Bonferroni-Based Size-Correction for Nonstandard Testing Problems[*]

Adam McCloskey[†]
Brown University

October 2011; This Version: October 2012

Abstract

We develop powerful new size-correction procedures for nonstandard hypothesis testing environments in which the asymptotic distribution of a test statistic is discontinuous in a parameter under the null hypothesis. Examples of this form of testing problem are pervasive in econometrics and complicate inference by making size difficult to control. This paper introduces two sets of new size-correction methods that correspond to two different general hypothesis testing frameworks. The new methods are designed to maximize the power of the underlying test while maintaining correct asymptotic size uniformly over the parameter space specified by the null hypothesis. They involve the construction of critical values that make use of reasoning derived from Bonferroni bounds. The first set of new methods provides complementary alternatives to existing size-correction methods, entailing substantially higher power for many testing problems. The second set of new methods provides the first available asymptotically size-correct tests for the general class of testing problems to which it applies. This class includes hypothesis tests on parameters after consistent model selection and tests on super-efficient/hard-thresholding estimators. We detail the construction and performance of the new tests in three specific examples: testing after conservative model selection, testing when a nuisance parameter may be on a boundary and testing after consistent model selection.

Keywords: Hypothesis testing, uniform inference, asymptotic size, exact size, power, size-correction, model selection, boundary problems, local asymptotics

1 Introduction

Nonstandard econometric testing problems have gained substantial attention in recent years. In this paper, we focus on a very broad class of these problems: those for which the null limit distribution of a test statistic is discontinuous in a parameter. The problems falling into this class range from tests in the potential presence of identification failure (e.g., Staiger and Stock, 1997 and Andrews and Cheng, 2012) to tests after pretesting or model selection (e.g., Leeb and Pötscher, 2005 and Guggenberger, 2010) to tests when a parameter may lie on the boundary of its parameter space (e.g., Andrews, 1999, 2001). Though test statistics that do not exhibit this type of discontinuity exist for some problems (e.g., Kleibergen, 2002), they do not for others. Moreover, such test statistics may not necessarily be preferred to parameter-discontinuous ones when good size-correction procedures are available, as they can have low power. However, we sidestep the important issue of choosing a test statistic in this paper, taking it as given.

The usual approximation to the size of a test is the asymptotic probability of rejecting a true null hypothesis (null rejection probability) at a fixed parameter value. For the types of problems studied in this paper, such an approximation is grossly misleading. In fact, the discrepancy between this point-wise null rejection probability (NRP) and the (asymptotic) size of a test can reach unity less the nominal level. This problem does not disappear, and often worsens, as the sample size grows. The inadequacy of point-wise asymptotic approximations and the resulting pitfalls for inference have been studied extensively in the literature. See, for example, Dufour (1997) in the context of inference on the two-stage least squares estimator, Leeb and Pötscher (2005) in the context of inference after model selection, Stoye (2009) in the context of inference on partially identified parameters and Andrews and Guggenberger (2009a) (AG henceforth) in the context of inference on the autoregressive parameter in a first-order autoregressive model. In the parameter-discontinuous testing framework of this paper, one must examine the maximal NRP *uniformly*, over the entire parameter space, as the sample size grows in order to determine the asymptotic size of the test (see e.g., the work of Mikusheva, 2007 and Andrews and Guggenberger, 2009b, 2010b). When using the test statistics considered in this paper, one typically takes a conservative approach to control size, leading to highly non-similar tests, i.e., tests for which the point-wise NRP differs substantially across parameter values. This often results in very poor power.

In this paper, we develop novel size-correction methods with the goal of minimizing the

degree of "conservativeness" of the test, and hence maximizing its power, while maintaining correct asymptotic size. We do so under two different frameworks that allow for different null limiting behavior of a given test statistic. For the first, termed the single localized limit distribution (SLLD) framework, we adopt the framework studied by AG as it is quite broad in scope. For the second, termed the multiple localized limit distributions (MLLDs) framework, we generalize the SLLD framework in order to accommodate certain complicated asymptotic behaviors of test statistics. To our knowledge, this latter, more general framework has not yet been studied. It includes examples of testing after consistent model selection and testing on super-efficient/hard-thresholding estimators.

