$$

Proof of Lemma B.1.2 In the case of X_t, we have that $X_t^m = X_t$ if X_t is not determined recursively (i.e. lags of dependent variable). If it is determined recursively, we only need to show the results hold for y_t^m with respect to y_t.

$$\sup_{z\in\mathbb{R}}|\mathbb{P}_m(y_t^{m,s}\leq z)-\mathbb{P}_{\theta_n}(y_t(\theta_n)\leq z)| \tag{B.7}$$

$$\leq \sup_{z\in\mathbb{R}}|\mathbb{P}_m(n^{-1/2}\sum_{t=1}^n y_t^{m,s}\leq z)-\mathbb{P}_{\theta_0}(n^{-1/2}\sum_{t=1}^n y_t(\theta_0)\leq z)| \tag{B.8}$$

$$+ \sup_{z\in\mathbb{R}}|\mathbb{P}_{\theta_0}(n^{-1/2}\sum_{t=1}^n y_t(\theta_0)\leq z)-\mathbb{P}_{\theta_n}(n^{-1/2}\sum_{t=1}^n y_t(\theta_n)\leq z)| \tag{B.9}$$

The second term of the inequality is $o(1)$ as the non-random sequence $\theta_n \to \theta_0$, and the continuity of measures property theorem. Therefore the distribution functions of $y_t(\theta_n)$ converges to the distribution function of $y_t(\theta_0)$. Now we consider the first term of the inequality.

$$y_t^{m,s} = \hat{\zeta}_n^{m,s\prime}X_{1,t}+\hat{\beta}_n^{m,s\prime}h(X_{2,t},\hat{\pi}_n^{m,s})+z_t^m\epsilon_t(\hat{\theta}_n) \tag{B.10}$$

$$= (\hat{\zeta}_n^{m,s}-\zeta_n)'X_{1,t}+\hat{\beta}_n^{m,s\prime}h(X_{2,t},\hat{\pi}_n^{m,s})-\beta_n'h(X_{2,t},\pi_n) \tag{B.11}$$

$$+\zeta_n'X_{1,t}+\beta_n'h(X_{2,t},\pi_n)+z_t^m\epsilon_t(\hat{\theta}_n) \tag{B.12}$$

$$= \zeta_n'X_{1,t}+\beta_n'h(X_{2,t},\pi_n)+z_t^m\epsilon_t(\hat{\theta}_n)+o_p(1) \tag{B.13}$$

The last equality follows by the same argument as in Equation (B.5). Notice that

$$\mathbb{E}_{\theta_n}(z_t^m \epsilon_t(\hat{\theta}_n)) = \mathbb{E}_{\theta_n}(\mathbb{E}(z_t^m|X_t)\epsilon_t(\hat{\theta}_n)) = 0 \tag{B.14}$$

$$Var_{\theta_n}(z_t^m \epsilon_t(\hat{\theta}_n)) = \mathbb{E}(\mathbb{E}((z_t^m)^2|X_t)\epsilon_t^2(\hat{\theta}_n)) = \Gamma_0(\theta_n) + o_p(1) \tag{B.15}$$

$$Cov_{\theta_n}(z_t^m \epsilon_{t-j}(\hat{\theta}_n), z_t^m \epsilon_{t-j}(\hat{\theta}_n)) = \Gamma_j(\theta_n) + o_p(1) \tag{B.16}$$

The variance equality follows from Lemma B.1.1, and the covariance equality follows similarly from the construction of the Dependent Wild bootstrap (see Shao (2010)). The Central Limit Theorem of Wooldridge and White (1988) follows as ϵ_t has finite $4+\iota$ moments and is mixing size $-r/(r-2)$ for $r > 2$. The central limit theorem follows,

$$n^{-1/2}\sum_{t=1}^{n} z_t^m \epsilon_t(\hat{\theta}_n) \xrightarrow{d} N(0, V_\epsilon(\theta_0)) \tag{B.17}$$

where $V_\epsilon(\theta_0) = \Gamma_0(\theta_0) + 2\sum_{j=1}^{\infty}\Gamma_j(\theta_0)$ denotes the variance covariance matrix of ϵ_t. In consequence,

$$n^{-1/2}\sum_{t=1}^{n} y_t^{m,s} = n^{-1/2}\sum_{t=1}^{n}(\zeta_n' X_{1,t} + \beta_n' h(X_{2,t}, \pi_n) + z_t^m \epsilon_t(\hat{\theta}_n)) \tag{B.18}$$

The first two terms of $y_t^{m,s}$ are equivalent to the terms of $y_t(\theta_n) = \zeta_n' X_{1,t} + \beta_n' h(X_{2,t}, \pi_n) + \epsilon_t$, while the last term converges in distribution to ϵ_t. Then $y_t^{m,s}$ and $y_t(\theta_n)$ have the same distribution in the limit, (conditional on the sample X_t, see Giné and Zinn (1990)), i.e.

$$\sup_{z\in\mathbb{R}}|\mathbb{P}_m(n^{-1/2}\sum_{t=1}^{n} y_t^{m,s} \leq z) - \mathbb{P}_{\theta_0}(n^{-1/2}\sum_{t=1}^{n} y_t(\theta_0) \leq z)| \xrightarrow{p} 0 \tag{B.19}$$

To prove joint convergence use the Cramer Wold device. Let $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}^d$. Then

$$\lambda_1 y_t^m + \lambda_2' X_t^m \xrightarrow{d} \lambda_1 y_t(\theta_n) + \lambda_2' X_t(\theta_n) \tag{B.20}$$

by the Continuous Mapping Theorem. ■

The previous lemmas are used to show the next proposition.

Proposition 2.4.1 *Suppose that Assumptions (J) to (N) hold and suppose that the true data generating process is strongly identified. Let $M_n \to \infty$ as $n \to \infty$. The estimators obtained following Step 1 SI to Step 4 SI have the following distribution,*

$$\sup_{z\in\mathbb{R}^{d_\theta}} \left| \mathbb{P}_m(\sqrt{M_n}B(\hat{\beta}_n^{m,s})(\hat{\theta}_n^{m,s} - \hat{\theta}_n) \leq z) - \mathbb{P}_{\theta_n}(\sqrt{n}B(\hat{\beta}_n)(\hat{\theta}_n - \theta_n) \leq z) \right| \xrightarrow{p} 0 \tag{2.19}$$

Proof of Proposition 2.4.1 By Lemma B.1.2, $\sup_{z\in\mathbb{R}^{d+1}} ||\mathbb{P}_m(W_t^{m,s} \leq z) - \mathbb{P}_{\theta_n}(W_t(\theta_n) \leq z)|| = o_p(1)$. In other words, the bootstrapped samples and original samples have the same distribution in the limit. As $Q_n^m(W_t^{m,s}, \theta) = n^{-1}\sum_{t=1}^n \epsilon_t(W_t^{m,s}, \theta)^2 = n^{-1}\sum_{t=1}^n \epsilon_t(W_t, \theta)^2 + o_p(1)$ then by the classic results of extremum estimators Newey (1991); Newey and McFadden (1994), the bootstrapped estimators $\hat{\theta}_n^{m,s}$ converge to the limit distribution which is normal. The distribution result follows Theorem 3.2 of Andrews and Cheng (2012). We must show that the conditions of Andrews and Cheng (2012) are satisfied. This is proven in Proposition 1.3.1 and Proposition 1.3.2 of Chapter 1. ■

Lemma B.1.3 *To prove Step 1 WI, we must show,*

(a) $\sup_{\pi\in\Pi} ||\hat{\psi}_n(\pi) - \psi_n|| \xrightarrow{p} 0$

(b) $\sup_{\pi\in\Pi} |\epsilon_t(\hat{\psi}_n(\pi), \pi) - \epsilon_t(\psi_n)| \xrightarrow{p} 0$

(c) $||\epsilon_t(\hat{\psi}_{0,n}) - \epsilon_t(\psi_n)|| \xrightarrow{p} 0$

Proof of Lemma B.1.3 (a) This result follows from Lemma 3.1 of Andrews and Cheng (2012). We show that the conditions of this result are satisfied in Proposition 1.3.1 and Proposition 1.3.2 of Chapter 1. Specifically, we need to show that the objective function $Q_n(\psi, \pi)$ satisfies a pointwise law of large numbers, is stochastic equicontinuous, and that the limit objective function $Q(\psi, \pi; \theta_0)$ is equicontinuous.

(b) Consider,

$$\sup_{\pi\in\Pi} ||\epsilon_t(\hat{\psi}_n(\pi), \pi) - \epsilon_t(\psi_n)|| \tag{B.21}$$

$$= \sup_{\pi\in\Pi} ||(\hat{\zeta}_n(\pi) - \zeta_n)X_{1,t} + \hat{\beta}_n(\pi)h(X_{2,t},\pi) - \beta_n h(X_{2,t},\pi_n)|| \tag{B.22}$$

$$\le \sup_{\pi\in\Pi} ||(\hat{\zeta}_n(\pi) - \zeta_n)X_{1,t}|| + \sup_{\pi\in\Pi} ||\hat{\beta}_n(\pi)h(X_{2,t},\pi) - \beta_n h(X_{2,t},\pi_n)|| \tag{B.23}$$

$$\le \sup_{\pi\in\Pi} ||(\hat{\zeta}_n(\pi) - \zeta_n)|| \sup_{\pi\in\Pi} ||X_{1,t}|| + \sup_{\pi\in\Pi} ||\hat{\beta}_n(\pi) - \beta_n|| \sup_{\pi\in\Pi} ||h(X_{2,t},\pi)|| \tag{B.24}$$

$$= O_{p,\pi}(n^{-1/2})O_{p,\pi}(1) + O_{p,\pi}(n^{-1/2})O_{p,\pi}(1) = O_{p,\pi}(n^{-1/2}) \tag{B.25}$$

(c) Similarly,

$$\sup_{\pi\in\Pi} ||\epsilon_t(\hat{\psi}_{0,n}) - \epsilon_t(\psi_n)|| = ||y_t - \hat{\zeta}_n X_{1,t} - y_t + \zeta_n X_{1,t} + \beta_n h(X_{2,t},\pi_n)|| \tag{B.26}$$

$$= \sup_{\pi\in\Pi} ||(\hat{\zeta}_n(\pi) - \zeta_n)|| \sup_{\pi\in\Pi} ||X_{1,t}|| + \sup_{\pi\in\Pi} ||\beta_n|| \sup_{\pi\in\Pi} ||h(X_{2,t},\pi)|| \tag{B.27}$$

$$= O_p(n^{-1/2})O_p(1) + O_p(n^{-1/2})O_p(1) = O_p(n^{-1/2}) \quad \blacksquare \tag{B.28}$$

Lemma B.1.4 *To prove Step 2 WI, we must show,*

(a) $\sup_{z\in\mathbb{R}} |\mathbb{P}_m(n^{-1/2}\sum_{t=1}^n(\epsilon_t^m(\hat{\psi}_{0,n}) - \mathbb{E}_m(\epsilon_t^m(\hat{\psi}_{0,n}))) \le z) - \mathbb{P}_{\theta_n}(n^{-1/2}\sum_{t=1}^n \epsilon_t(\theta_n) \le z)| \xrightarrow{p} 0$

Proof of Lemma B.1.4 By Lemma B.1.3, $\epsilon_t(\hat{\psi}_{0,n}) = \epsilon_t(\psi_{0,n}) + o_p(1)$ (there is no influence from the weakly identified parameters π). Therefore, we will use $\psi_{0,n}$ in place of $\hat{\psi}_{0,n}$ for the arguments that follow. We notice that the residuals centered at the point of lack of identification are not necessarily mean zero by construction.

$$\sup_{z\in\mathbb{R}} |\mathbb{P}_m(n^{-1/2}\sum_{t=1}^n(\epsilon_t^m(\hat{\psi}_{0,n}) - \mathbb{E}_m(\epsilon_t^m(\hat{\psi}_{0,n}))) \le z) - \mathbb{P}_{\theta_n}(n^{-1/2}\sum_{t=1}^n \epsilon_t(\theta_n) \le z)| \tag{B.29}$$

$$\le \sup_{z\in\mathbb{R}} |\mathbb{P}_m(n^{-1/2}\sum_{t=1}^n(\epsilon_t^m(\hat{\psi}_{0,n}) - \mathbb{E}_m(\epsilon_t^m(\hat{\psi}_{0,n}))) \le z) - \Phi_\epsilon(z)| \tag{B.30}$$

$$+ \sup_{z\in\mathbb{R}} |\mathbb{P}_{\theta_n}(n^{-1/2}\sum_{t=1}^n \epsilon_t(\theta_n) \le z) - \Phi_\epsilon(z)| \tag{B.31}$$

where $\Phi_\epsilon(z)$ is the limit distribution of the scaled average of $\epsilon_t(\theta_0)$, that is, a mean zero normal distribution with variance $V_\epsilon(\theta_0) = \Gamma_0(\theta_0) + 2\sum_{j=1}^\infty \Gamma_j(\theta_0)$. Equation (B.31) is $o_p(1)$ as the sequence of non-random numbers $\theta_n \to \theta_0$, continuity of measures, and the residuals do not depend on π. More specifically, we show it using a central limit theorem argument. Notice that $\mathbb{E}_{\theta_n}(\epsilon_t(\theta_n)) = 0$ and $Var_{\theta_n}(\epsilon_t(\theta_n)) = \Gamma_0(\theta_n)$. The limit variance is bounded away from zero, i.e. for some $\delta > 0$

$$\lim_{n\to\infty} n^{-1} Var_{\theta_n}(\sum_{t=1}^{n} \epsilon_t(\theta_n)) = \lim_{n\to\infty}[n^{-1}\sum_{t=1}^{n} Var_{\theta_n}(\epsilon_t(\theta_n)) + 2\ n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} Cov_{\theta_n}(\epsilon_t(\theta_n), \epsilon_s(\theta_n))] \tag{B.32}$$

$$= \lim_{n\to\infty}[\Gamma_0(\theta_n) + 2\sum_{j=1}^{n}(\frac{n-j}{n})\Gamma_j(\theta_n)] = \Gamma_0(\theta_0) + 2\sum_{j=1}^{n}\Gamma_j(\theta_0) > \delta > 0 \tag{B.33}$$

As the limit variance is bounded away from zero and $\epsilon_t(\theta_n)$ is L_p bounded for $p = 4 + \iota$, then the Wooldridge and White (1988) central limit theorem applies,

$$n^{-1/2}\sum_{t=1}^{n}\epsilon_t(\theta_n) \xrightarrow{d} N(0, V_\epsilon(\theta_0)) \tag{B.34}$$

The central limit theorem result proves Equation (B.31) is $o_p(1)$. To prove Equation (B.30) is $o_p(1)$, consider

$$n^{-1/2}\sum_{t=1}^{n}[\epsilon_t(\psi_{0,n}) - \mathbb{E}_m(\epsilon_t(\psi_{0,n}))] = n^{-1/2}\sum_{t=1}^{n}[z_t^m\epsilon_t(\psi_{0,n}) - \mathbb{E}_m(z_t^m\epsilon_t(\psi_{0,n}))] \tag{B.35}$$

$$= n^{-1/2}\sum_{t=1}^{n}[z_t^m\epsilon_t(\theta_n) + z_t^m\beta_n h(X_{2,t},\pi_n) - \mathbb{E}_m(z_t^m\epsilon_t(\theta_n) + z_t^m\beta_n h(X_{2,t},\pi_n))] \tag{B.36}$$

$$= \underbrace{n^{-1/2}\sum_{t=1}^{n}[z_t^m\epsilon_t(\theta_n) - \mathbb{E}_m(z_t^m\epsilon_t(\theta_n))]}_{(I)} - \underbrace{b\ n^{-1}\sum_{t=1}^{n}[z_t^m h(X_{2,t},\pi_n) - \mathbb{E}_m(z_t^m h(X_{2,t},\pi_n))]}_{(I)} \tag{B.37}$$

First we show (II) follows a law of large numbers. Let $\mathbb{E}_m$ be the expectation conditional on W_t (see Giné and Zinn (1990)),

$$\mathbb{E}_m(z_t^m h(X_{2,t},\pi_n)) = \mathbb{E}_{\theta_n}(z_t^m|W_t)h(X_{2,t},\pi_n)) = 0 \tag{B.38}$$

The process $z_t^m h(X_{2,t}, \pi_n)$ is a sequence of stationary and mixing random variables, and

$$\mathbb{E}_{\theta_n}(||z_t^m h(X_{2,t}, \pi_n)||^{1+\iota}) < (\mathbb{E}_{\theta_n}(||z_t^m||^{2+\iota}))^{2+\iota}(\mathbb{E}_{\theta_n}(||h(X_{2,t}, \pi_n)||^{2+\iota}))^{2+\iota} \leq \infty \tag{B.39}$$

which follows from Assumption K. By McLeish et al. (1975),

$$||n^{-1} \sum_{t=1}^{n} (z_t^m h(X_{2,t}, \pi_n) - \mathbb{E}_m(z_t^m h(X_{2,t}, \pi_n)))|| \xrightarrow{p} 0 \tag{B.40}$$

This implies $\mathbb{E}_m(z_t^m \epsilon_t(\psi_{0,n})) = \mathbb{E}_{\theta_n}(z_t^m \epsilon_t(\psi_{0,n})) + o_p(1)$. Now we consider (I) of Equation (B.37), we show that $\{z_t^m \epsilon_t(\theta_n)\}$ satisfies a central limit theorem as $n \to \infty$. If $z_t \overset{d}{\sim} iid$ as in Liu et al. (1988), the expectation and variances of the bootstrapped samples are: $\mathbb{E}_m(z_t^m \epsilon_t(\theta_n)) = \mathbb{E}_m(z_t^m | W_t)\epsilon_t(\theta_n))$ and $Var_m(z_t^m \epsilon_t(\theta_n)) = \mathbb{E}_m(z_t^m \epsilon_t(\theta_n))^2 = \Gamma_0(\theta_n)$. Moreover, the sequence is L_p bounded for $p = 2 + \iota$ by Assumption K and stationary and mixing. If z_t are dependent, to prove the central limit theorem we must also prove the covariances are converging, which can be shown using the large block, small block argument. We refer to Theorem 3.1 of Shao (2010), particularly Equation (A.3), for a proof. Having shown the conditions of the central limit theorem,

$$n^{-1/2} \sum_{t=1}^{n} [z_t^m \epsilon_t(\theta_n) - \mathbb{E}_m(z_t^m \epsilon_t(\theta_n))] \xrightarrow{d} N(0, V_\epsilon(\theta_0)) \tag{B.41}$$

which implies that (I) is $o_p(1)$. ■

Before we show the next lemma, we need to introduce a few concepts. Let $\{Y_n^m(\hat{\psi}_n(\pi), \pi) : \pi \in \Pi\}$ be a stochastic process of the bootstrapped sample, that is $Y_n^m(\pi) = n^{-1/2} \sum_{t=1}^n y_t^m(\hat{\psi}_n(\pi), \pi)$. Let $\{Y(\psi_0, \pi) : \pi \in \Pi\}$ be the limit Gaussian process of $y_t(\psi_0, \pi)$ as a function of π. That is, a stochastic process with the following mean and variance-covariance kernel,

$$\mathbb{E}_{\theta_0}(Y(\psi_0, \pi)) = \mathbb{E}_{\theta_0}(y_t(\psi_0, \pi)) \tag{B.42}$$

$$Var_{\theta_0}(Y(\psi_0, \pi)) = \mathbb{E}(y_t(\psi_0, \pi) y_t(\psi_0, \pi)') \tag{B.43}$$

Similarly let $\{W_n^m(\hat{\psi}_n(\pi),\pi):\pi\in\Pi\}$ and $\{W(\psi_0,\pi):\pi\in\Pi\}$ be defined analogously with $W(\psi,\pi)=\{(Y(\psi,\pi),X(\psi,\pi):\pi\in\Pi\}$ and $\{X(\psi,\pi):\pi\in\Pi\}$ depends on $\pi\in\Pi$ only if X_t includes lags of y_t. In the next Lemma we show the bootstrap sample converges weakly to the limit stochastic process.