The basic idea behind the size-corrections we introduce is to adaptively learn from the data how far the true parameters are from the point that causes the discontinuity in the asymptotic behavior of the test statistic in order to construct critical values (CVs) that control the size of the test but are not overly conservative. We do this under a "drifting sequence" framework by embedding the true parameter values in a sequence indexed by the sample size and a localization parameter. Within this framework, we estimate a corresponding localization parameter to find a set of drifting sequences of parameters relevant to the testing problem at hand. We then examine the CVs corresponding to the null limiting quantiles of the test statistic that obtain under the drifting sequences within this set. Though the localization parameter cannot be consistently estimated under these drifting sequences, it is often possible to obtain estimators that are asymptotically centered about their true values and hence to construct asymptotically valid confidence sets for the true localization parameter.

Based upon this estimator and corresponding confidence sets, we examine three different size-correction methods in increasing order of computational complexity. For the first, we search for the maximal CV over a confidence set, rather than the maximal CV over the entire space of localization parameters, to reduce the degree of conservativeness of a given hypothesis test. Inherent in this construction are two levels of uncertainty: one for the localization parameter and one for the test statistic itself. We use procedures based on Bonferroni bounds to account for both. For the second, we also search for a maximal CV over a confidence set but, instead of using Bonferroni bounds, we account for the two levels of uncertainty by adjusting CV levels according to the asymptotic distributions that arise under drifting parameter sequences. This method compensates for the asymptotic dependence structure between the test statistic and the CV, leading to more powerful tests. For the third, we find the smallest CVs over sets of those justified by the first and second,

leading to tests with high power over most of the parameter space.

For testing problems within the scope of the SLLD framework, our new size-correction methods can be constructed to be either uniformly more powerful asymptotically than existing least-favorable (LF) methods or are more powerful over most of the relevant parameter space. In the latter case, the portions of the parameter space for which LF methods dominate tend to be very small and such dominance tends to be nearly undetectable even within these portions. The finite-sample power dominance of our new methods can be very pronounced, sometimes reaching nearly 100% over most of the parameter space. Our size-corrections can also be constructed to direct power toward different regions of the parameter space while sacrificing very little in others. We also develop the first size-correction procedures we are aware of to provide tests with correct asymptotic size for all testing problems falling within the MLLDs framework. Since they are adapted from the size-correction procedures used in the SLLD framework to this generalized framework, they also adopt many desirable power properties.

The scope of problems to which our size-correction methods may be applied is quite wide. For illustration, we provide detailed applications to three nonstandard testing problems. Two of these examples concern testing after model selection/pretesting. The first, taken from AG, involves testing after conservative model selection and falls within the scope of the SLLD framework. The second, taken from Leeb and Pötscher (2005), considers testing after consistent model selection and falls within the scope of the MLLDs framework. The other example we detail concerns testing when a nuisance parameter may be on a boundary of its parameter space, taken from Andrews and Guggenberger (2010b). We also briefly discuss a subset of the numerous other examples for which our size-corrections can be used. We focus on testing after model selection in the examples because at present, available uniformly valid inference methods tend to be extremely conservative. Inference after model selection is an important issue that is all too frequently ignored, sometimes being referred to as "the quiet scandal of statistics". See, for example, Hansen (2005a) for a discussion of the importance of this issue. Moreover, excluding the results of AG and Kabaila (1998), the literature has been quite negative with regards to solving this inference problem (e.g., Andrews and Guggenberger, 2009b, Leeb and Pötscher, 2005, 2006 and 2008), especially with regards to inference after consistent model selection.

The recently developed methods for uniform inference of AG are closely related to the methods developed here. AG also study a given test statistic and adjust CVs according to a drifting parameter sequence framework. However, our methods tend to be (oftentimes,

much) less conservative than theirs. Existing inference procedures that make use of Bonferroni bounds are also related to those developed here. Some of the CVs we employ can also be interpreted as smoothed versions of those based on binary decision rules that use an inconsistent estimator of the localization parameter. Our CVs are also related to those that use a "transition function" approach to interpolate between LF and standard CVs. In contrast to these approaches, ours do not necessitate an ad hoc choice of transition function. Rather, they use the data and the limiting behavior of the test statistic to adaptively transition between CVs. See Sections 3.1 and 3.3 for details and references on these related procedures. Finally, the recent work of Elliott et al. (2012) takes a somewhat different approach to some of the problems discussed here by attempting to numerically determine tests that approach an asymptotic power bound.