Lemma B.1.5 *Let $\{Y_n^m(\hat{\psi}_n(\pi),\pi):\pi\in\Pi\}$ and $\{Y(\psi_0,\pi):\pi\in\Pi\}$ be the stochastic processes defined in the previous paragraph. Letting $M_n\to\infty$ as $n\to\infty$,*

(a) $\{Y_n^m(\hat{\psi}_n(\pi),\pi):\pi\in\Pi\}\Rightarrow\{Y(\psi_0,\pi):\pi\in\Pi\}$

(b) $\{W_n^m(\hat{\psi}_n(\pi),\pi):\pi\in\Pi\}\Rightarrow\{W(\psi_0,\pi):\pi\in\Pi\}$

Proof of Lemma B.1.5 As Π is a compact set, to show the weak convergence result to the Gaussian process $Y(\psi_0,\pi)$ and $W(\psi_0,\pi)$, we must show convergence in finite dimensional distributions and stochastic equicontinuity Dudley (1978); Pollard (1990).

(a) First we notice that, $\{n^{-1/2}\sum_{t=1}^n y_t(\psi_n,\pi):\pi\in\Pi\}\Rightarrow\{Y(\psi_0,\pi):\pi\in\Pi\}$, because $Y(\psi_0,\pi)$ is the limit Gaussian process for fixed $\pi\in\Pi$. This follows as $\psi_n\to\psi_0$ and therefore the sequence of $y_t(\psi_n,\pi)$ follows a central limit theorem for each $\pi\in\Pi$ and Stochastic Equicontinuity. This proof is verbatim to the proof below replacing $y_t^m(\psi_n,\pi)$ with $y_t(\psi_n,\pi)$.

Now we consider, $y_t^m(\hat{\psi}_n(\pi),\pi)$. By Lemma B.1.3, we can replace $\hat{\psi}_n(\pi)$ by ψ_n, letting the stochastic process take the form $y_t^m(\hat{\psi}_n(\pi),\pi)=y_t^m(\psi_n,\pi)+o_{p,\pi}(1)$ for each $\pi\in\Pi$. Similar to the previous lemma, we assume $z_t\overset{d}{\sim}iid$. We refer to Theorem 3.1 of Shao (2010), particularly Equation (A.3), for a proof when z_t is dependent. To show convergence in finite dimensional distributions, notice that,

$$\mathbb{E}_m(y_t^m(\psi_n,\pi))=\mathbb{E}_{\theta_n}(y_t^m(\psi_n,\pi)|X_t) \tag{B.44}$$

$$=\zeta_n'X_{1,t}+\beta_n'h(X_{2,t},\pi)+\mathbb{E}_{\theta_n}(z_t^m(\epsilon_t(\psi_{0,n})-n^{-1}\sum_{t=1}^n\epsilon_t(\psi_{0,n}))) \tag{B.45}$$

$$=\zeta_n'X_{1,t}+\beta_n'h(X_{2,t},\pi) \tag{B.46}$$

$$Var_{\theta_n}(z_t^m\epsilon_t(\psi_{0,n}))=\mathbb{E}_{\theta_n}((z_t^m)^2\epsilon_t^2(\psi_{0,n}))=\mathbb{E}_{\theta_n}((z_t^m)^2(y_t-\zeta_n'X_{1,t})^2) \tag{B.47}$$

$$= \mathbb{E}_{\theta_n}((z_t^m)^2(y_t - \zeta_n' X_{1,t} - \beta_n' h(X_{2,t}, \pi_n) + \beta_n' h(X_{2,t}, \pi_n))^2) \tag{B.48}$$

$$= \mathbb{E}_{\theta_n}((z_t^m)^2(\epsilon_t^2(\theta_n) + (\beta_n' h(X_{2,t}, \pi_n)^2 + \beta_n' h(X_{2,t}, \pi_n)\epsilon_t(\theta_n))) \tag{B.49}$$

$$= \mathbb{E}_{\theta_n}(\mathbb{E}_{\theta_n}((z_t^m)^2|W_t)\epsilon_t^2(\theta_n)) + O_p(n^{-1/2}) = \Gamma_0(\theta_n) \tag{B.50}$$

$$Var_{\theta_n}(y_t^m|X_t) = Var_{\theta_n}(y_t^m(\psi_n, \pi)|X_t) = \Gamma_0(\theta_n) \tag{B.51}$$

$$Cov_{\theta_n}(y_t^m, y_{t-j}^m|X_t) = Cov_{\theta_n}(y_t^m(\psi_n, \pi), y_{t-j}^m(\psi_n, \pi)|X_t) = \Gamma_j(\theta_n) \tag{B.52}$$

This implies $\mathbb{E}_{\theta_n}(y_t^m(\psi_n, \pi)) = \mathbb{E}_{\theta_n}(y_t(\psi_n, \pi))$ and $Var_{\theta_n}(y_t^m(\psi_n, \pi)) = Var_{\theta_n}(y_t(\psi_n, \pi))$. By Assumption J and Assumption K, for each $\pi \in \Pi$, $y_t^m(\psi_n, \pi)$ is a stationary and mixing process, and for some $\delta > 0$

$$\lim_{n\to\infty} n^{-1} Var_{\theta_n}(\sum_{t=1}^{n} y_t(\psi_n, \pi)) \tag{B.53}$$

$$= \lim_{n\to\infty}[n^{-1}\sum_{t=1}^{n} Var_{\theta_n}(y_t(\psi_n, \pi)) + 2\ n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} Cov_{\theta_n}(y_t(\psi_n, \pi), y_{t-j}(\psi_n, \pi))] \tag{B.54}$$

$$= \lim_{n\to\infty}[\Gamma_{y,0}(\theta_n) + 2\sum_{j=1}^{n}(\frac{n-j}{n})\Gamma_{y,j}(\theta_n)] = \Gamma_{y,0}(\theta_0) + 2\sum_{j=1}^{n}\Gamma_{y,j}(\theta_0) > \delta > 0 \tag{B.55}$$

Therefore the limit variance is positive definite, and we can apply Wooldridge and White (1988) central limit theorem. This shows convergence in distribution for fixed $\pi \in \Pi$,

$$n^{-1/2}\sum_{t=1}^{n} y_t^m(\psi_n, \pi) \xrightarrow{d} N(0, V_y(\theta_0)) \tag{B.56}$$

where $V_y(\theta_0) = \Gamma_{y,0}(\theta_0) + 2\sum_{j=1}^{n}\Gamma_{y,j}(\theta_0)$.

Now we show Stochastic Equicontinuity, which follows by the Lipschitz condition in Assumption K and the fact that $y_t^m(\psi_n, \pi)$ is uniformly continuous with respect to π in a compact set. Let $\pi, \tilde{\pi} \in \Pi$ and $\delta, \eta > 0$,

$$\mathbb{P}_m(\sup_{||\pi-\tilde{\pi}||<\delta} ||y_t^m(\psi_n, \pi) - y_t^m(\psi_n, \tilde{\pi})|| > \eta) \tag{B.57}$$

$$\leq \mathbb{P}_{\theta_n}(\sup_{||\pi-\tilde{\pi}||<\delta} ||\beta_n' h(X_{2,t}, \pi) - \beta_n' h(X_{2,t}, \tilde{\pi})|| > \eta) \tag{B.58}$$

$$\leq \frac{1}{\eta}\mathbb{E}_{\theta_n}(\sup_{||\pi-\tilde{\pi}||<\delta} C_n||\pi-\tilde{\pi}||) \tag{B.59}$$

$$\leq \frac{1}{\eta}\delta\mathbb{E}_{\theta_n}(C_n) \tag{B.60}$$

The second line follows from the fact that the second expression does not depend on the bootstrapped sample z_t^m, it only depends on the randomness of X_t. In consequence we can substitute the bootstrapped probability measure with the original probability measure. The third line follows by Markov's inequality and $C_n = O_p(1)$ is the Lipschitz constant. Moreover as $\mathbb{E}_{\theta_n}(C_n) = O(1)$ by Assumption K. Let $\epsilon, \eta > 0$ and $\delta = \epsilon\eta/\mathbb{E}_{\theta_n}(C_n)$, then

$$\mathbb{P}_{\theta_n}(\sup_{||\pi-\tilde{\pi}||<\delta} ||y_t^m(\psi_n,\pi) - y_t^m(\psi_n,\tilde{\pi})|| > \eta) < \epsilon \tag{B.61}$$

which proves stochastic equicontinuity.

(b) If X_t does not include lags of y_t, the result follows directly from (a) and the fact that $X_t^m = X_t$. If X_t includes lags of y_t, then the proof is verbatim to (a) replacing $y_t^m(\pi)$ by $y_{t-1}^m(\pi)$. Joint convergence follows using the Cramer-Wold device. ■

For the rest of the proofs, we assume that X_t^m is not autoregressive and therefore does not depend on π for any of the bootstrapped samples. The proof for $X_t(\pi)$ follows the same argument as in Lemma B.1.5(b).

Lemma B.1.6 *In addition to Lemma B.1.5, Step 3 WI requires,*

(a) $\sup_{\pi\in\Pi} ||\hat{\psi}_n^m(\pi,\pi_n) - \psi_n|| \xrightarrow{p} 0$

(b) Fore each $\pi_k \in \tilde{\Pi}$, $\sup_{\pi\in\Pi} ||\hat{\psi}_n^m(\pi,\pi_k) - \psi_n|| \xrightarrow{p} 0$

Proof of Lemma B.1.6 Notice that we assume that π_n is known, therefore $\hat{\psi}_n^m(\pi,\pi_n)$ is estimated using bootstrapped samples that are generated using consistent estimators.

(a) Consider the above term,

$$\sup_{\pi\in\Pi} ||\hat{\psi}_n^m(\pi,\pi_n) - \psi_n|| \leq \underbrace{\sup_{\pi\in\Pi} ||\hat{\psi}_n^m(\pi,\pi_n) - \hat{\psi}_n(\pi)||}_{(I)} + \underbrace{\sup_{\pi\in\Pi} ||\hat{\psi}_n(\pi) - \psi_n||}_{(II)} \tag{B.62}$$

The term (II) in Equation (B.62) is $o_{p,\pi}(1)$ by Lemma B.1.3(a). Consider term (I). We notice that $\hat{\psi}_n^m(\pi)$ is constructed analogously to $\hat{\psi}_n(\pi)$, using the bootstrapped data W_t^m instead of W_t. By Lemma B.1.5, for each fixed $\pi \in \Pi$, $W_t^m(\pi) \overset{d}{\sim} W_t(\pi)$. Then,

$$Q_n(W_t, \theta) = n^{-1} \sum_{t=1}^{n} (y_t - \zeta' X_{1,t} - \beta' h(X_{2,t}^m, \pi))^2 \tag{B.63}$$

$$\overset{d}{\sim} n^{-1} \sum_{t=1}^{n} (y_t^m(\pi_n) - \zeta' X_{1,t}^m - \beta' h(X_{2,t}^m, \pi))^2 = Q_n(W_t^m, \theta) \tag{B.64}$$

Where $y_t^m(\pi_n)$ denotes that the bootstrapped samples are generated setting $\pi = \pi_n$. In particular, as $Q_n(W_t, \theta)$ is converging to a real value $Q_0(W_t, \theta) = \mathbb{E}_{\theta_0}(y_t - \zeta' X_{1,t} - \beta' h(X_{2,t}, \pi))^2$, then both objective functions are converging in probability to the same value using the same arguments of Newey and McFadden (1994). By the mapping theorem Van Der Vaart and Wellner (1996), $\min_{\psi \in \Psi(\pi)} Q_n(W_t, \theta) \xrightarrow{p} \min_{\psi \in \Psi(\pi)} \mathbb{E}_{\theta_0}(Q_0(W_t, \theta))$ and therefore $\forall \pi \in \Pi$,

$$\sup_{\pi \in \Pi} || \min_{\psi \in \Psi(\pi)} Q_n(W_t^m(\pi_n), \theta) - \min_{\psi \in \Psi(\pi)} Q_n(W_t, \theta)|| = \sup_{\pi \in \Pi} ||\hat{\psi}_n^m(\pi, \pi_n) - \hat{\psi}_n(\pi)|| \xrightarrow{p} 0 \quad \blacksquare \tag{B.65}$$

(b) The bootstrapped samples generated in Step 3 WI are generated without the knowledge of π_n along the grid. Notice that,

$$y_t^m(\hat{\psi}_n(\pi), \pi_k) = \hat{\zeta}_n(\pi)' + \hat{\beta}_n(\pi)' h(X_{2,t}, \pi_k) + \epsilon_t^m(\hat{\psi}_{0,n}) \tag{B.66}$$

$$= \zeta_n' X_{1,t} + \beta_n' h(X_{2,t}, \pi_n) + \epsilon_t^m(\hat{\psi}_{0,n}) + (\hat{\zeta}_n(\pi) - \zeta_n)' X_{1,t} \tag{B.67}$$

$$+ (\hat{\beta}_n(\pi) - \beta_n)' h(X_{2,t}, \pi_k) + \beta_n' h(X_{2,t}, \pi_k) - \beta_n' h(X_{2,t}, \pi_n) \tag{B.68}$$

$$= y_t(\theta_n) + 1/\sqrt{n}[\sqrt{n}(\hat{\zeta}_n(\pi) - \zeta_n)' X_{1,t} + \sqrt{n}(\hat{\beta}_n(\pi) - \beta_n)' h(X_{2,t}, \pi_k) \tag{B.69}$$

$$+ \sqrt{n}\beta_n' h(X_{2,t}, \pi_k) - \sqrt{n}\beta_n' h(X_{2,t}, \pi_n)] \tag{B.70}$$

$$= y_t^m(\theta_n) + O_{p,\pi}(n^{-1/2}) \tag{B.71}$$

That is, $\sup_{\pi \in \Pi} ||y_t^m(\hat{\psi}_n(\pi), \pi_k) - y_t^m(\theta_n)|| \xrightarrow{p} 0$. By Lemma B.1.5, $y_t^m(\hat{\psi}_n(\pi), \pi_k) \overset{d}{\sim} y_t$. Using the same argument as in Lemma B.1.6(a), $Q_n(W_t^m(\pi_k), \theta)$ depends on $y_t^m(\hat{\psi}_n(\pi), \pi_k)$ and $X_{1,t}^m$

which have the same distribution as y_t, X_t respectively. In consequence $Q_n(W_t^m(\pi_k), \theta) \xrightarrow{p} Q_0(W_t, \theta) = \mathbb{E}_{\theta_0}(y_t - \zeta' X_{1,t} - \beta' h(X_{2,t}, \pi))^2$ and by the mapping theorem of Van Der Vaart and Wellner (1996),

$$\sup_{\pi \in \Pi} || \min_{\psi \in \Psi(\pi)} Q_n(W_t^m(\pi_k), \theta) - \min_{\psi \in \Psi(\pi)} Q_n(W_t, \theta) || = \sup_{\pi \in \Pi} || \hat{\psi}_n^m(\pi, \pi_k) - \hat{\psi}_n(\pi) || \xrightarrow{p} 0 \quad \blacksquare \tag{B.72}$$

Proposition 2.4.2 *Suppose Assumption J to Assumption N hold. Suppose that the true value of the weakly identified parameter π_n is known and imposed on Step 3 WI. The following holds letting $M_n \to \infty$ as $n \to \infty$,*

$$\sup_{z \in \mathbb{R}^{d_\theta}} \left| \mathbb{P}_m \begin{pmatrix} \sqrt{M_n}(\hat{\psi}_n^m(\hat{\pi}_n^m, \pi_n) - \hat{\psi}_n(\pi_n)) \leq z_1 \\ \hat{\pi}_n^m(\pi_n) \leq z_2 \end{pmatrix} - \mathbb{P}_{\theta_n} \begin{pmatrix} \sqrt{n}(\hat{\psi}_n - \psi_n) \leq z_1 \\ \hat{\pi}_n \leq z_2 \end{pmatrix} \right| \xrightarrow{p} 0 \tag{2.26}$$

where $\mathbb{P}_m$ is the bootstrap induced probability measure.

Proof of Proposition 2.4.2 The first step is to notice that,

$$\sup_{z \in \mathbb{R}^{d_\theta}} \mid \mathbb{P}_m(\sqrt{n}(\hat{\psi}_n^m(\hat{\pi}_n^m) - \psi_n) \leq z) - \mathbb{P}_n(\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n) \leq z) \mid \tag{B.73}$$

$$\leq \sup_{z \in \mathbb{R}^{d_\theta}} \mid \mathbb{P}_m(\sqrt{n}(\hat{\psi}_n^m(\hat{\pi}_n^m) - \psi_n) \leq z) - \mathbb{P}_{\theta_0}(\tau(\pi^*, \theta_0, b) \leq z) \mid +$$

$$\sup_{z \in \mathbb{R}^{d_\theta}} \mid \mathbb{P}_n(\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n) \leq z) - \mathbb{P}_{\theta_0}(\tau(\pi^*, \theta_0, b) \leq z) \mid \tag{B.74}$$

Similarly,

$$\sup_{z \in \mathbb{R}^{d_\theta}} \mid \mathbb{P}_m(\hat{\pi}_n^m \leq z) - \mathbb{P}_n(\hat{\pi}_n \leq z) \mid \tag{B.75}$$

$$\leq \sup_{z \in \mathbb{R}^{d_\theta}} \mid \mathbb{P}_m(\hat{\pi}_n^m \leq z) - \mathbb{P}_{\theta_0}(\pi^*(\theta_0, b) \leq z) \mid + \tag{B.76}$$

$$\sup_{z \in \mathbb{R}^{d_\theta}} \mid \mathbb{P}_n(\hat{\pi}_n \leq z) - \mathbb{P}_{\theta_0}(\pi^*(\theta_0, b) \leq z) \mid \tag{B.77}$$

Equation (B.74) and Equation (B.76) converge to zero by Theorem 3.1 of Andrews and Cheng (2012) or in our context by Proposition 1.3.2 of Chapter 1.

Part 1. We begin by showing that $\hat{\pi}_n^m \xrightarrow{d} \pi^*(\theta_0, b)$ where $\pi^*(\theta_0, b) = \min_{\pi \in \Pi} \xi(\pi)$ in Equation (B.76). In Proposition 1.3.2 of Chapter 1 we show that $\{\xi_n(\pi) : \pi \in \Pi\} \Rightarrow \{\xi(\pi) : \pi \in \Pi\}$, where $\xi(\pi)$ is defined in as the limit distribution such that $\hat{\pi}_n \xrightarrow{d} \min_{\pi \in \Pi} \xi(\pi)$ (see the Appendix A for the definition of the stochastic process $\xi(\pi)$). Using the fact that $y_t^m(\hat{\psi}_n^m(\pi_n), \pi_n) = \hat{\zeta}_n^m(\pi_n)' X_{1,t} + \hat{\beta}_n^m(\pi_n)' h(X_{2,t}, \pi_n) + \epsilon_t^m(\hat{\psi}_{0,n})$ is generated with known value π_n. Let,

$$\xi_n^m(\pi) = n(Q_n^{c,m}(\cdot) - Q_{0,n}^m) \tag{B.78}$$

$$Q_n^{c,m}(\cdot) = n^{-1} \sum_{t=1}^{n} (y_t^m(\hat{\psi}_n^m(\pi_n), \pi_n) - \hat{\zeta}_n^m(\cdot)' X_{1,t} - \hat{\beta}_n^m(\cdot)' h(X_{2,t}, \cdot))^2 \tag{B.79}$$

$$Q_{0,n}^m = n^{-1} \sum_{t=1}^{n} (y_t^m(\hat{\psi}_n^m(\pi_n), \pi_n) - \hat{\zeta}_{0,n}^{m\prime} X_{1,t})^2 \tag{B.80}$$

No we show that $\xi_n^m(\pi) \Rightarrow \xi(\pi)$. Consider $Q_n^{c,m}(\cdot) = O_{p,\pi}(n^{-1})$,

$$Q_n^{c,m}(\cdot) = n^{-1} \sum_{t=1}^{n} (y_t^m(\hat{\psi}_n^m(\pi_n), \pi_n) - \hat{\zeta}_n^m(\cdot)' X_{1,t} - \hat{\beta}_n^m(\cdot)' h(X_{2,t}, \cdot))^2 \tag{B.81}$$

$$= n^{-1} \sum_{t=1}^{n} (\zeta_n' X_{1,t} + \beta_n' h(X_{2,t}, \pi_n) + \epsilon_t^m(\hat{\psi}_{0,n}) - \hat{\zeta}_n^m(\cdot)' X_{1,t} - \hat{\beta}_n^m(\cdot)' h(X_{2,t}, \cdot)$$

$$+ (\hat{\zeta}_n^m(\pi_n) - \zeta_n)' X_{1,t} + (\hat{\beta}_n^m(\pi_n) \ - \beta_n)' h(X_{2,t}, \pi_n))^2 \tag{B.82}$$

$$= n^{-1} \sum_{t=1}^{n} (y_t^m(\theta_n) - \hat{\zeta}_n^m(\cdot)' X_{1,t} - \hat{\beta}_n^m(\cdot)' h(X_{2,t}, \cdot))^2 \tag{B.83}$$

$$+ \underbrace{n^{-1} \sum_{t=1}^{n} ((\hat{\zeta}_n^m(\pi_n) - \zeta_n)' X_{1,t} + (\hat{\beta}_n^m(\pi_n) \ - \beta_n)' h(X_{2,t}, \pi_n))^2}_{(I)} \tag{B.84}$$

$$+ \underbrace{n^{-1} \sum_{t=1}^{n} (y_t^m(\theta_n) - \hat{\zeta}_n^m(\cdot)' X_{1,t} - \hat{\beta}_n^m(\cdot)' h(X_{2,t}, \cdot))}_{(II)} \tag{B.85}$$

$$\underbrace{[(\hat{\zeta}_n^m(\pi_n) - \zeta_n)' X_{1,t} + (\hat{\beta}_n^m(\pi_n) \ - \beta_n)' h(X_{2,t}, \pi_n)]}_{(II)} \tag{B.86}$$

where $y_t^m(\theta_n) = \zeta_n' X_{1,t} + \beta_n' h(X_{2,t}, \pi_n) + \epsilon_t^m(\hat{\psi}_{0,n})$. By Lemma B.1.4, $y_t^m(\theta_n) \overset{d}{\sim} y_t$, that is, the boostrapped sample of y_t using the true parameters has the same distribution as y_t. First, consider (I),

$$(I) = \frac{1}{n^2}\sum_{t=1}^{n}[(\sqrt{n}(\hat{\zeta}_n^m(\pi_n) - \zeta_n)'X_{1,t})^2 + (\sqrt{n}(\hat{\beta}_n^m(\pi_n) \ - \beta_n)'h(X_{2,t},\pi_n))^2$$
$$+ (\sqrt{n}(\hat{\zeta}_n^m(\pi_n) - \zeta_n)'X_{1,t})(\sqrt{n}(\hat{\beta}_n^m(\pi_n) \ - \beta_n)'h(X_{2,t},\pi_n))] = O_p(n^{-3/2}) \tag{B.87}$$

Next consider (II), i.e. the product of Equations (B.85) and (B.86)

$$(II) = \frac{1}{n^{3/2}}\sum_{t=1}^{n}[(y_t^m(\theta_n) - \hat{\zeta}_n^m(\cdot)'X_{1,t} - \hat{\beta}_n^m(\cdot)'h(X_{2,t},\cdot))$$
$$* (\sqrt{n}(\hat{\zeta}_n^m(\pi_n) - \zeta_n)'X_{1,t} + \sqrt{n}(\hat{\beta}_n^m(\pi_n) - \beta_n)'h(X_{2,t},\pi_n)] = O_{p,\pi}(n^{-1}) \tag{B.88}$$

Now we use the other term of $\xi(\pi)$. Consider $Q_{0,n}^m$

$$Q_{0,n}^m = n^{-1}\sum_{t=1}^{n}(y_t^m(\hat{\psi}_n^m(\pi_n),\pi_n) - \hat{\zeta}_{0,n}^{m\prime}X_{1,t})^2 \tag{B.89}$$
$$= n^{-1}\sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_{0,n}^{m\prime}X_{1,t})^2 \tag{B.90}$$
$$+ \underbrace{n^{-1}\sum_{t=1}^{n}((\hat{\zeta}_n^m(\pi_n) - \zeta_n)'X_{1,t} + (\hat{\beta}_n^m(\pi_n) \ - \beta_n)'h(X_{2,t},\pi_n))^2}_{(I)} \tag{B.91}$$
$$+ \underbrace{\frac{1}{n^{3/2}}\sum_{t=1}^{n}[(y_t^m(\theta_n) - \hat{\zeta}_{0,n}^{m\prime}X_{1,t})(\sqrt{n}(\hat{\zeta}_n^m(\pi_n) - \zeta_n)'X_{1,t} + \sqrt{n}(\hat{\beta}_n^m(\pi_n) \ - \beta_n)'h(X_{2,t},\pi_n))]}_{(III)} \tag{B.92}$$

By Equation (B.87), we have $(I) = O_p(n^{-3/2})$. Taking (II) and (III),

$$(II) - (III) = \frac{1}{n^2}\sum_{t=1}^{n}(\sqrt{n}(\hat{\zeta}_{0,n}^m - \hat{\zeta}_n^m(\cdot)'X_{1,t} - \sqrt{n}\hat{\beta}_n^m(\cdot)'h(X_{2,t},\cdot))$$
$$* (\sqrt{n}(\hat{\zeta}_n^m(\pi_n) - \zeta_n)'X_{1,t} + \sqrt{n}(\hat{\beta}_n^m(\pi_n) - \beta_n)'h(X_{2,t},\pi_n)) = O_{p,\pi}(n^{-3/2}) \tag{B.93}$$

Combining the results from $(I), (II)$ and (III),

$$
\begin{aligned}
(Q_n^{c,m}(\cdot) - Q_{0,n}^m)) &= n^{-1}\sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_n^m(\cdot)'X_{1,t} - \hat{\beta}_n^m(\cdot)'h(X_{2,t},\cdot))^2 - n^{-1}\sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_{0,n}^{m\prime}X_{1,t})^2 \\
&= (I) + (II) - (I) - (III) && \text{(B.94)} \\
&= n^{-1}\sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_n^m(\cdot)'X_{1,t} && \text{(B.95)} \\
&- \hat{\beta}_n^m(\cdot)'h(X_{2,t},\cdot))^2 - n^{-1}\sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_{0,n}^{m\prime}X_{1,t})^2 + O_{p,\pi}(n^{-3/2}) && \text{(B.96)}
\end{aligned}
$$

Which is what we needed for the results as,

$$
\begin{aligned}
\xi_n^m(\cdot) &= n(Q_n^{c,m}(\cdot) - Q_{0,n}^m)) && \text{(B.97)} \\
&= \sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_n^m(\cdot)'X_{1,t} && \text{(B.98)} \\
&- \hat{\beta}_n^m(\cdot)'h(X_{2,t},\cdot))^2 - \sum_{t=1}^{n}(y_t^m(\theta_n) - \hat{\zeta}_{0,n}^{m\prime}X_{1,t})^2 + O_{p,\pi}(n^{-1/2}) && \text{(B.99)} \\
&= \xi_n(\cdot) + O_{p,\pi}(n^{-1/2}) \Rightarrow \xi(\cdot) && \text{(B.100)}
\end{aligned}
$$

By Lemma A.2.0.6 of Appendix A as $y_t^m(\theta_n) \overset{d}{\sim} y_t$, and $\hat{\psi}_n^m$ are obtained identically to $\hat{\psi}_n$, using $y_t^m(\theta_n)$ instead of y_t. We use the equivalence in distribution argument as in Lemma B.1.4, conditional on the sample X_t from Giné and Zinn (1990). The stochastic processes $\xi_n^m(\pi)$ and $\xi_n(\pi)$ are constructed analogously, and by Equation (B.100) $\{\xi_n^m(\pi) : \pi \in \Pi\} \Rightarrow \{\xi(\pi) : \pi \in \Pi\}$, and in consequence, $\hat{\pi}_n^m \xrightarrow{d} \pi^*(\theta_0, b)$ by the mapping theorem of Van Der Vaart and Wellner (1996).

Part 2. To show $\sqrt{n}(\hat{\psi}_n^m(\pi) - \psi_n) \xrightarrow{d} \tau(\pi^*(\theta_0, b))$, the proof follows the lines of Part 1. Consider,

$$
\begin{aligned}
\hat{\psi}_n^m(\pi) &= \inf_{\psi\in\Psi(\pi)} n^{-1}\sum_{t=1}^{n}(y_t^m(\hat{\psi}_n(\pi_n),\pi_n) - \zeta'X_{1,t} - \beta'h(X_{2,t},\pi_n))^2 && \text{(B.101)} \\
&= \inf_{\psi\in\Psi(\pi)} n^{-1}\sum_{t=1}^{n}(y_t^m(\theta_n) + (\hat{\zeta}_n(\pi_n) - \zeta_n)'X_{1,t} + (\hat{\beta}_n(\pi_n) - \beta_n)'h(X_{2,t},\pi_n) \\
&- \zeta'X_{1,t} - \beta'h(X_{2,t},\pi_n))^2 && \text{(B.102)}
\end{aligned}
$$

$$= \inf_{\psi \in \Psi(\pi)} n^{-1/2} \sum_{t=1}^{n} (y_t^m(\theta_n) - \zeta' X_{1,t} - \beta' h(X_{2,t}, \pi_n))^2 + O_p(n^{-1/2}) \tag{B.103}$$

The last line follow multiplying by $n^{1/2}$. Moreover, the first term is $O_p(1)$ while the second is $O_p(n^{-1/2})$, using a similar argument as (I) and (II) of Part 1. More specifically, take the cross product of Equation (B.102),

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} (y_t^m(\theta_n) - \zeta' X_{1,t} - \beta' h(X_{2,t}, \pi_n)) \tag{B.104}$$

$$* (\sqrt{n}(\hat{\zeta}_n(\pi_n) - \zeta_n)' X_{1,t} + \sqrt{n}(\hat{\beta}_n(\pi_n) - \beta_n)' h(X_{2,t}, \pi_n)) = O_p(n^{-1}) \tag{B.105}$$

By Equation (B.103), $\hat{\psi}_n^m(\pi_n)$ is obtained using $y_t(\theta_n) \overset{d}{\sim} y_t$ and the same objective function. Moreover, the parameters are all strongly identified as the weakly identified parameter has been fixed to its true value π_n. This proves that $\hat{\psi}_n^m(\pi_n)$ is getting closer and closer to $\hat{\psi}_n$, as in Part 1 of the proof. Therefore for large enough n, $\hat{\psi}_n^m$ has the same distribution as $\hat{\psi}_n$. In consequence $\sqrt{n}(\hat{\psi}_n^m(\pi) - \psi_n) \xrightarrow{d} \tau(\pi^*(\theta_0, b).\theta_0, b)$ by Lemma A.2.0.5. of Appendix A. Joint convergence of ψ and π follows from the Cramer Wold Device. ■

Proposition 2.4.3 *Suppose Assumption J to Assumption N hold. Let π_n be unknown. Under weak identification, the following holds letting $M_n \to \infty$ as $n \to \infty$.*

$$\liminf_{n\to\infty} \ \mathbb{P}_m \begin{pmatrix} \sqrt{M_n}(\hat{\psi}_n^m(\hat{\pi}_n^m, \pi_k) - \hat{\psi}_n(\pi_k)) & \leq & \sup\limits_{\pi_k \in \tilde{\Pi}} c_{1-\alpha}^{\psi}(\pi_k) \\ \hat{\pi}_n^m(\pi_k) & \leq & \sup\limits_{\pi_k \in \tilde{\Pi}} c_{1-\alpha}^{\pi}(\pi_k) \end{pmatrix} \geq 1 - \alpha \tag{2.27}$$

with probability approaching one.