The remainder of this paper is composed as follows. Section 2 describes the general class of nonstandard hypothesis testing problems we study, subsequently detailing the SLLD and MLLDs frameworks and providing examples. Section 3 goes on to specify the size-correction methods of this paper under the two localized limit distribution frameworks and provides the conditions under which some of these size-corrections yield correct asymptotic size. To conserve space, some of the conditions used to show size-correctness of procedures in the MLLDs framework are relegated to Appendix III, which also contains some auxiliary sufficient conditions. Specifics on how to construct some of the size-corrected CVs are provided for three econometric examples in Sections 4, 5 and 6. The finite sample performance of two of the examples, corresponding to testing after model selection, is also analyzed there. Section 7 concludes. Appendix I contains proofs of the main results of this paper while Appendix II is composed of derivations used to show how the example testing problems fit the assumptions of the paper. All tables and figures can be found at the end of the document.

To simplify notation, we will occasionally abuse it by letting (a_1, a_2) denote the vector $(a_1', a_2')'$. The sample size is denoted by n and all limits are taken to mean as $n \to \infty$. Let $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_- = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_{-,\infty} = \mathbb{R}_- \cup \{-\infty\}$, $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{\infty\}$ and $\mathbb{R}_\infty = \mathbb{R} \cup \{-\infty, \infty\}$. $\mathbf{1}(\cdot)$ denotes the indicator function. $\Phi(\cdot)$ and $\phi(\cdot)$ are the usual notation for the distribution and density functions of the standard normal distribution. "$\xrightarrow{d}$" and "$\xrightarrow{p}$" denote weak convergence and convergence in probability while $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ denote the usual (stochasitc) orders of magnitude.

4

2 Parameter-Discontinuous Asymptotic Distributions

In this paper, we are interested in performing hypothesis tests when the asymptotic distribution of the test statistic is discontinuous in a parameter under the null hypothesis. We take the test statistic as given and examine the tasks of controlling size and maximizing power for the given statistic. The important separate issue of choosing a test statistic depends on the specific testing problem at hand and is not the focus of this paper. In order to analyze this problem, we adopt the same general testing framework as AG. Consider some generic test statistic $T_n(\theta_0)$ used for testing $H_0 : \theta = \theta_0$ for some finite-dimensional parameter $\theta \in \mathbb{R}^d$. Under H_0, $T_n(\theta_0)$ and its asymptotic distribution depend on some parameter $\gamma \in \Gamma$. Refer to this limit distribution as F_γ. We decompose γ into three components, viz. $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, depending on how each component affects F_γ as follows. The distribution F_γ is discontinuous in γ_1, a parameter in $\Gamma_1 \in \mathbb{R}^p$, when one or more of the elements of γ_1 is equal to zero. It also depends on γ_2, a parameter in $\Gamma_2 \in \mathbb{R}^q$, but γ_2 does not affect the distance of γ to the point of discontinuity in F_γ. The third component γ_3 may be finite- or infinite-dimensional, lying in some general parameter space $\Gamma_3(\gamma_1, \gamma_2)$ that may depend on γ_1 and γ_2. The component γ_3 does not affect the limit distribution F_γ but may affect the properties of $T_n(\theta_0)$ in finite samples. Formally, the parameter space for γ is given by

$$\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}. \tag{1}$$

To complete the preliminary setup, we impose the following product space assumption on Γ_1. This assumption is identical to Assumption A of AG. Let $\lfloor$ ($\rfloor$) denote the left (right) endpoint of an interval that may be open or closed.

Assumption D. *(i)* Γ *satisfies* (1) *and (ii)* $\Gamma_1 = \prod_{m=1}^p \Gamma_{1,m}$, *where* $\Gamma_{1,m} = \lfloor \gamma_{1,m}^l, \gamma_{1,m}^u \rfloor$ *for some* $-\infty \le \gamma_{1,m}^l < \gamma_{1,m}^u \le \infty$ *that satisfy* $\gamma_{1,m}^l \le 0 \le \gamma_{1,m}^u$ *for* $m = 1, \ldots, p$.

This paper introduces testing methods that are asymptotically size-controlled. Asymptotic size control requires one to asymptotically bound the NRP *uniformly* over the parameter space admissible under the null hypothesis. In order to assess the uniform limiting behavior of a test, one must examine its behavior along drifting sequences of parameters (see e.g., Andrews and Guggenberger, 2010b and Andrews et al., 2011). In this vein, we allow γ to depend on the sample size, and emphasize this dependence by denoting it $\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3})$, where $h = (h_1, h_2) \in H \equiv H_1 \times H_2$ is a localization parameter that describes the limiting

behavior of the sequence. The sets H_1 and H_2 depend on Γ_1 and Γ_2 as follows:

$$H_1 = \prod_{m=1}^{p} \begin{cases} \mathbb{R}_{+,\infty}, & \text{if } \gamma_{1,m}^l = 0, \\ \mathbb{R}_{-,\infty}, & \text{if } \gamma_{1,m}^u = 0, \\ \mathbb{R}_{\infty}, & \text{if } \gamma_{1,m}^l < 0 \text{ and } \gamma_{1,m}^u > 0, \end{cases}$$
$$H_2 = \text{cl}(\Gamma_2),$$

where $\text{cl}(\Gamma_2)$ is the closure of Γ_2 with respect to $\mathbb{R}_{\infty}^q$. Given $r > 0$ and $h \in H$, define $\{\gamma_{n,h}\}$ as the sequence of parameters in Γ for which $n^r \gamma_{n,h,1} \to h_1$ and $\gamma_{n,h,2} \to h_2$. In this paper, we consider two broad classes of testing problems: one for which the limiting behavior of the test statistic is fully characterized by h under *any* drifting sequence of parameters $\{\gamma_{n,h}\}$ and the other for which this limiting behavior depends upon both h and the limiting behavior of $\{\gamma_{n,h,1}\}$ relative to another sequence.

2.1 Single Localized Limit Distribution Framework

We begin the analysis with the simpler of the two cases just described. This class of testing problems can be broadly characterized by Assumption D and the following assumption.

Assumption S-B.1. *There exists a single fixed $r > 0$ such that for all $h \in H$ and corresponding sequences $\{\gamma_{n,h}\}$, $T_n(\theta_0) \xrightarrow{d} W_h$ under H_0 and $\{\gamma_{n,h}\}$.*

Denote the limit distribution function for a given h as J_h, i.e., $P(W_h \leq x) = J_h(x)$ and the $(1-\alpha)^{th}$ quantile of W_h by $c_h(1-\alpha)$. We refer to J_h as a "localized" limit distribution as it obtains under a drifting sequence of parameters indexed by the localization parameter h. Assumption S-B.1 is identical to Assumption B of AG. For every sequence of parameters $\{\gamma_{n,h}\}$ indexed by the same localization parameter h, the same limit distribution J_h obtains, hence the term "single" in SLLD. We now introduce some new assumptions, noting that they are applicable to most of the same econometric applications that satisfy the assumptions imposed by AG.

Assumption S-B.2. *Consider some fixed $\delta \in (0, 1)$.*
(i) As a function in h from H into $\mathbb{R}$, $c_h(1 - \delta)$ is continuous.
(ii) For any $h \in H$, $J_h(\cdot)$ is continuous at $c_h(1 - \delta)$.

Assumption S-B.2 is a mild continuity assumption. To obtain stronger results, we strengthen part (i) as follows.

Assumption S-BM.1. *For some fixed $\alpha \in (0,1)$ and pair $(\underline{\delta}, \bar{\delta}) \in [0, \alpha - \bar{\delta}] \times [0, \alpha - \underline{\delta}]$, as a function of h and δ, $c_h(1 - \delta)$ is continuous over H and $[\underline{\delta}, \alpha - \bar{\delta}]$.*

The quantity $\underline{\delta}$ serves as a lower bound and $\alpha - \bar{\delta}$ serves as an upper bound on the points δ for which $c_h(1 - \delta)$ must be continuous. In many examples of interest, W_h is a continuous random variable with infinite support so that $c_h(1) = \infty$. For such examples $\underline{\delta}$ can be set arbitrarily close, but not equal, to zero. Assumptions D, S-B.1, S-B.2 and S-BM.1 and the other assumptions for this framework introduced later in Sections 3.1 through 3.3 hold in many nonstandard econometric testing problems of interest. The following are simple, illustrative examples of such problems.