Proof of Proposition 2.4.3 It is easier to prove the inequality for each marginal distribution. We follow this approach and argue that the joint result follows by a similar argument. First we show it for π, i.e. we show, $\liminf_{n\to\infty} \mathbb{P}_{m,\pi}(\hat{\pi}_n(\pi_k) \leq \sup\limits_{\pi_k \in \tilde{\Pi}} c_{1-\alpha}^{\pi}(\pi_k)) \geq 1 - \alpha$ w.p.a.1, where $\mathbb{P}_{m,\pi}$ denotes the marginal distribution probability measure with respect to π, which is well defined as the joint distribution exists. Let $\hat{\pi}_n(\pi_k)$ be the estimators of π constructed with the data generating process that sets$\pi = \pi_k$, specifically

$$y_t(\hat{\psi}_n, \psi_k) = \hat{\zeta}_n' X_{1,t} + \hat{\beta}_n' h(X_{2,t}, \pi_k) + \epsilon_t(\hat{\psi_{0,n}}) \tag{B.106}$$

Consider the following terms,

$$\vartheta_n^{m,\pi}(\pi_k) = \hat{\pi}_n^m(\pi_k) \tag{B.107}$$

$$\vartheta_n^{\pi}(\pi_k) = \hat{\pi}_n(\pi_k) \tag{B.108}$$

$$\vartheta^{\pi}(\pi_k) = \pi^*(\theta_0, b; \pi_k) \tag{B.109}$$

The expression $\vartheta_n^{m,\pi}(\pi_k)$ depends on π_k because we assume that the bootstrapped estimators are generated using the data generating process setting π_k in Step 3 WI. We consider the right tailed critical values,

$$c_{n,1-\alpha}^{m,\pi}(\pi_k) = \inf\{c \geq 0 : \mathbb{P}_{m,\pi}(\vartheta_n^{m,\pi}(\pi_k) \leq c) \geq 1-\alpha\} \tag{B.110}$$

$$c_{n,1-\alpha}^{\pi}(\pi_k) = \inf\{c \geq 0 : \mathbb{P}_{\theta_n,\pi}(\vartheta_n^{\pi}(\pi_k) \leq c) \geq 1-\alpha\} \tag{B.111}$$

$$c_{1-\alpha}^{\pi}(\pi_k) = \inf\{c \geq 0 : \mathbb{P}_{\theta_0,\pi}(\vartheta^{\pi}(\pi_k) \leq c) \geq 1-\alpha\} \tag{B.112}$$

where $\mathbb{P}_{\theta_0,\pi}$ and $\mathbb{P}_{\theta_n,\pi}$ are the marginal distribution probability measure with respect to θ_0 and θ_n respectively. Consider the supremum of the one tailed critical values with respect to π_k, that is, we take the largest critical value with size $1-\alpha$ from all possible data generating process generated along the grid of $\tilde{\Pi}$. We construct the coverage probabilities,

$$CP_{n,1-\alpha}^{m,\pi}(\pi_k) = \mathbb{P}_{m,\pi}(\vartheta_n^{m,\pi}(\pi_k) \leq \sup_{\pi_k \in \tilde{\Pi}} c_{n,1-\alpha}^{m,\pi}(\pi_k)) \tag{B.113}$$

$$CP_{n,1-\alpha}^{\pi}(\pi_k) = \mathbb{P}_{\theta_n,\pi}(\vartheta_n^{\pi}(\pi_k) \leq \sup_{\pi_k \in \tilde{\Pi}} c_{n,1-\alpha}^{\pi}(\pi_k)) \tag{B.114}$$

$$CP_{1-\alpha}^{\pi}(\pi_k) = \mathbb{P}_{\theta_0,\pi}(\vartheta^{\pi}(\pi_k) \leq \sup_{\pi_k \in \tilde{\Pi}} c_{1-\alpha}^{\pi}(\pi_k)) \tag{B.115}$$

Clearly, $\mathbb{P}_{m,\pi}(\vartheta_n^{m,\pi}(\pi_k) \leq c_{n,1-\alpha}^{m,\pi}(\pi_k)) = 1-\alpha$. All coverage probabilities are well defined as $\xi(\pi)$ is a stochastic process with continuous sample paths a.s. and therefore the critical values are well defined and unique. By Proposition 1.3.2 of Chapter 1, $c^{\pi}_{n,1-\alpha}(\pi_k) \to c^{\pi}_{1-\alpha}(\pi_k)$ as $n \to \infty$ for each $\pi_k \in \tilde{\Pi}$, and by Theorem 1.4.2 of the same paper, $c^{m,\pi}_{n,1-\alpha}(\pi_k) = c^{\pi}_{n,1-\alpha}(\pi_k) + o_p(1) \xrightarrow{p} c^{\pi}_{1-\alpha}(\pi_k)$. Notice that $c^{\pi}_{1-\alpha}(\pi_k)$ depends on the optimal draw of $\pi^*(\theta_{0,k}, b)$, where $\theta_{0,k}$ denotes the limit of $\theta_{n,k}$ which imposes $\pi = \pi_k$ as the true value of π. Therefore, $\sup_{\pi_k \in \tilde{\Pi}} c^{\pi}_{n,1-\alpha}(\pi_k) \to \sup_{\pi_k \in \tilde{\Pi}} c^{\pi}_{1-\alpha}(\pi_k)$ and $\sup_{\pi_k \in \tilde{\Pi}} c^{m,\pi}_{n,1-\alpha}(\pi_k) \xrightarrow{p} \sup_{\pi_k \in \tilde{\Pi}} c^{\pi}_{n,1-\alpha}(\pi_k)$ as $n \to \infty$. Assumption Q states that the coverage probabilities of the marginal distribution with respect to π converge, i.e. $CP^{\pi}_{n,1-\alpha}(\theta_k) \to CP^{\pi}_{1-\alpha}(\theta_k, p)$, for some $p \in \mathcal{P}$. Using this assumption and Theorem 1.4.2, $CP^{m,\pi}_{n}(\pi_k) = CP^{\pi}_{n,1-\alpha}(\pi_k) + o_p(1)$. In consequence, $CP^{m,\pi}_{n,1-\alpha}(\pi_k) = CP^{\pi}_{1-\alpha}(\pi_k, p) + o_p(1) \geq 1 - \alpha + o_p(1)$, i.e. $\liminf_{n\to\infty} CP^{m,\pi}_{n,1-\alpha}(\pi_k) \geq 1 - \alpha$ w.p.a.1.

Now we prove the result with respect to the margnial distribution of ψ. For every $\pi_k \in \tilde{\Pi}$, define the following,

$$\vartheta^{m,\psi}_n(\pi, \pi_k) = \sqrt{n}(\hat{\psi}^m_n(\pi, \pi_k) - \hat{\psi}_n(\pi_k)) \tag{B.116}$$

$$\vartheta^{\psi}_n(\pi, \pi_k) = \sqrt{n}(\hat{\psi}_n(\pi, \pi_k) - \hat{\psi}_n) \tag{B.117}$$

$$\vartheta^{\psi}(\pi, \pi_k) = \tau(\pi, \theta_0, b; \pi_k) \tag{B.118}$$

where $\tau(\pi, \theta_0, b)$ is defined on Lemma A.2.0.5 of Appendix . We consider one tailed critical values,

$$c^{m,\psi}_{n,1-\alpha}(\pi, \pi_k) = \inf\{c \geq 0 : \mathbb{P}_{m,\psi}(\vartheta^{m,\psi}_n(\pi, \pi_k) \leq c) \geq 1 - \alpha\} \tag{B.119}$$

$$c^{\psi}_{n,1-\alpha}(\pi, \pi_k) = \inf\{c \geq 0 : \mathbb{P}_{\theta_n,\psi}(\vartheta^{\psi}_n(\pi, \pi_k) \leq c) \geq 1 - \alpha\} \tag{B.120}$$

$$c^{\psi}_{1-\alpha}(\pi, \pi_k) = \inf\{c \geq 0 : \mathbb{P}_{\theta_0,\psi}(\vartheta^{\psi}(\pi, \pi_k) \leq c) \geq 1 - \alpha\} \tag{B.121}$$

Now define the coverage probabilities with respect to ψ,

$$CP^{m}_{n,1-\alpha}(\pi, \pi_k) = \mathbb{P}_{m,\psi}(v^{m,\pi}_n(\pi, \pi_k) \leq \sup_{\pi_k \in \tilde{\Pi}} c^{m,\pi}_{n,1-\alpha}(\pi, \pi_k)) \tag{B.122}$$

$$CP_{n,1-\alpha}(\pi,\pi_k) = \mathbb{P}_{\theta_n,\psi}(v_n^{\pi}(\pi,\pi_k) \leq \sup_{\pi_k\in\tilde{\Pi}} c^{\pi}_{n,1-\alpha}(\pi,\pi_k)) \tag{B.123}$$

$$CP_{1-\alpha}(\pi,\pi_k) = \mathbb{P}_{\theta_0,\psi}(v^{\pi}(\pi,\pi_k) \leq \sup_{\pi_k\in\tilde{\Pi}} c^{\pi}_{1-\alpha}(\pi,\pi_k)) \tag{B.124}$$

All coverage probabilities are well defined as $\tau(\pi,\theta_0,b)$ is a stochastic process with continuous sample paths a.s. and therefore the critical values are well defined and unique. By Proposition 1.3.2 of Chapter 1, $c^{\psi}_{n,1-\alpha}(\pi_n,\pi_k) \to c^{\psi}_{1-\alpha}(\pi_k)$ for each $\pi_k \in \tilde{\Pi}$. Notice that, $c^{\psi}_{1-\alpha}$ depends on the random draw of π when we set the data generating process with $\pi = \pi_k$, that is $c^{\psi}_{1-\alpha}(\pi_k) = c^{\psi}_{1-\alpha}(\pi^*(\theta_{0,k},b),\pi_k)$, where $\theta_{0,k}$ denotes the limit sequence of θ setting $\pi = \pi_k$. By Theorem 1.4.2 of the same paper, $c^{m,\psi}_{n,1-\alpha}(\hat{\pi}^m_n,\pi_k) = c^{\psi}_{n,1-\alpha}(\pi_n,\pi_k) + o_p(1) \xrightarrow{p} c^{\psi}_{1-\alpha}(\pi_k)$. Therefore, $\sup_{\pi_k\in\tilde{\Pi}} c^{\psi}_{n,1-\alpha}(\pi_n,\pi_k) \to \sup_{\pi_k\in\tilde{\Pi}} c^{\psi}_{1-\alpha}(\pi_k)$ and in consequence, $\sup_{\pi_k\in\tilde{\Pi}} c^{m,\psi}_{n,1-\alpha}(\hat{\pi}^m_n,\pi_k) \xrightarrow{p} \sup_{\pi_k\in\tilde{\Pi}} c^{\psi}_{1-\alpha}(\pi_k)$. Assumption Q states that, $CP^{\psi}_{n,1-\alpha}(\theta_k) \to CP^{\psi}_{1-\alpha}(\theta_k,p)$, for some $p \in \mathcal{P}$. Using this assumption and Theorem 1.4.2, (with respect to the coverage probabilities of the marginal distribution with respect to ψ) $CP^{m,\psi}_n(\pi_k) = CP^{\psi}_{n,1-\alpha}(\pi_k) + o_p(1)$. In consequence, $CP^{m,\psi}_{n,1-\alpha}(\pi_k) = CP^{\psi}_{1-\alpha}(\pi_k,p) + o_p(1) \geq 1-\alpha + o_p(1)$, i.e. $\liminf_{n\to\infty} CP^{m,\psi}_{n,1-\alpha}(\pi_k) \geq 1-\alpha$ w.p.a.1.

The argument with respect to the joint distribution of ψ,π follows a verbatim using the joint distribution instead of the marginals. All critical values are well defined and are unique as the stochastic process have continuous sample paths (a.s) on a compact set. Another way to prove it is using the Cramer Wold Device. The proof that considers the left tailed or two-tailed critical values follows the same argument considering $\inf_{\pi_k\in\Pi} c^{\psi}_{1-\alpha}(\pi_k)$ and $\inf_{\pi_k\in\Pi} c^{\pi}_{1-\alpha}(\pi_k)$. ■

Theorem 2.6.1 *Suppose Assumption J to Assumption Q are satisfied. Let π_n be known. Using the parametric bootstrap introduced in from Step 1 WI to Step 4 WI and Step 1 SI to Step 4 WI, and letting $M_n \to \infty$ as $n \to \infty$ for each identification category,*

(a) Under weak identification with $\dim(r_\pi(\theta)) = 0$, $T^{\psi,m}_n(\hat{\pi}^m_n(\pi_n)) \xrightarrow{d} T^{\psi}(\pi^*(\theta_0,b);\theta_0,b)$

(b) Under weak identification with $\dim(r_\pi(\theta)) = 1$, $T^{\pi,m}_n(\hat{\pi}^m_n(\pi_n)) \xrightarrow{d} T^{\pi}(\pi^*(\theta_0,b);\theta_0,b)$

(c) Under strong identification, $T_n^{\theta,m} \xrightarrow{d} N(0,1)$

Proof of Theorem 2.6.1 As we are assuming that the true π_n is known, the proof follows almost verbatim from Proposition 1.3.2 and Proposition 1.3.3 of Chapter 1 and using the bootstrap convergence result of Proposition 2.4.2. See also Theorem 3.1 and Theorem 4.1 of Andrews and Cheng (2012). In particular, as we have shown that $\hat{\pi}_n^m(\pi_k) \xrightarrow{d} \pi^*(\theta_0, b)$, the result follows from the continuous mapping theorem and Lemma A.2.0.7 and Lemma A.2.0.8 of Appendix A. The variance-covariance matrices $\Sigma_n(\theta), \Sigma(\theta)$ are well defined. The convergence in probability to the limit covariance matrices follows by Lemma A.2.0.7 and Lemma A.2.0.8. See the simulation exercise of Chapter 1 for an example of matrices $J_n(\theta), V_n(\theta), J(\theta), V(\theta)$ using an exponential smoothing regression function. ■

Theorem 2.6.2 *Let $M_n \to \infty$ as $n \to \infty$. Under the null hypothesis $H_0 : r(\theta) = q$, the LF and ICS_0 critical values of the t-test have correct asymptotic size w.p.a.1,*

(a) $AsySz^{LF,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c_{n,1-\alpha}^{LF,m}(r(\theta))) = 1-\alpha$

(b) $AsySz^{ICS_0,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta: r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c_{n,1-\alpha}^{ICS_0,m}(r(\theta))) = 1-\alpha$

(c) If H_0 is false, that is $r(\theta_n) \neq q$, then $T_n(r(\theta)) \xrightarrow{p} \infty$

Proof of Theorem 2.6.2 The proof of this theorem uses Lemma 2.1 of Andrews and Cheng (2012). This results states that,

$$AsySz = \min\{\inf_{p\in\mathcal{P}} CP(p), CP_\infty\} \tag{B.125}$$

where $CP(p)$ is the limit coverage probability of the t-test under weak identification.

(a) To construct the least favorable critical values, consider the t-test critical values,

$$c_{n,1-\alpha}^{m,T^a}(\pi_k) = \inf\{c \geq 0 : \mathbb{P}_m(T_n^{a,m}(\pi_k) \leq c) \geq 1-\alpha\} \tag{B.126}$$

$$c_{n,1-\alpha}^{,T^a}(\pi_k) = \inf\{c \geq 0 : \mathbb{P}_{\theta_n}(T_n^{a}(\pi_k) \leq c) \geq 1-\alpha\} \tag{B.127}$$

$$c_{1-\alpha}^{,T^a}(\pi_k) = \inf\{c \geq 0 : \mathbb{P}_{\theta_0}(T^{a}(\pi_k) \leq c) \geq 1-\alpha\} \tag{B.128}$$

where T^a can be either T^ψ or T^π according to which parameter is tested. Notice that the least favourable critical values take the following form,

$$c_{n,1-\alpha}^{LF^a,m} = \max\{\sup_{\pi_k\in\Pi} c_{n,1-\alpha}^{m,T^a}(\pi_k), c_{n,1-\alpha}^{m,T^a,s}\} \tag{B.129}$$

where $c_{n,1-\alpha}^{m,T^a,s}$ denotes the critical values of the t-statistic under strong identification, $a = \psi, \pi$. By Equation (B.125) and Theorem 2.6.1,

$$AsySz^{LF^a,m} = \liminf_{n\to\infty} \inf_{\theta\in\Theta:r(\theta)=q} \mathbb{P}_m(T_n(r(\theta)) \leq c_{n,1-\alpha}^{LF,m}(r(\theta))) \tag{B.130}$$

$$= \min\{\mathbb{P}_{\theta_0}(T^a(r(\theta)) \leq c_{1-\alpha}^{LF^a}), \mathbb{P}_{\theta_0}(N(0,1) \leq c_{1-\alpha}^{LF^a})\} + o_p(1) \tag{B.131}$$

as the t-statistic is pivotal under strong identification and the $o_p(1)$ term follows from Theorem 2.6.1. Notice that under weak identification $\mathbb{P}_{\theta_0}(T^a(r(\theta)) \leq c_{1-\alpha}^{LF^a}) \geq 1-\alpha$ and under strong identification $\mathbb{P}_{\theta_0}(N(0,1) \leq c_{1-\alpha}^{LF^a}) = 1-\alpha$. In either case we have that $AsySz^{LF^a,m} \geq 1-\alpha$ w.p.a.1.