2.1.1 Testing After Conservative Model Selection

Various forms of hypothesis tests after model selection exemplify testing problems with parameter-discontinuous null limit distributions. Conducting a t-test on a parameter of interest after *conservative* model selection falls within the framework of Section 2.1, having a SLLD. Conservative model selection includes, among others, methods based on the Akaike information criterion (AIC) and standard pre-testing techniques. As an illustrative example, consider the following problem described by AG. We have a model given by

$$y_i = x_{1i}^* \theta + x_{2i}^* \beta_2 + x_{3i}^{*\prime} \beta_3 + \sigma \varepsilon_i, \tag{2}$$

for $i = 1, \ldots, n$, where $x_i^* \equiv (x_{1i}^*, x_{2i}^*, x_{3i}^{*\prime})' \in \mathbb{R}^k$, $\beta \equiv (\theta, \beta_2, \beta_3')' \in \mathbb{R}^k$, $x_{1i}^*, x_{2i}^*, \theta, \beta_2, \sigma, \varepsilon_i \in \mathbb{R}$, $x_{3i}^*, \beta_3 \in \mathbb{R}^{k-2}$, the observations $\{(y_i, x_i^*)\}$ are i.i.d. and ε_i has mean zero and unit variance conditional on x_i^*. We are interested in testing $H_0 : \theta = \theta_0$ after determining whether to include x_{2i}^* in the regression model (2), that is, after determining whether to impose the restriction $\beta_2 = 0$. This decision is based on whether the absolute value of the pretest t-statistic

$$|T_{n,2}| \equiv \left| \frac{n^{1/2} \hat{\beta}_2}{\hat{\sigma}(n^{-1} X_2^{*\prime} M_{[X_1^*:X_3^*]} X_2^*)^{-1/2}} \right|$$

exceeds a pretest CV $c > 0$, where c is fixed (i.e., does not depend on n), $\hat{\beta}_2$ is the standard unrestricted OLS estimator of β_2 in the regression (2) and $\hat{\sigma}^2 \equiv (n - k)^{-1} Y' M_{[X_1^*:X_2^*:X_3^*]} Y$ with $Y \equiv (y_1, \ldots, y_n)'$, $X_j^* \equiv [x_{j1}^* : \ldots : x_{jn}^*]'$ for $j = 1, 2, 3$ and $M_A \equiv I - A(A'A)^{-1}A'$ for some generic full-rank matrix A and conformable identity matrix I. The model selection pretest rejects $\beta_2 = 0$ when $|T_{n,2}| > c$ and the subsequent t-statistic for testing H_0 is based

on the unrestricted version of (2) and is given by

$$\hat{T}_{n,1}(\theta_0) \equiv \frac{n^{1/2}(\hat{\theta} - \theta_0)}{\hat{\sigma}(n^{-1}X_1^{*\prime}M_{[X_2^*:X_3^*]}X_1^*)^{-1/2}},$$

where $\hat{\theta}$ is the unrestricted OLS estiamtor from regression (2). Conversely, the model selection pretest selects the model without x_{2i}^*, or equivalently restricts $\beta_2 = 0$, when $|T_{n,2}| \leq c$ and the resulting t-statistic for H_0 is given by

$$\widetilde{T}_{n,1}(\theta_0) = \frac{n^{1/2}(\widetilde{\theta} - \theta_0)}{\hat{\sigma}(n^{-1}X_1^{*\prime}M_{X_3^*}X_1^*)^{-1/2}},$$

where $\widetilde{\theta}$ is the restricted OLS estimator from regression (2) with β_2 restricted to equal zero. Hence, for a two-sided test, the post-conservative model selection test statistic for testing H_0 is given by

$$T_n(\theta_0) = |\widetilde{T}_{n,1}(\theta_0)|\mathbf{1}(|T_{n,2}| \leq c) + |\hat{T}_{n,1}(\theta_0)|\mathbf{1}(|T_{n,2}| > c).$$

With straightforward modification, the results described below also apply to one-sided testing for this problem. See AG and Andrews and Guggenberger (2009c) for more details.[1]