(b) Consider the ICS_0 critical value. In the limit, we assume that we are able to correctly conclude if the model is weakly identified or strongly identified. Under weak identification, $\mathbb{P}(\hat{A}_n \leq \kappa_n) \xrightarrow{p} 1$ as $n \to \infty$. This would imply that $c_{n,1-\alpha}^{ICS,m} = c_{n,1-\alpha}^{LF,m} = c_{1-\alpha}(\pi_k) + o_p(1) \geq \sup_{\pi_k\in\Pi} c_{1-\alpha}(\pi_k) + o_p(1)$ whose size is greater or equal to $1-\alpha$ by Proposition 2.4.2. In the case of strong identification, $\mathbb{P}(\hat{A}_n \leq \kappa_n) \xrightarrow{p} 0$ and therefore $c_{n,1-\alpha}^{ICS,m} = c_{1-\alpha}^s + o_p(1)$ whose size is equal to $1-\alpha$. In consequence, letting $\inf_{p\in\mathcal{P}} CP(p)$ being the coverage probability using critical value $\sup_{\pi_k\in\Pi} c_{1-\alpha}(\pi_k)$ and CP_∞ being the coverage probability using critical value $c_{1-\alpha}^s$, then $AsySz = \min\{\inf_{p\in\mathcal{P}} CP(p), CP_\infty\} = 1-\alpha$ w.p.a.1.

(c) First consider T_n, only for restrictions of ψ, that is with $\dim(r_\pi) = 0$. This proof follows the argument of Theorem 1.4.3 in Chapter 1. If H_0 is false $r(\theta_n) \neq q_n$,

$$r(\hat{\theta}_n) - r(\theta_n) = r(\hat{\theta}_n) - r(\theta_n) + r(\theta_n) - q_n = CI_k + o_p(1) \tag{B.132}$$

for some $C \neq 0$ and identity matrix I_k and large enough n and by consistency of the numerator. Then we have

$$T_n = \frac{\sqrt{n}C}{r_\psi(\hat{\theta}_n)\Sigma_n r_\psi(\hat{\theta}_n)} + o_p(1) \to \infty \text{ as } n \to \infty \tag{B.133}$$

as the denominator converges in probability to $r_\psi(\theta_0)\Sigma(\theta_0)r_\psi(\theta_0)$ which is finite and non-random.

Similarly consider T_n with $\dim(r_\pi) = 1$,

$$r(\hat{\theta}_n) - r(\hat{\theta}_n) = r(\hat{\psi}_n, \hat{\pi}_n) - r(\psi_n, \pi_n) + r(\psi_n, \pi_n) - q_n = CI_k + O_p(1) \tag{B.134}$$

Using the arguments in the proof of Proposition 1.3.3 of Chapter 1,

$$T_n = \frac{||\sqrt{n}\hat{\beta}_n||(r(\psi_n, \hat{\pi}_n) - r(\psi_n, \pi_n))}{[r_\pi(\psi_n, \hat{\pi}_n)\Sigma_n r_\pi(\psi_n, \hat{\pi}_n)]^{1/2}} + o_p(1) \tag{B.135}$$

$$= \frac{\sqrt{n}C}{r_\psi(\hat{\theta}_n)\Sigma_n r_\psi(\hat{\theta}_n)} + o_p(1) \xrightarrow{p} \infty \text{ as } n \to \infty \tag{B.136}$$

as the denominator converges in probability to $r_\pi(\theta_0)\Sigma(\theta_0)r_\pi(\theta_0)$ which is finite and non-random. ■

B.2 Supplemental Tables and Figures

	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
Strongly Identified $n = 100$ $(\beta_n = 0.5)$															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.081	0.086	0.090	0.090	0.091	0.093	0.110	0.159	0.178	0.170	0.138	0.128	0.148	0.132	0.144
5%	0.041	0.042	0.044	0.044	0.044	0.067	0.078	0.101	0.128	0.115	0.080	0.065	0.085	0.066	0.085
1%	0.006	0.011	0.007	0.012	0.007	0.042	0.036	0.045	0.058	0.047	0.033	0.014	0.034	0.016	0.034
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.240	0.245	0.256	0.256	0.257	0.198	0.315	0.323	0.318	0.328	0.873	0.863	0.884	0.866	0.877
5%	0.141	0.153	0.150	0.160	0.150	0.156	0.221	0.235	0.229	0.246	0.833	0.823	0.842	0.827	0.838
1%	0.045	0.056	0.051	0.063	0.051	0.105	0.105	0.138	0.127	0.143	0.763	0.684	0.771	0.692	0.767
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.894	0.892	0.902	0.896	0.902	0.850	0.923	0.933	0.923	0.933	0.956	0.957	0.957	0.957	0.956
5%	0.823	0.827	0.833	0.834	0.832	0.798	0.867	0.894	0.867	0.894	0.951	0.952	0.953	0.952	0.952
1%	0.617	0.622	0.644	0.647	0.644	0.633	0.607	0.776	0.612	0.776	0.944	0.945	0.946	0.946	0.944
Weakly Identified $n = 100$ $(\beta_n = 0.5/\sqrt{n})$															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.101	0.105	0.107	0.107	0.108	0.115	0.112	0.137	0.182	0.254	0.057	0.088	0.096	0.095	0.065
5%	0.057	0.062	0.066	0.065	0.059	0.068	0.057	0.088	0.095	0.156	0.029	0.040	0.047	0.045	0.034
1%	0.009	0.018	0.016	0.021	0.013	0.021	0.012	0.035	0.022	0.041	0.005	0.007	0.006	0.008	0.006
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.232	0.239	0.244	0.247	0.238	0.361	0.386	0.364	0.398	0.424	0.357	0.413	0.437	0.435	0.364
5%	0.154	0.153	0.161	0.158	0.163	0.324	0.350	0.325	0.363	0.390	0.278	0.329	0.360	0.352	0.280
1%	0.055	0.064	0.066	0.066	0.064	0.210	0.222	0.210	0.249	0.303	0.155	0.182	0.258	0.212	0.157
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.891	0.898	0.901	0.903	0.905	0.923	0.943	0.945	0.963	0.974	0.799	0.833	0.841	0.844	0.799
5%	0.810	0.824	0.823	0.826	0.828	0.878	0.910	0.902	0.931	0.946	0.742	0.789	0.798	0.802	0.747
1%	0.612	0.625	0.639	0.632	0.632	0.775	0.792	0.784	0.811	0.862	0.641	0.688	0.726	0.720	0.644
Non-Identified $n = 100$ $(\beta_n = 0)$															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.093	0.101	0.101	0.104	0.101	0.092	0.102	0.094	0.110	0.232	0.040	0.073	0.084	0.082	0.045
5%	0.041	0.050	0.049	0.050	0.046	0.048	0.050	0.074	0.070	0.113	0.017	0.031	0.031	0.035	0.019
1%	0.007	0.009	0.009	0.009	0.008	0.016	0.013	0.026	0.017	0.026	0.001	0.003	0.002	0.003	0.001
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.248	0.257	0.257	0.262	0.264	0.427	0.430	0.430	0.436	0.486	0.350	0.405	0.429	0.426	0.359
5%	0.151	0.162	0.167	0.164	0.164	0.352	0.348	0.354	0.355	0.441	0.273	0.314	0.361	0.346	0.279
1%	0.049	0.055	0.056	0.058	0.056	0.199	0.187	0.199	0.197	0.256	0.134	0.165	0.242	0.188	0.134
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.897	0.905	0.904	0.908	0.907	0.939	0.944	0.945	0.949	0.971	0.764	0.812	0.813	0.818	0.766
5%	0.825	0.833	0.841	0.838	0.837	0.894	0.898	0.898	0.903	0.949	0.722	0.755	0.765	0.771	0.726
1%	0.633	0.657	0.655	0.670	0.653	0.804	0.789	0.806	0.798	0.849	0.627	0.675	0.707	0.703	0.629

Table B.1: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with Normal(0,1) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 100$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

	Strongly Identified $n = 100$ $(\beta_n = 0.5)$														
	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.112	0.112	0.119	0.122	0.119	0.086	0.092	0.130	0.200	0.147	0.133	0.122	0.142	0.127	0.138
5%	0.048	0.058	0.061	0.062	0.058	0.059	0.067	0.081	0.139	0.093	0.083	0.078	0.093	0.081	0.088
1%	0.007	0.011	0.012	0.013	0.011	0.030	0.029	0.042	0.057	0.047	0.034	0.026	0.039	0.027	0.038
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.256	0.256	0.267	0.269	0.266	0.184	0.306	0.297	0.333	0.327	0.860	0.852	0.883	0.862	0.867
5%	0.167	0.173	0.176	0.181	0.174	0.143	0.204	0.204	0.250	0.236	0.797	0.790	0.834	0.802	0.808
1%	0.059	0.063	0.063	0.071	0.066	0.088	0.106	0.108	0.162	0.120	0.663	0.587	0.703	0.610	0.674
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.889	0.890	0.894	0.894	0.896	0.832	0.919	0.931	0.919	0.932	0.960	0.960	0.960	0.960	0.960
5%	0.828	0.832	0.836	0.839	0.838	0.762	0.848	0.873	0.851	0.877	0.952	0.957	0.958	0.957	0.952
1%	0.642	0.660	0.666	0.680	0.667	0.612	0.587	0.723	0.611	0.729	0.944	0.942	0.948	0.944	0.944
	Weakly Identified $n = 100$ $(\beta_n = 0.5/\sqrt{n})$														
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.096	0.101	0.107	0.103	0.101	0.110	0.101	0.131	0.167	0.252	0.074	0.113	0.112	0.124	0.088
5%	0.044	0.043	0.047	0.044	0.044	0.076	0.050	0.093	0.081	0.142	0.037	0.067	0.059	0.078	0.048
1%	0.009	0.012	0.014	0.013	0.011	0.016	0.011	0.028	0.022	0.028	0.008	0.013	0.013	0.018	0.011
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.298	0.314	0.310	0.318	0.312	0.358	0.374	0.359	0.386	0.413	0.359	0.424	0.451	0.453	0.367
5%	0.202	0.216	0.222	0.221	0.212	0.308	0.333	0.310	0.358	0.378	0.267	0.317	0.370	0.350	0.273
1%	0.073	0.083	0.089	0.087	0.087	0.190	0.207	0.190	0.231	0.276	0.159	0.187	0.254	0.222	0.169
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.912	0.914	0.916	0.916	0.919	0.899	0.928	0.919	0.949	0.954	0.782	0.817	0.818	0.831	0.787
5%	0.864	0.873	0.870	0.877	0.874	0.859	0.881	0.876	0.903	0.923	0.723	0.780	0.789	0.798	0.729
1%	0.710	0.734	0.741	0.744	0.727	0.752	0.773	0.762	0.796	0.833	0.637	0.679	0.716	0.703	0.643
	Non-Identified $n = 100$ $(\beta_n = 0)$														
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.116	0.122	0.122	0.127	0.127	0.104	0.097	0.104	0.106	0.212	0.052	0.091	0.094	0.106	0.057
5%	0.063	0.067	0.070	0.072	0.070	0.061	0.058	0.086	0.073	0.108	0.031	0.046	0.049	0.053	0.036
1%	0.016	0.020	0.018	0.020	0.021	0.019	0.009	0.029	0.023	0.029	0.006	0.012	0.007	0.014	0.007
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.299	0.306	0.314	0.314	0.316	0.428	0.434	0.433	0.446	0.493	0.353	0.403	0.453	0.431	0.359
5%	0.198	0.202	0.208	0.209	0.217	0.370	0.361	0.371	0.372	0.438	0.286	0.324	0.374	0.357	0.291
1%	0.076	0.086	0.089	0.088	0.082	0.182	0.173	0.188	0.191	0.271	0.159	0.194	0.263	0.228	0.166
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.911	0.918	0.920	0.919	0.914	0.916	0.918	0.926	0.928	0.952	0.769	0.824	0.823	0.834	0.776
5%	0.864	0.869	0.871	0.874	0.872	0.884	0.887	0.889	0.896	0.926	0.721	0.769	0.788	0.789	0.724
1%	0.709	0.727	0.727	0.733	0.732	0.811	0.794	0.811	0.801	0.838	0.626	0.673	0.704	0.700	0.636

Table B.2: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with t(4) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 100$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

	Strongly Identified $n=100$ $(\beta_n=0.5)$														
	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
	$H_{0,1}:\zeta=\zeta_n$					$H_{0,1}:\beta=\beta_n$					$H_{0,1}:\pi=\pi_n$				
10%	0.094	0.090	0.107	0.114	0.107	0.123	0.117	0.175	0.204	0.183	0.159	0.147	0.187	0.163	0.185
5%	0.050	0.044	0.053	0.062	0.053	0.089	0.085	0.128	0.152	0.132	0.105	0.090	0.125	0.103	0.125
1%	0.009	0.012	0.013	0.020	0.012	0.053	0.046	0.061	0.081	0.064	0.048	0.042	0.061	0.049	0.061
	$H_{0,2}:\zeta=\zeta_n+\sigma_\zeta$					$H_{0,2}:\beta=\beta_n+\sigma_\beta$					$H_{0,2}:\pi=\pi_n+\sigma_\pi$				
10%	0.325	0.321	0.344	0.352	0.343	0.301	0.359	0.364	0.381	0.372	0.873	0.850	0.885	0.860	0.879
5%	0.216	0.219	0.234	0.256	0.233	0.248	0.281	0.289	0.311	0.295	0.836	0.795	0.849	0.811	0.845
1%	0.082	0.083	0.094	0.125	0.095	0.167	0.162	0.176	0.193	0.179	0.722	0.632	0.750	0.654	0.740
	$H_{0,3}:\zeta=\zeta_n+3\sigma_\zeta$					$H_{0,3}:\beta=\beta_n+3\sigma_\beta$					$H_{0,3}:\pi=\pi_n+3\sigma_\pi$				
10%	0.911	0.914	0.917	0.917	0.917	0.910	0.921	0.934	0.926	0.934	0.962	0.962	0.964	0.963	0.963
5%	0.853	0.849	0.861	0.873	0.860	0.878	0.875	0.895	0.879	0.895	0.959	0.955	0.959	0.958	0.959
1%	0.726	0.740	0.742	0.785	0.739	0.781	0.673	0.797	0.691	0.797	0.942	0.928	0.948	0.935	0.946

	Weakly Identified $n=100$ $(\beta_n=0.5/\sqrt{n})$														
	$H_{0,1}:\zeta=\zeta_n$					$H_{0,1}:\beta=\beta_n$					$H_{0,1}:\pi=\pi_n$				
10%	0.113	0.115	0.118	0.121	0.121	0.148	0.142	0.188	0.219	0.335	0.078	0.093	0.109	0.134	0.117
5%	0.066	0.064	0.075	0.074	0.073	0.087	0.087	0.119	0.134	0.187	0.045	0.042	0.062	0.061	0.060
1%	0.021	0.021	0.025	0.026	0.024	0.036	0.044	0.057	0.065	0.060	0.013	0.009	0.024	0.015	0.026
	$H_{0,2}:\zeta=\zeta_n+\sigma_\zeta$					$H_{0,2}:\beta=\beta_n+\sigma_\beta$					$H_{0,2}:\pi=\pi_n+\sigma_\pi$				
10%	0.262	0.266	0.277	0.277	0.273	0.409	0.430	0.413	0.445	0.479	0.388	0.385	0.456	0.464	0.404
5%	0.171	0.177	0.187	0.197	0.180	0.374	0.385	0.375	0.397	0.436	0.303	0.291	0.374	0.380	0.334
1%	0.082	0.086	0.097	0.089	0.086	0.283	0.286	0.283	0.305	0.359	0.175	0.131	0.256	0.221	0.193
	$H_{0,3}:\zeta=\zeta_n+3\sigma_\zeta$					$H_{0,3}:\beta=\beta_n+3\sigma_\beta$					$H_{0,3}:\pi=\pi_n+3\sigma_\pi$				
10%	0.899	0.898	0.909	0.903	0.901	0.934	0.942	0.943	0.948	0.956	0.779	0.807	0.828	0.845	0.801
5%	0.858	0.857	0.868	0.869	0.866	0.914	0.919	0.921	0.926	0.936	0.736	0.746	0.781	0.800	0.760
1%	0.685	0.706	0.717	0.727	0.713	0.854	0.848	0.864	0.870	0.885	0.656	0.641	0.710	0.715	0.678

	Non-Identified $n=100$ $(\beta_n=0)$														
	$H_{0,1}:\zeta=\zeta_n$					$H_{0,1}:\beta=\beta_n$					$H_{0,1}:\pi=\pi_n$				
10%	0.098	0.105	0.106	0.111	0.109	0.139	0.148	0.148	0.176	0.310	0.066	0.075	0.107	0.132	0.097
5%	0.045	0.041	0.053	0.052	0.056	0.091	0.098	0.128	0.128	0.168	0.036	0.032	0.056	0.061	0.053
1%	0.008	0.009	0.009	0.011	0.009	0.026	0.032	0.061	0.070	0.061	0.008	0.004	0.012	0.008	0.015
	$H_{0,2}:\zeta=\zeta_n+\sigma_\zeta$					$H_{0,2}:\beta=\beta_n+\sigma_\beta$					$H_{0,2}:\pi=\pi_n+\sigma_\pi$				
10%	0.302	0.309	0.319	0.322	0.315	0.486	0.474	0.490	0.483	0.528	0.392	0.397	0.464	0.468	0.420
5%	0.211	0.205	0.225	0.224	0.217	0.419	0.409	0.419	0.419	0.494	0.303	0.311	0.385	0.392	0.335
1%	0.068	0.075	0.082	0.091	0.081	0.293	0.258	0.294	0.268	0.350	0.175	0.142	0.262	0.234	0.197
	$H_{0,3}:\zeta=\zeta_n+3\sigma_\zeta$					$H_{0,3}:\beta=\beta_n+3\sigma_\beta$					$H_{0,3}:\pi=\pi_n+3\sigma_\pi$				
10%	0.918	0.917	0.921	0.919	0.921	0.942	0.940	0.946	0.951	0.960	0.809	0.811	0.844	0.861	0.829
5%	0.870	0.873	0.877	0.881	0.878	0.922	0.917	0.925	0.927	0.935	0.764	0.764	0.809	0.821	0.776
1%	0.727	0.738	0.755	0.754	0.740	0.870	0.853	0.878	0.869	0.878	0.660	0.645	0.725	0.728	0.678

Table B.3: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n=\log(n)^{1/2}$. Sample size $n=100$, bootstrapped samples $M_n=500$ simulations $S=1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0=0$.

	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
Strongly Identified $n = 250$ $(\beta_n = 0.5)$															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.105	0.103	0.113	0.106	0.113	0.067	0.071	0.135	0.130	0.135	0.105	0.100	0.107	0.100	0.107
5%	0.049	0.057	0.056	0.059	0.056	0.035	0.039	0.070	0.066	0.070	0.064	0.056	0.064	0.060	0.064
1%	0.008	0.012	0.009	0.013	0.009	0.022	0.022	0.028	0.029	0.028	0.019	0.014	0.019	0.015	0.019
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.261	0.269	0.270	0.275	0.270	0.141	0.289	0.294	0.292	0.294	0.930	0.920	0.932	0.923	0.932
5%	0.164	0.168	0.173	0.176	0.173	0.095	0.203	0.201	0.204	0.201	0.887	0.864	0.887	0.869	0.887
1%	0.052	0.064	0.060	0.070	0.060	0.043	0.079	0.089	0.079	0.089	0.770	0.691	0.770	0.700	0.770
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.904	0.904	0.910	0.910	0.910	0.847	0.945	0.945	0.945	0.945	0.990	0.990	0.990	0.990	0.990
5%	0.840	0.846	0.849	0.855	0.849	0.752	0.903	0.903	0.903	0.903	0.989	0.989	0.989	0.989	0.989
1%	0.668	0.679	0.692	0.697	0.692	0.566	0.704	0.734	0.706	0.734	0.989	0.989	0.989	0.989	0.989
Weakly Identified $n = 250$ $(\beta_n = 0.5/\sqrt{n})$															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.089	0.097	0.096	0.099	0.096	0.124	0.109	0.131	0.175	0.268	0.055	0.079	0.106	0.091	0.056
5%	0.047	0.053	0.051	0.055	0.053	0.054	0.055	0.080	0.099	0.157	0.017	0.036	0.037	0.041	0.017
1%	0.006	0.012	0.008	0.012	0.008	0.009	0.008	0.015	0.017	0.026	0.000	0.004	0.001	0.008	0.000
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.243	0.257	0.257	0.262	0.259	0.362	0.382	0.364	0.387	0.409	0.381	0.446	0.462	0.465	0.383
5%	0.163	0.167	0.172	0.170	0.173	0.323	0.348	0.324	0.369	0.385	0.278	0.341	0.378	0.362	0.282
1%	0.053	0.067	0.066	0.067	0.057	0.202	0.218	0.202	0.255	0.315	0.134	0.176	0.240	0.198	0.134
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.912	0.913	0.917	0.915	0.921	0.913	0.939	0.930	0.957	0.977	0.819	0.851	0.853	0.858	0.819
5%	0.833	0.842	0.845	0.849	0.843	0.862	0.896	0.872	0.915	0.949	0.765	0.815	0.817	0.818	0.767
1%	0.627	0.656	0.660	0.666	0.649	0.758	0.785	0.759	0.802	0.844	0.671	0.726	0.744	0.741	0.673
Non-Identified $n = 250$ $(\beta_n = 0)$															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.096	0.106	0.107	0.107	0.098	0.095	0.104	0.095	0.105	0.246	0.052	0.087	0.111	0.098	0.052
5%	0.048	0.058	0.062	0.059	0.055	0.046	0.057	0.049	0.061	0.121	0.017	0.032	0.047	0.044	0.017
1%	0.009	0.010	0.013	0.010	0.012	0.012	0.015	0.029	0.018	0.029	0.003	0.002	0.005	0.004	0.003
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.269	0.273	0.280	0.276	0.284	0.445	0.444	0.448	0.452	0.503	0.379	0.429	0.462	0.445	0.380
5%	0.169	0.177	0.179	0.178	0.182	0.377	0.373	0.377	0.377	0.457	0.293	0.351	0.392	0.368	0.294
1%	0.056	0.064	0.065	0.067	0.063	0.210	0.204	0.214	0.210	0.284	0.162	0.201	0.274	0.223	0.163
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.914	0.918	0.917	0.918	0.917	0.921	0.935	0.924	0.940	0.968	0.802	0.847	0.848	0.857	0.803
5%	0.846	0.859	0.861	0.861	0.856	0.893	0.896	0.894	0.898	0.938	0.749	0.808	0.818	0.822	0.749
1%	0.649	0.667	0.674	0.673	0.677	0.772	0.782	0.774	0.791	0.841	0.647	0.703	0.724	0.718	0.647

Table B.4: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with Normal(0,1) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
Strongly Identified $n = 250$ $(\beta_n = 0.5)$															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.080	0.083	0.093	0.083	0.094	0.065	0.074	0.116	0.119	0.117	0.114	0.103	0.118	0.105	0.116
5%	0.033	0.035	0.040	0.040	0.040	0.050	0.045	0.068	0.072	0.070	0.068	0.056	0.071	0.056	0.068
1%	0.004	0.010	0.005	0.011	0.005	0.034	0.031	0.041	0.043	0.042	0.019	0.005	0.019	0.006	0.019
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.244	0.251	0.255	0.253	0.256	0.145	0.291	0.294	0.299	0.294	0.916	0.906	0.923	0.908	0.920
5%	0.157	0.168	0.161	0.172	0.161	0.101	0.194	0.202	0.199	0.203	0.873	0.851	0.875	0.856	0.874
1%	0.042	0.050	0.052	0.055	0.052	0.053	0.075	0.086	0.079	0.086	0.768	0.690	0.772	0.703	0.770
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.889	0.890	0.903	0.895	0.903	0.871	0.947	0.949	0.947	0.949	0.981	0.980	0.981	0.980	0.981
5%	0.820	0.819	0.831	0.825	0.831	0.788	0.905	0.910	0.905	0.910	0.979	0.979	0.979	0.980	0.979
1%	0.635	0.652	0.660	0.664	0.660	0.607	0.663	0.752	0.663	0.752	0.978	0.974	0.978	0.974	0.978
Weakly Identified $n = 250$ $(\beta_n = 0.5/\sqrt{n})$															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.098	0.103	0.107	0.106	0.112	0.095	0.101	0.096	0.133	0.241	0.048	0.077	0.095	0.092	0.050
5%	0.047	0.052	0.056	0.055	0.050	0.050	0.050	0.061	0.078	0.128	0.021	0.037	0.047	0.042	0.022
1%	0.007	0.010	0.012	0.010	0.010	0.011	0.012	0.022	0.017	0.030	0.006	0.007	0.009	0.007	0.006
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.251	0.259	0.261	0.265	0.266	0.378	0.398	0.379	0.409	0.427	0.350	0.413	0.443	0.433	0.351
5%	0.161	0.173	0.175	0.175	0.178	0.325	0.349	0.326	0.370	0.398	0.264	0.321	0.363	0.339	0.265
1%	0.067	0.078	0.078	0.079	0.073	0.203	0.209	0.204	0.245	0.299	0.144	0.184	0.231	0.196	0.148
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.889	0.895	0.896	0.896	0.897	0.927	0.945	0.938	0.956	0.981	0.795	0.841	0.846	0.849	0.798
5%	0.825	0.829	0.829	0.831	0.833	0.882	0.907	0.886	0.927	0.945	0.734	0.787	0.801	0.803	0.734
1%	0.663	0.690	0.694	0.696	0.695	0.761	0.781	0.763	0.806	0.858	0.630	0.692	0.705	0.713	0.631
Non-Identified $n = 250$ $(\beta_n = 0)$															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.105	0.113	0.114	0.114	0.111	0.096	0.098	0.096	0.104	0.212	0.046	0.071	0.084	0.082	0.050
5%	0.050	0.058	0.060	0.061	0.053	0.059	0.056	0.060	0.060	0.112	0.019	0.024	0.039	0.028	0.019
1%	0.014	0.016	0.017	0.018	0.016	0.013	0.014	0.032	0.020	0.032	0.000	0.002	0.002	0.004	0.000
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.240	0.259	0.259	0.261	0.260	0.426	0.433	0.427	0.444	0.492	0.347	0.397	0.425	0.428	0.349
5%	0.140	0.154	0.153	0.157	0.154	0.355	0.348	0.358	0.355	0.444	0.266	0.316	0.366	0.342	0.270
1%	0.055	0.064	0.062	0.064	0.061	0.175	0.169	0.176	0.183	0.249	0.156	0.189	0.240	0.215	0.158
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.897	0.901	0.903	0.902	0.906	0.929	0.935	0.930	0.940	0.964	0.771	0.805	0.816	0.814	0.771
5%	0.845	0.851	0.851	0.852	0.856	0.889	0.891	0.890	0.893	0.938	0.705	0.759	0.775	0.770	0.706
1%	0.674	0.706	0.705	0.709	0.697	0.782	0.796	0.786	0.802	0.839	0.609	0.662	0.694	0.682	0.609

Table B.5: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with t(4) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

	LF_ζ AC	LF_ζ PB	$ICS_{0,\zeta}$ AC	$ICS_{0,\zeta}$ PB	Std	LF_β AC	LF_β PB	$ICS_{0,\beta}$ AC	$ICS_{0,\beta}$ PB	Std	LF_π AC	LF_π PB	$ICS_{0,\pi}$ AC	$ICS_{0,\pi}$ PB	Std
Strongly Identified $n = 250$ ($\beta_n = 0.5$)															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.095	0.098	0.107	0.114	0.108	0.108	0.100	0.171	0.176	0.174	0.126	0.125	0.140	0.137	0.138
5%	0.054	0.051	0.057	0.061	0.057	0.069	0.069	0.109	0.122	0.111	0.075	0.067	0.084	0.076	0.083
1%	0.009	0.013	0.010	0.024	0.010	0.043	0.038	0.048	0.060	0.049	0.026	0.019	0.036	0.025	0.036
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.293	0.292	0.307	0.308	0.307	0.277	0.383	0.388	0.392	0.391	0.935	0.928	0.940	0.930	0.942
5%	0.195	0.192	0.209	0.224	0.209	0.226	0.295	0.298	0.301	0.299	0.918	0.895	0.924	0.899	0.922
1%	0.060	0.063	0.068	0.095	0.068	0.119	0.156	0.158	0.163	0.160	0.853	0.797	0.856	0.811	0.855
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.912	0.909	0.918	0.913	0.919	0.926	0.932	0.942	0.936	0.944	0.973	0.973	0.973	0.973	0.974
5%	0.854	0.859	0.866	0.866	0.867	0.895	0.897	0.913	0.903	0.912	0.972	0.969	0.973	0.972	0.972
1%	0.687	0.683	0.712	0.723	0.712	0.809	0.769	0.829	0.775	0.829	0.965	0.961	0.969	0.966	0.966
Weakly Identified $n = 250$ ($\beta_n = 0.5/\sqrt{n}$)															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.090	0.095	0.098	0.100	0.100	0.141	0.147	0.153	0.204	0.306	0.071	0.089	0.115	0.121	0.089
5%	0.040	0.042	0.048	0.048	0.046	0.084	0.077	0.105	0.130	0.191	0.034	0.035	0.059	0.061	0.042
1%	0.011	0.010	0.012	0.011	0.013	0.019	0.022	0.039	0.046	0.049	0.008	0.004	0.011	0.010	0.014
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.252	0.256	0.264	0.264	0.264	0.389	0.412	0.395	0.417	0.463	0.371	0.376	0.437	0.445	0.381
5%	0.156	0.165	0.171	0.171	0.175	0.349	0.368	0.349	0.373	0.415	0.291	0.290	0.354	0.355	0.303
1%	0.054	0.064	0.066	0.068	0.067	0.265	0.277	0.265	0.293	0.334	0.165	0.135	0.240	0.203	0.173
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.895	0.889	0.900	0.899	0.900	0.937	0.943	0.942	0.947	0.961	0.809	0.828	0.842	0.855	0.825
5%	0.819	0.817	0.833	0.826	0.830	0.918	0.923	0.924	0.931	0.940	0.771	0.774	0.815	0.818	0.779
1%	0.618	0.641	0.667	0.661	0.644	0.856	0.856	0.864	0.882	0.892	0.671	0.676	0.728	0.739	0.678
Non-Identified $n = 250$ ($\beta_n = 0$)															
	$H_{0,1}: \zeta = \zeta_n$					$H_{0,1}: \beta = \beta_n$					$H_{0,1}: \pi = \pi_n$				
10%	0.096	0.097	0.103	0.102	0.105	0.124	0.130	0.124	0.138	0.274	0.057	0.081	0.100	0.123	0.076
5%	0.043	0.046	0.047	0.049	0.054	0.069	0.067	0.082	0.087	0.161	0.024	0.030	0.048	0.047	0.037
1%	0.010	0.011	0.013	0.011	0.011	0.014	0.017	0.040	0.046	0.040	0.004	0.002	0.004	0.006	0.006
	$H_{0,2}: \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2}: \beta = \beta_n + \sigma_\beta$					$H_{0,2}: \pi = \pi_n + \sigma_\pi$				
10%	0.285	0.285	0.299	0.295	0.293	0.446	0.446	0.454	0.449	0.500	0.383	0.394	0.458	0.465	0.394
5%	0.187	0.185	0.200	0.198	0.201	0.394	0.402	0.395	0.405	0.459	0.302	0.304	0.386	0.378	0.308
1%	0.068	0.077	0.083	0.081	0.075	0.270	0.245	0.271	0.255	0.339	0.189	0.175	0.268	0.235	0.198
	$H_{0,3}: \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3}: \beta = \beta_n + 3\sigma_\beta$					$H_{0,3}: \pi = \pi_n + 3\sigma_\pi$				
10%	0.913	0.913	0.920	0.919	0.918	0.933	0.930	0.935	0.936	0.957	0.807	0.821	0.839	0.848	0.815
5%	0.856	0.854	0.872	0.867	0.862	0.913	0.913	0.917	0.917	0.932	0.759	0.779	0.807	0.818	0.766
1%	0.667	0.677	0.698	0.693	0.696	0.859	0.830	0.865	0.841	0.873	0.657	0.661	0.724	0.738	0.668

Table B.6: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with GARCH(0.1,0.3,0.6) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 250$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

	LF_ζ		$ICS_{0,\zeta}$			LF_β		$ICS_{0,\beta}$			LF_π		$ICS_{0,\pi}$		
	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std	AC	PB	AC	PB	Std
Strongly Identified $n = 500$ ($\beta_n = 0.5$)															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.106	0.107	0.113	0.111	0.113	0.045	0.056	0.094	0.097	0.094	0.091	0.092	0.095	0.093	0.095
5%	0.051	0.049	0.053	0.054	0.053	0.026	0.029	0.052	0.051	0.052	0.048	0.045	0.048	0.045	0.048
1%	0.013	0.014	0.017	0.016	0.017	0.009	0.013	0.011	0.015	0.011	0.009	0.006	0.009	0.006	0.009
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.260	0.263	0.274	0.271	0.274	0.120	0.279	0.269	0.282	0.269	0.790	0.783	0.792	0.786	0.792
5%	0.167	0.178	0.178	0.183	0.178	0.072	0.182	0.176	0.184	0.176	0.699	0.691	0.701	0.695	0.701
1%	0.056	0.063	0.066	0.064	0.066	0.030	0.070	0.057	0.071	0.057	0.515	0.479	0.520	0.485	0.520
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.901	0.902	0.907	0.906	0.907	0.818	0.924	0.920	0.924	0.920	0.998	0.998	0.998	0.998	0.998
5%	0.840	0.844	0.848	0.852	0.848	0.717	0.875	0.871	0.875	0.871	0.997	0.995	0.997	0.995	0.997
1%	0.659	0.675	0.676	0.687	0.676	0.526	0.694	0.685	0.697	0.685	0.995	0.994	0.995	0.994	0.995
Weakly Identified $n = 500$ ($\beta_n = 0.5/\sqrt{n}$)															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.102	0.106	0.118	0.110	0.110	0.078	0.072	0.078	0.091	0.197	0.034	0.074	0.082	0.080	0.037
5%	0.049	0.056	0.057	0.057	0.055	0.045	0.041	0.049	0.061	0.097	0.016	0.024	0.040	0.031	0.016
1%	0.011	0.014	0.014	0.015	0.016	0.015	0.014	0.023	0.018	0.030	0.001	0.000	0.003	0.001	0.001
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.280	0.287	0.299	0.291	0.291	0.366	0.382	0.369	0.393	0.425	0.360	0.432	0.438	0.444	0.363
5%	0.170	0.186	0.190	0.189	0.183	0.297	0.334	0.299	0.355	0.385	0.256	0.322	0.372	0.351	0.257
1%	0.047	0.054	0.057	0.054	0.057	0.147	0.161	0.148	0.203	0.262	0.123	0.166	0.231	0.180	0.123
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.905	0.908	0.909	0.908	0.906	0.915	0.932	0.917	0.943	0.964	0.779	0.825	0.824	0.830	0.779
5%	0.860	0.863	0.863	0.866	0.869	0.868	0.902	0.869	0.916	0.938	0.728	0.771	0.782	0.782	0.729
1%	0.662	0.683	0.689	0.687	0.694	0.726	0.751	0.728	0.774	0.848	0.639	0.699	0.703	0.713	0.641
Non-Identified $n = 500$ ($\beta_n = 0$)															
	$H_{0,1} : \zeta = \zeta_n$					$H_{0,1} : \beta = \beta_n$					$H_{0,1} : \pi = \pi_n$				
10%	0.086	0.097	0.099	0.098	0.092	0.107	0.110	0.107	0.111	0.224	0.048	0.082	0.097	0.095	0.049
5%	0.044	0.047	0.049	0.049	0.046	0.051	0.051	0.051	0.055	0.123	0.016	0.031	0.046	0.037	0.017
1%	0.007	0.006	0.008	0.006	0.008	0.017	0.011	0.023	0.016	0.023	0.001	0.003	0.006	0.006	0.001
	$H_{0,2} : \zeta = \zeta_n + \sigma_\zeta$					$H_{0,2} : \beta = \beta_n + \sigma_\beta$					$H_{0,2} : \pi = \pi_n + \sigma_\pi$				
10%	0.244	0.257	0.256	0.262	0.257	0.402	0.405	0.408	0.409	0.454	0.369	0.421	0.447	0.430	0.371
5%	0.152	0.166	0.170	0.169	0.162	0.330	0.337	0.332	0.344	0.416	0.270	0.334	0.374	0.346	0.270
1%	0.044	0.053	0.054	0.053	0.054	0.172	0.177	0.172	0.182	0.247	0.137	0.177	0.241	0.200	0.137
	$H_{0,3} : \zeta = \zeta_n + 3\sigma_\zeta$					$H_{0,3} : \beta = \beta_n + 3\sigma_\beta$					$H_{0,3} : \pi = \pi_n + 3\sigma_\pi$				
10%	0.902	0.907	0.907	0.907	0.908	0.924	0.929	0.928	0.937	0.969	0.783	0.847	0.847	0.856	0.785
5%	0.822	0.829	0.826	0.830	0.834	0.890	0.889	0.890	0.891	0.939	0.739	0.782	0.803	0.798	0.739
1%	0.640	0.672	0.675	0.676	0.661	0.762	0.764	0.764	0.778	0.828	0.633	0.695	0.716	0.710	0.633