Results in AG show how this testing problem satisfies Assumptions D and S-B.1. Specifically, let G denote the distribution of (ε_i, x_i^*) and define the following quantities

$$x_i^\perp = \begin{pmatrix} x_{1i}^* - x_{3i}^{*\prime}(E_G x_{3i}^* x_{3i}^{*\prime})^{-1}x_{3i}^* x_{1i}^* \\ x_{2i}^* - x_{3i}^{*\prime}(E_G x_{3i}^* x_{3i}^{*\prime})^{-1}x_{3i}^* x_{2i}^* \end{pmatrix}$$

$$Q = E_G x_i^\perp x_i^{\perp\prime}, \quad \text{and} \quad Q^{-1} = \begin{pmatrix} Q^{11} & Q^{12} \\ Q^{12} & Q^{22} \end{pmatrix}.$$

Then for this example, $\gamma_1 = \beta_2/\sigma(Q^{22})^{1/2}$, $\gamma_2 = Q^{12}/(Q^{11}Q^{22})^{1/2}$ and $\gamma_3 = (\beta_2, \beta_3, \sigma, G)'$ and

$$T_n(\theta_0) \xrightarrow{d} \begin{cases} |\widetilde{Z}_1|\mathbf{1}(|Z_2| \leq c) + |\hat{Z}_1|\mathbf{1}(|Z_2| > c), & \text{if } \gamma_1 = 0 \\ |\hat{Z}_1|, & \text{if } \gamma_1 \neq 0, \end{cases}$$

where $\widetilde{Z}_1$, $\hat{Z}_1$ and Z_2 are standard normal random variables with $\widetilde{Z}_1$ independent of Z_2 and $\text{Corr}(\hat{Z}_1, Z_2) = \gamma_2$. The parameter spaces in (1) are given by $\Gamma_1 = \mathbb{R}$, $\Gamma_2 = [-1 + \omega, 1 - \omega]$ for some $\omega > 0$ and

$$\Gamma_3(\gamma_1, \gamma_2) = \{(\beta_2, \beta_3, \sigma, G) : \beta_2 \in \mathbb{R}, \beta_3 \in \mathbb{R}^{k-2}, \sigma \in (0, \infty), \gamma_1 = \beta_2/\sigma(Q^{22})^{1/2},$$

---

[1]The testing problem described here also applies to testing a linear combination of regression coefficients after conservative model selection. This involves a reparameterization described in Andrews and Guggenberger (2009c).

$$\gamma_2 = Q^{12}/(Q^{11}Q^{22})^{1/2}, \lambda_{\min}(Q) \geq \kappa, \lambda_{\min}(E_G x^*_{3i} x^{*\prime}_{3i}) \geq \kappa, E_G\|x^*_i\|^{2+\delta} \leq M,$$
$$E_G\|\varepsilon_i x^*_i\|^{2+\delta} \leq M, E_G(\varepsilon_i|x^*_i) = 0 \text{ a.s.}, E_G(\varepsilon_i^2|x^*_i) = 1 \text{ a.s.}\}$$

for some $\kappa, \delta > 0$ and $M < \infty$, where $\lambda_{\min}(A)$ is the smallest eigenvalue of generic matrix A. From this parameter space, it is clear that Assumption D holds for this example. Moreover, Andrews and Guggenberger (2009c) show that Assumption S-B.1 holds with $r = 1/2$ and

$$J_h(x) = \Delta(h_1 h_2(1 - h_2^2)^{-1/2}, x)\Delta(h_1, c) + \int_{-x}^{x}\left(1 - \Delta\left(\frac{h_1 + h_2 t}{(1 - h_2^2)^{1/2}}, \frac{c}{(1 - h_2^2)^{1/2}}\right)\right)\phi(t)dt,$$

where $\Delta(a, b) = \Phi(a+b) - \Phi(a-b)$. As defined, $H_1 = \mathbb{R}_\infty$ and $H_2 = [-1+\omega, 1-\omega]$. Turning to the new assumptions, the lower bound $\underline{\delta}$ of Assumption S-BM.1 can be set arbitrarily close to zero and $\bar{\delta}$ can be set to any quantity in its admissible range $[0, \alpha - \underline{\delta}]$ since W_h is a continuous random variable with support over the entire real line for any $h \in H$, which can be seen by examining $J_h(\cdot)$. This fact similarly implies Assumption S-B.2(ii) holds. An assumption imposed later will necessitate a restriction on $\bar{\delta}$ that we will discuss in Section 4. Continuity of $c_h(1 - \delta)$ in h follows from the facts that $c_h(1 - \delta) = J_h^{-1}(1 - \delta)$ and $J_h(x)$ is continuous in $h \in H$.

The SLLD framework applies to many more complex examples of testing after conservative model selection. For example, results in Leeb (2006) and Leeb and Pötscher (2008) can be used to verify the assumptions of this paper for a sequential general-to-specific model selection procedure with multiple potential control variables.

2.1.2 Testing when a Nuisance Parameter may be on a Boundary

We now explore an illustrative example of a testing problem in which a nuisance parameter may be on the boundary of its parameter space under the null hypothesis and show how it also falls within the framework of Section 2.1. This problem is considered by Andrews and Guggenberger (2010b) and can be described as follows. We have a sample of size n of an i.i.d. bivariate random vector $X_i = (X_{i1}, X_{i2})'$ with distribution F. Under F, the first two moments of X_i exist and are given by

$$E_F(X_i) = \begin{pmatrix} \theta \\ \mu \end{pmatrix} \quad \text{and} \quad \text{Var}_F(X_i) = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

Say we are interested in the null hypothesis $H_0 : \theta = \theta_0$ and we know that $\mu \geq 0$. Now suppose we use the Gaussian quasi-maximum likelihood estimator of $(\theta, \mu, \sigma_1, \sigma_2, \rho)$ under

the restriction $\mu \geq 0$, denoted by $(\hat{\theta}_n, \hat{\mu}_n, \hat{\sigma}_{n1}, \hat{\sigma}_{n2}, \hat{\rho}_n)$, to construct a lower one-sided t-test of H_0.[2] That is, $T_n(\theta_0) = -n^{1/2}(\hat{\theta}_n - \theta_0)/\hat{\sigma}_{n1}$, where $\hat{\theta}_n = \bar{X}_{n1} - (\hat{\rho}_n\hat{\sigma}_{n1})\min(0, \bar{X}_{n2}/\hat{\sigma}_{n2})$ with $\bar{X}_{nj} = n^{-1}\sum_{i=1}^n X_{ij}$ for $j = 1, 2$. As in the previous example, results for upper one-sided and two-sided tests are quite similar.

Results in Andrews and Guggenberger (2010b) provide that this testing problem also satisfies Assumptions D and S-B.1. Here, $\gamma_1 = \mu/\sigma_2$, $\gamma_2 = \rho$ and $\gamma_3 = (\sigma_1, \sigma_2, F)$ and

$$T_n(\theta_0) \overset{d}{\longrightarrow} \begin{cases} -Z_1 + \gamma_2 \min\{0, Z_2\}, & \text{if } \gamma_1 = 0 \\ -Z_1, & \text{if } \gamma_1 \neq 0, \end{cases}$$

where Z_1 and Z_2 are standard normal random variables with $\text{Corr}(Z_1, Z_2) = \gamma_2$. The corresponding parameter spaces are $\Gamma_1 = \mathbb{R}_+$, $\Gamma_2 = [-1 + \omega, 1 - \omega]$ for some $\omega > 0$ and

$$\Gamma_3(\gamma_1, \gamma_2) = \{(\sigma_1, \sigma_2, F) : \sigma_1, \sigma_2 \in (0, \infty), E_f\|X_i\|^{2+\delta} \leq M, \theta = 0, \gamma_1 = \mu/\sigma_2, \gamma_2 = \rho\}$$

for some $M < \infty$ and $\delta > 0$.[3] From these definitions, it is immediate that Assumption D holds. Assumption S-B.1 holds for this example with $r = 1/2$ and

$$W_h \overset{d}{\sim} -Z_{h_2,1} + h_2 \min\{0, Z_{h_2,2} + h_1\}, \quad \text{where} \tag{3}$$

$$Z_{h_2} = \begin{pmatrix} Z_{h_2,1} \\ Z_{h_2,2} \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & h_2 \\ h_2 & 1 \end{pmatrix}\right).$$