Table B.7: Size and power of t-statistic under Strong, Weak and Non-Identification for the Smoothing Exponential model with t(4) errors. Numerical values are rejection probabilities at the given level. AC, PB and Std denote Andrews and Cheng, Parametric Bootstrapped and Standard (strong identification) t-statistic rejection probabilities respectively. The AC critical values are infeasible because they assume we know the nuisance parameters. LF denotes Least Favourable critical value and ICS_0 denotes Type 1 identification category with $\kappa_n = \log(n)^{1/2}$. Sample size $n = 500$, bootstrapped samples $M_n = 500$ simulations $S = 1000$. The grid $\tilde{\Pi}$ is constructed using 501 equally spaced values inside the interval [-5,5] with $\pi_0 = 0$.

APPENDIX C

SUPPLEMENTAL APPENDIX OF "THE RISK RETURN TRADE-OFF UNDER WEAK IDENTIFICATION"

C.1 Risk-return QML estimation

In this section, we provide the expressions used to estimate the values and variance of the parameters using QML. Let $V_t(\pi)$ be the midas estimator of the conditional variance and $D = 252$ the number of daily observations which roughly represents the number of trading days in a year.

$$R_{t+1} \sim N(\zeta + \beta V_t(\pi), V_t(\pi)) \tag{C.1}$$

$$w(d, \pi_1, \pi_2) = \frac{\exp(\pi_1 d + \pi_2 d^2)}{\sum_{k=0}^{D-1} \exp(\pi_1 k + \pi_2 k^2)} \tag{C.2}$$

$$V_t(\pi) = A \sum_{d=0}^{D-1} w(d, \pi_1, \pi_2) r_{t-d}^2 \tag{C.3}$$

Define $\epsilon_t(\theta) = R_{t+1} - \zeta - \beta V_t(\pi)$, where $\theta = (\zeta, \beta, \pi)'$. From Equation (C.1) we derive the log-likelihood objective function to be maximized. Let n be the number of monthly observations.

$$Q_n(\theta) = n^{-1} \sum_{t=1}^{n} Q_t(\theta) = -\frac{1}{2n} \sum_{t=1}^{n} \left[\log(V_t(\pi) + \frac{\epsilon_t(\theta)^2}{V_t(\pi)} \right] \tag{C.4}$$

The gradient of the likelihood function $\frac{\partial Q_n(\theta)}{\partial \theta}$ takes the following form.

$$[\zeta]: \quad n^{-1} \sum_{t=1}^{n} \frac{\epsilon_t(\theta)}{V_t(\pi)} \tag{C.5}$$

$$[\beta]: \quad n^{-1} \sum_{t=1}^{n} \epsilon_t(\theta) \tag{C.6}$$

$$[\pi]: \quad -\frac{1}{2} n^{-1} \sum_{t=1}^{n} \left[V_t(\pi)^{-1} - 2\beta \frac{\epsilon_t(\theta)}{V_t(\pi)} - \frac{\epsilon_t(\theta)^2}{V_t(\pi)^2} \right] \frac{\partial V_t(\pi)}{\partial \pi} \tag{C.7}$$

where the previous expressions take the form

$$\frac{\partial V_t(\pi)}{\partial \pi} = A \sum_{d=0}^{D-1} \frac{\partial w(d,\pi_1,\pi_2)}{\partial \pi} r_{t-d}^2 \tag{C.8}$$

$$\frac{\partial w(d,\pi_1,\pi_2)}{\partial \pi} = \frac{Num_d(\pi)}{Den(\pi)} \begin{bmatrix} d \\ d^2 \end{bmatrix} - \frac{Num_d(\pi)}{Den(\pi)^2} \frac{\partial Den(\pi)}{\partial \pi} \tag{C.9}$$

$$Num_d(\pi) = \exp(\pi_1 d + \pi_2 d^2) \tag{C.10}$$

$$Den(\pi) = \sum_{k=0}^{D-1} \exp(\pi_1 k + \pi_2 k^2) \tag{C.11}$$

$$\frac{\partial Den(\pi)}{\partial \pi} = \sum_{k=0}^{D-1} \exp(\pi_1 k + \pi_2 k^2) \begin{bmatrix} k \\ k^2 \end{bmatrix} \tag{C.12}$$

$$\frac{\partial Num(\pi)}{\partial \pi} = \exp(\pi_1 d + \pi_2 d^2) \begin{bmatrix} d \\ d^2 \end{bmatrix} \tag{C.13}$$

To obtain the variance of the parameters, we require the Hessian matrix of the likelihood function $\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'}$. The second derivatives of the log likelihood are as follows.

$$[\zeta,\zeta]: \quad n^{-1} \sum_{t=1}^{n} -V_t(\pi)^{-1} \tag{C.14}$$

$$[\zeta,\beta]: \quad -1 \tag{C.15}$$

$$[\zeta,\pi]: \quad n^{-1} \sum_{t=1}^{n} -\left[\beta V_t(\pi)^{-1} + \epsilon_t(\theta) V_t(\pi)^{-2}\right] \frac{\partial V_t(\pi)}{\partial \pi'} \tag{C.16}$$

$$[\beta,\beta]: \quad n^{-1} \sum_{t=1}^{n} -V_t(\pi) \tag{C.17}$$

$$[\beta,\pi]: \quad n^{-1} \sum_{t=1}^{n} -\beta \frac{\partial V_t(\pi)}{\partial \pi'} \tag{C.18}$$

$$\begin{aligned}[\pi,\pi]: \frac{-1}{2} n^{-1} \sum_{t=1}^{n} \Bigg\{ &\left[V_t(\pi)^{-1} - 2\beta \frac{\epsilon_t(\theta)}{V_t(\pi)} - \frac{\epsilon_t(\theta)^2}{V_t(\pi)^2}\right] \frac{\partial^2 V_t(\pi)}{\partial\pi\partial\pi'} + \\ &\left[- V_t(\pi)^{-2} \frac{\partial V_t(\pi)}{\partial \pi} - 2\beta \frac{\partial}{\partial \pi} \frac{\epsilon_t(\theta)}{V_t(\pi)} - 2 \frac{\epsilon_t(\theta)}{V_t(\pi)} \frac{\partial}{\partial \pi} \frac{\epsilon_t(\theta)}{V_t(\pi)}\right] \frac{\partial V_t(\pi)}{\partial \pi'} \Bigg\}\end{aligned} \tag{C.19}$$

where the previous expressions take the form,

$$\frac{\partial}{\partial \pi}\frac{\epsilon_t(\theta)}{V_t(\pi)} = -V_t(\pi)^{-1}\beta\frac{\partial V_t(\pi)}{\partial \pi} - \epsilon_t(\theta)V_t(\pi)^{-2}\frac{\partial V_t(\pi)}{\partial \pi} \tag{C.20}$$

$$\frac{\partial^2 V_t(\pi)}{\partial \pi \partial \pi'} = \sum_{d=0}^{D-1}\frac{\partial^2 w(d,\pi_1,\pi_2)}{\partial \pi \partial \pi'}r_{t-d}^2 \tag{C.21}$$

$$\begin{aligned}\frac{\partial^2 w(d,\pi_1,\pi_2)}{\partial \pi \partial \pi'} &= \begin{bmatrix} d \\ d^2 \end{bmatrix} \frac{\partial}{\partial \pi'}\frac{Num_d(\pi)}{Den(\pi)} - \frac{\partial Den(\pi)}{\partial \pi}\frac{\partial}{\partial \pi'}\frac{Num_d(\pi)}{Den(\pi)^2} \\ &- \frac{Num_d(\pi)}{Den(\pi)^2}\sum_{k=0}^{D-1}\exp(\pi_1 k + \pi_2 k^2)\begin{bmatrix} k^2 & k^3 \\ k^3 & k^4 \end{bmatrix}\end{aligned} \tag{C.22}$$

$$\frac{\partial}{\partial \pi}\frac{Num_d(\pi)}{Den(\pi)} = Den^{-1}(\pi)\frac{\partial Num_d(\pi)}{\partial \pi} - Den(\pi)^{-2}Num_d(\pi)\frac{\partial Den(\pi)}{\partial \pi} \tag{C.23}$$

$$\frac{\partial}{\partial \pi}\frac{Num_d(\pi)}{Den(\pi)^2} = Den^{-2}(\pi)\frac{\partial Num_d(\pi)}{\partial \pi} - 2Den(\pi)^{-3}Num_d(\pi)\frac{\partial Den(\pi)}{\partial \pi} \tag{C.24}$$

By Wooldridge and White (1988), the QMLE converges in probability to the pseudo-true value θ^* and its asymptotic distribution takes the following form. The estimator is consistent with respect to the pseudo-true value because the log-likelihood is misspecified, and therefore it could be inconsistent with respect to the true value θ_0.

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} N(0, \tilde{J}^{-1}(\theta^*)\tilde{V}(\theta^*)\tilde{J}^{-1}(\theta^*)) \tag{C.25}$$

$$\tilde{J}(\theta^*) = n^{-1}\sum_{t=1}^{n}\frac{\partial Q_t(\theta^*)}{\partial \theta}\frac{\partial Q_t(\theta^*)}{\partial \theta'} \tag{C.26}$$

$$\tilde{V}(\theta^*) = n^{-1}\sum_{t=1}^{n}\frac{\partial^2 Q_t(\theta^*)}{\partial \theta \partial \theta'} \tag{C.27}$$

C.2 Supplemental Tables and Figures

Table C.1: Risk Return Trade-off wild bootstrapped t-test, absolute returns at monthly frequency.

Monthly		ζ $(\times 10^2)$	β	π_1 $(\times 10^2)$	π_2 $(\times 10^4)$	R^2
$1928-2000$	Coeff.	0.541	0.463	-4.289	1.551	0.000
	Std. Dev.	0.476	3.753	2.057	0.776	
	Std. p-value	0.000	0.902	0.030	0.000	
	WB p-value	0.000	0.960	0.090	0.412	
$1928-1963$	Coeff.	1.232	-2.791	-2.953	1.189	-0.003
	Std. Dev.	0.700	5.200	3.571	1.341	
	Std. p-value	0.000	0.591	0.388	0.000	
	WB p-value	0.000	0.746	0.404	0.200	
$1964-2000$	Coeff.	0.061	2.989	2.925	-19.859	0.003
	Std. Dev.	0.524	4.185	16.084	29.348	
	Std. p-value	0.731	0.475	0.842	0.001	
	WB p-value	0.722	0.452	0.850	0.000	
$1928-2016$	Coeff.	0.542	0.293	-5.270	1.846	0.000
	Std. Dev.	0.387	2.973	1.729	0.683	
	Std. p-value	0.000	0.921	0.002	0.000	
	WB p-value	0.000	0.978	0.138	0.354	
$1928-2000^{FTS}$	Coeff.	0.380	2.745	-6.063	2.029	-0.003
	Std. Dev.	0.326	2.374	1.722	0.753	
	Std. p-value	0.004	0.248	0.000	0.000	
	WB p-value	0.022	0.486	0.016	0.046	
$1964-2000^{FTS}$	Coeff.	-0.617	8.594	3.220	-10.933	0.023
	Std. Dev.	0.536	3.841	13.298	22.221	
	Std. p-value	0.001	0.025	0.804	0.035	
	WB p-value	0.010	0.118	0.798	0.026	

This table presents the estimates, standard deviation and p-values of the standard and wild bootstrapped t-test of the MIDAS model using monthly frequency and absolute returns. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

Table C.2: Risk Return Trade-off wild bootstrapped t-test, absolute returns at quarterly frequency.

Quarterly		ζ ($\times 10^2$)	β	π_1 ($\times 10^2$)	π_2 ($\times 10^4$)	R^2
1928 – 2000	Coeff.	0.906	2.003	-5.454	1.975	0.003
	Std. Dev.	1.629	4.326	7.593	2.972	
	Std. p-value	0.049	0.643	0.466	0.000	
	WB p-value	0.118	0.780	0.486	0.072	
1928 – 1963	Coeff.	3.796	-2.949	68.589	-22.629	-0.014
	Std. Dev.	2.237	5.600	514.754	170.316	
	Std. p-value	0.000	0.595	0.893	0.000	
	WB p-value	0.000	0.724	0.892	0.000	
1964 – 2000	Coeff.	-1.312	6.690	10.428	-28.412	0.040
	Std. Dev.	1.267	3.562	22.398	43.001	
	Std. p-value	0.012	0.060	0.628	0.019	
	WB p-value	0.056	0.128	0.634	0.026	
1928 – 2016	Coeff.	1.026	1.546	-6.287	2.245	0.002
	Std. Dev.	1.310	3.391	5.917	2.371	
	Std. p-value	0.012	0.648	0.284	0.000	
	WB p-value	0.032	0.786	0.278	0.090	
$1928-2000^{FTS}$	Coeff.	3.839	-3.955	180.311	-48.788	0.038
	Std. Dev.	0.865	2.077	63.795	17.077	
	Std. p-value	0.000	0.056	0.003	0.000	
	WB p-value	0.000	0.142	0.004	0.000	
$1964-2000^{FTS}$	Coeff.	-2.285	10.589	-0.354	-2.612	0.057
	Std. Dev.	1.602	4.175	4.761	3.274	
	Std. p-value	0.000	0.011	0.941	0.019	
	WB p-value	0.018	0.138	0.950	0.016	

This table presents the estimates, standard deviation and p-values of the standard and wild bootstrapped t-test of the MIDAS model using quarterly frequency and absolute returns. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

Table C.3: Risk Return Trade-off parametric bootstrapped t-test, monthly frequency.

Monthly		ζ $(\times 10^2)$	β	π_1 $(\times 10^2)$	π_2 $(\times 10^4)$	R^2
$1928-2000$	Coeff.	0.537	0.482	-4.258	1.539	0.000
	Std. Dev.	0.476	3.754	2.053	0.773	
	Std. p-value	0.000	0.898	0.031	0.000	
	PB p-value	0.000	0.912	0.676	0.232	
$1928-1963$	Coeff.	1.212	-2.646	-2.820	1.138	-0.003
	Std. Dev.	0.700	5.197	3.615	1.352	
	Std. p-value	0.000	0.611	0.414	0.000	
	PB p-value	0.000	0.634	0.508	0.024	
$1964-2000$	Coeff.	0.022	3.314	3.051	-19.843	0.003
	Std. Dev.	0.530	4.223	16.334	29.638	
	Std. p-value	0.903	0.432	0.836	0.001	
	PB p-value	0.854	0.404	0.804	0.030	
$1928-2016$	Coeff.	0.544	0.292	-5.165	1.803	0.000
	Std. Dev.	0.388	2.979	1.706	0.673	
	Std. p-value	0.000	0.922	0.002	0.000	
	PB p-value	0.000	0.914	0.654	0.334	
$1928-2000^{FTS}$	Coeff.	0.374	2.795	-6.080	2.036	-0.003
	Std. Dev.	0.326	2.374	1.725	0.754	
	Std. p-value	0.004	0.239	0.000	0.000	
	PB p-value	0.006	0.218	0.096	0.006	
$1964-2000^{FTS}$	Coeff.	-0.619	8.608	3.538	-11.463	0.023
	Std. Dev.	0.536	3.839	13.861	23.209	
	Std. p-value	0.001	0.025	0.793	0.032	
	PB p-value	0.000	0.000	0.760	0.040	

This table presents the estimates, standard deviation and p-values of the standard and parametric bootstrapped t-test of the MIDAS model using monthly frequency and absolute returns. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

Table C.4: Risk Return Trade-off parametric bootstrapped t-test, absolute returns at quarterly frequency.

Quarterly		ζ ($\times 10^2$)	β	π_1 ($\times 10^2$)	π_2 ($\times 10^4$)	R^2
$1928-2000$	Coeff.	0.900	2.021	-5.362	1.938	0.003
	Std. Dev.	1.643	4.356	7.597	2.972	
	Std. p-value	0.051	0.643	0.471	0.000	
	PB p-value	0.046	0.676	0.636	0.020	
$1928-1963$	Coeff.	3.749	-2.831	128.824	-42.667	-0.013
	Std. Dev.	2.034	5.195	1033.203	343.220	
	Std. p-value	0.000	0.586	0.901	0.000	
	PB p-value	0.000	0.572	0.920	0.012	
$1964-2000$	Coeff.	-1.309	6.681	10.401	-28.356	0.040
	Std. Dev.	1.266	3.562	22.346	42.879	
	Std. p-value	0.012	0.061	0.628	0.019	
	PB p-value	0.012	0.002	0.612	0.054	
$1928-2016$	Coeff.	1.024	1.549	-6.314	2.256	0.002
	Std. Dev.	1.310	3.390	5.929	2.376	
	Std. p-value	0.012	0.648	0.283	0.000	
	PB p-value	0.010	0.650	0.538	0.022	
$1928-2000^{FTS}$	Coeff.	3.533	-3.272	1.327	-0.454	0.025
	Std. Dev.	1.787	4.097	9.673	3.223	
	Std. p-value	0.000	0.424	0.796	0.125	
	PB p-value	0.000	0.404	0.796	0.280	
$1964-2000^{FTS}$	Coeff.	-2.292	10.605	-0.312	-2.648	0.057
	Std. Dev.	1.601	4.175	4.763	3.279	
	Std. p-value	0.000	0.011	0.948	0.018	
	PB p-value	0.020	0.006	0.888	0.082	

This table presents the estimates, standard deviation and p-values of the standard and parametric bootstrapped t-test of the MIDAS model using quarterly frequency and absolute returns. The conditional variance estimator of returns is calculated using daily returns as in Equation (3.2). The variance of the coefficients is obtained using the sandwich formula of the QML estimator White (1982). R^2 is the coefficient of determination. The coefficients and standard deviation are multiplied by the value in the second row. FTS denotes Flight-to-Safety subsamples.

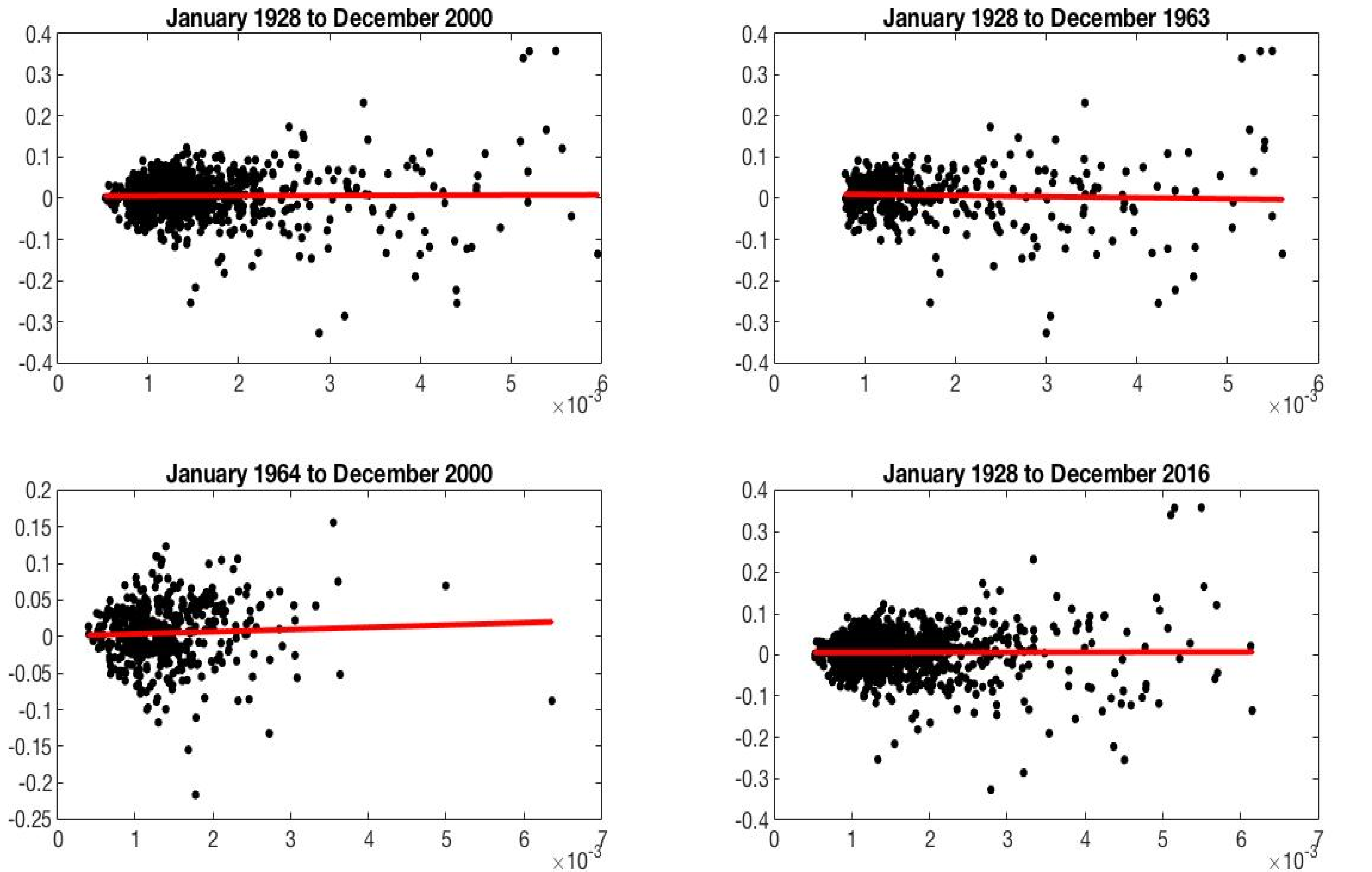

Figure C.1: Scatter plot and regression line of MIDAS estimation using absolute returns, monthly frequency. The figure plots the estimated conditional variance using MIDAS and the return at a monthly frequency for four sample periods. The line is constructed with the estimators obtained from the QML estimation of Equation (3.5) presented in Table 3.5.

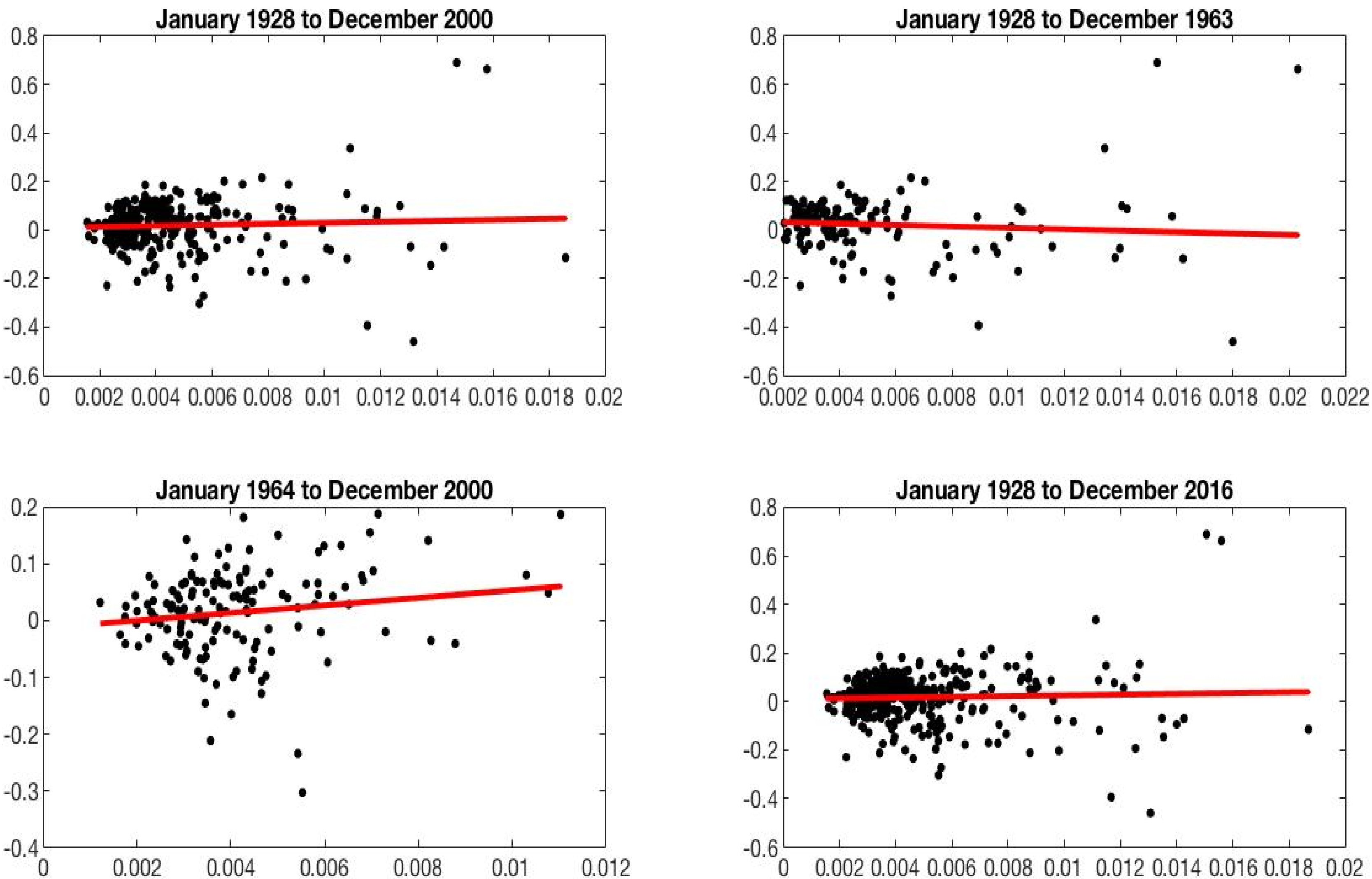

Figure C.2: Scatter plot and regression line of MIDAS estimation using absolute returns, quarterly frequency. The figure plots the estimated conditional variance using MIDAS and the return at a quarterly frequency for four sample periods. The line is constructed with the estimators obtained from the QML estimation of Equation (3.5) presented in Table 3.6.

REFERENCES

Andrews, D. W. (1994). Empirical process methods in econometrics. *Handbook of econometrics 4*, 2247–2294.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica 67*(6), 1341–1383.

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*(3), 683–734.

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Andrews, D. W. and X. Cheng (2013). Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics 173*(1), 36–56.

Andrews, D. W. and X. Cheng (2014). Gmm estimation and uniform subvector inference with possible identification failure. *Econometric Theory 30*(2), 287–333.

Andrews, D. W. and P. Guggenberger (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory 26*(2), 426–468.

Andrews, D. W. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, 1383–1414.

Andrews, I. and A. Mikusheva (2015). Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics 6*(1), 123–152.

Ansley, C. F. and P. Newbold (1980). Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics 13*(2), 159–183.

Antoine, B. and E. Renault (2009). Efficient gmm with nearly-weak instruments. *The Econometrics Journal 12*(s1).

Antoine, B. and E. Renault (2012). Efficient inference with poor instruments: a general framework.

Arrow, K. J. K. J. (1970). Essays in the theory of risk-bearing. Technical report.

Baillie, R. T. and R. P. DeGennaro (1990). Stock returns and volatility. *Journal of financial and Quantitative Analysis 25*(2), 203–214.

Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance 59*(4), 1481–1509.

Beran, R. (1986). Discussion: Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics 14*(4), 1295–1298.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica: Journal of the Econometric Society*, 1443–1458.

Bierens, H. J. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica: Journal of the Econometric Society*, 1129–1151.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of financial economics 18*(2), 373–399.

Caner, M. (2009). Testing, estimation in gmm and cue with nearly-weak identification. *Econometric Reviews 29*(3), 330–363.

Chen, H., Y. Fan, and R. Liu (2016). Inference for the correlation coefficient between potential outcomes in the gaussian switching regime model. *Journal of Econometrics 195*(2), 255–270.

Cheng, X. (2015). Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics 189*(1), 207–228.

Chernoff, H. and S. Zacks (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics 35*(3), 999–1018.

Davidson, J., A. Monticini, and D. Peel (2007). Implementing the wild bootstrap using a two-point distribution. *Economics Letters 96*(3), 309–315.

Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. *Journal of Econometrics 146*(1), 162–169.

Davidson, R. and J. G. MacKinnon (2014). Bootstrap confidence sets with weak instruments. *Econometric Reviews 33*(5-6), 651–675.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*(2), 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*(1), 33–43.

Dudley, R. M. (1978). Central limit theorems for empirical measures. *The Annals of Probability*, 899–929.

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica 73*(4), 1351–1365.

Efron, B. et al. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics 7*(1), 1–26.

Epstein, L. G. and S. E. Zin (2013). Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. In *HANDBOOK OF THE*

FUNDAMENTALS OF FINANCIAL DECISION MAKING: Part I, pp. 207–239. World Scientific.

Farber, H. S. (1978). Individual preferences and union wage determination: the case of the united mine workers. *Journal of Political Economy 86*(5), 923–942.

Forsberg, L. and E. Ghysels (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics 5*(1), 31–67.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 1218–1228.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987). Expected stock returns and volatility. *Journal of financial Economics 19*(1), 3–29.

Friend, I. and M. E. Blume (1975). The demand for risky assets. *The American Economic Review 65*(5), 900–922.

Ghysels, E., J. B. Hill, and K. Motegi (2016). Testing for granger causality with mixed frequency data. *Journal of Econometrics 192*(1), 207–230.

Ghysels, E., A. Plazzi, and R. I. Valkanov (2016). The risk-return relationship and financial crises.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The midas touch: Mixed data sampling regression models. *Finance*.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. *Journal of Financial Economics 76*(3), 509–548.

Giné, E. and J. Zinn (1990). Bootstrapping general empirical measures. *The Annals of Probability*, 851–869.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance 48*(5), 1779–1801.

Granger, C. W. and P. Newbold (1974). Spurious regressions in econometrics. *Journal of econometrics 2*(2), 111–120.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013). Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics 4*(2), 197–229.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, 413–430.

Härdle, W., J. Horowitz, and J.-P. Kreiss (2003). Bootstrap methods for time series. *International Statistical Review 71*(2), 435–459.

Hill, J. B. (2008). Consistent and non-degenerate model specification tests against smooth transition and neural network alternatives. *Annales d'Économie et de Statistique*, 145–179.

Hill, J. B. (2013). Stochastically weighted average conditional moment tests of functional form. *Studies in Nonlinear Dynamics and Econometrics 17*(2), 121–139.

Horowitz, J. L. (2001). The bootstrap. *Handbook of econometrics 5*, 3159–3228.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*(5), 1781–1803.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica 73*(4), 1103–1123.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, 1217–1241.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Lettau, M. and S. Ludvigson (2001a). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance 56*(3), 815–849.

Lettau, M. and S. C. Ludvigson (2001b). Measuring and modelling variation in the risk-return trade-off.

Liu, R. Y. et al. (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics 16*(4), 1696–1708.

Ludvigson, S. C. and S. Ng (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

Ma, J. and C. R. Nelson (2008). Valid inference for a class of models where standard inference performs poorly; including nonlinear regression, arma, garch, and unobserved components.

MacKinnon, J. G. (2002). Bootstrap inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique 35*(4), 615–645.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record 82*(s1).

MacKinnon, J. G. and H. White (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics 29*(3), 305–325.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The* *annals of statistics*, 255–285.

Markowitz, H. (1952). Portfolio selection. *The journal of finance 7*(1), 77–91.

McLeish, D. L. et al. (1975). A maximal inequality and dependent strong laws. *The Annals of probability 3*(5), 829–839.

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *Journal of monetary Economics 15*(2), 145–161.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867–887.

Moreira, M. J., J. R. Porter, and G. A. Suarez (2009). Bootstrap validity for the score test when instruments may be weak. *Journal of Econometrics 149*(1), 52–64.

Nelson, C. and R. Startz (1988). Some further results on the exact small sample properties of the instrumental variable estimator.

Nelson, C. R. and R. Startz (2007). The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics 138*(1), 47–62.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Paparoditis, E. and D. N. Politis (2001). Tapered block bootstrap. *Biometrika 88*(4), 1105–1119.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical association 89*(428), 1303–1313.

Pollard, D. (1990). Empirical processes: theory and applications. In *NSF-CBMS regional conference series in probability and statistics*, pp. i–86. JSTOR.

Rousseeuw, P. J. and A. M. Leroy (2005). *Robust regression and outlier detection*, Volume 589. John wiley & sons.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association 105*(489), 218–235.

Staiger, D. O. and J. H. Stock (1994). Instrumental variables regression with weak instruments.

Stinchcombe, M. B. and H. White (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory 14*(3), 295–325.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24–36.

Turner, C. M., R. Startz, and C. R. Nelson (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics 25*(1), 3–22.

Van Der Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak convergence and empirical processes*, pp. 16–28. Springer.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1–25.

White, H. (1989). An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. In *Proceedings of the international joint conference on neural networks*, Volume 2, pp. 451–455. Washington, DC.

Wooldridge, J. M. and H. White (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric theory 4*(2), 210–230.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, 1261–1295.