In order to verify the new assumptions, it is instructive to examine the distribution function $J_h(\cdot)$, which is given by

$$J_h(x) = \Phi\left(\frac{x - h_2 h_1}{\sqrt{1 - h_2^2}}\right)\Phi(-h_1) + \int_{-x}^{\infty}\left(1 - \Phi\left(\frac{-h_1 - h_2 z}{(1 - h_2^2)^{1/2}}\right)\right)\phi(z)dz \tag{4}$$

(see Appendix II for its derivation). By definition, $H_1 = \mathbb{R}_{+,\infty}$ and $H_2 = [-1 + \omega, 1 - \omega]$. Now, looking at the form of J_h, we can see that, as in the previous example, W_h is a continuous random variable with support over the entire real line for any $h \in H$ so that $\underline{\delta}$ of Assumption S-BM.1 can again be set arbitrarily close to zero and $\bar{\delta}$ is left unrestricted over its admissible

---

[2]The results below also allow for different estimators in this construction. See Andrews and Guggenberger (2010b) for details.

[3]For the purposes of this paper, we make a small departure from the exact setup used by Andrews and Guggenberger (2010b) in our definition of Γ_2, which they define as $[-1, 1]$. That is, we bound the possible correlation between X_{i1} and X_{i2} to be less than perfect. We do this in order to employ the size-corrections described later in this paper. Note that the analogous assumption is imposed in the above post-conservative model selection example.

range. We discuss a restriction imposed later on $\bar{\delta}$ in Section 5. For the same reasons given in the previous example, Assumption S-B.2(ii) holds as well.

More complicated testing problems when a nuisance parameter may be on a boundary can also be shown to fit the SLLD framework and later assumptions of this paper. For example, Andrews (1999, 2001) provides results for more complicated boundary examples that fit this framework.

2.1.3   Other Examples

There are many examples in the econometrics literature of testing problems that fit the SLLD framework. Apart from those discussed above, these include, but are not limited to, tests after pretests with fixed CVs (e.g., Guggenberger, 2010), testing when the parameter *of interest* may lie on the boundary of its parameter space (e.g., Andrews and Guggenberger, 2010a), tests on model-averaged estimators (e.g., Hansen, 2007), tests on certain types of shrinkage estimators (e.g., Hansen, 2012), tests in autoregressive models that may contain a unit root (e.g., AG and Mikusheva, 2007), Vuong tests for nonnested model selection (e.g., Shi, 2011), subvector tests allowing for weak identification (e.g., Guggenberger et al., 2012) and tests on break dates and coefficients in structural change models (e.g., Elliott et al., 2012 and Elliott and Müller, 2012).

The SLLD assumptions technically preclude certain testing problems with parameter-discontinuous null limit distributions such as testing in moment inequality models (e.g., Andrews and Soares, 2010) and certain tests allowing for weak identification (e.g., Staiger and Stock, 1997). Nevertheless, the SLLD framework of this paper can be modified in a problem-specific manner to incorporate some of these problems and apply the testing methods introduced later to them. For example, Assumption D does not allow for testing in the moment inequality context when the condition of one moment binding depends upon whether another moment binds. Yet the results of Andrews and Soares (2010), Andrews and Barwick (2011) and Romano et al. (2012) suggest that tailoring the assumptions to this context would permit analogous results to those presented later.

2.2   Multiple Localized Limit Distributions Framework

The MLLDs framework generalizes the SLLD framework. The motivation for this generalization comes from an important class of hypothesis testing problems with parameter-discontinuous null limit distributions that do not satisfy Assumption S-B.1 because under H_0 and a given $\{\gamma_{n,h}\}$, the asymptotic behavior of $T_n(\theta_0)$ is not fully characterized by h.

In this more general framework, we retain the description of the parameter space given in Assumption D as well as the description of H following it but weaken Assumption S-B.1 to the following.

Assumption M-B.1. *There is a sequence $\{k_n\}$, a set $K \subset H_1$ and a single fixed $r > 0$ such that for all $h \in H$ and corresponding sequences $\{\gamma_{n,h}\}$, under H_0:*

(i) if $\lim_{n \to \infty} \gamma_{n,h,1}/k_n \in K$, $T_n(\theta_0) \overset{d}{\longrightarrow} W_h^{(1)}$;

(ii) if $\lim_{n \to \infty} \gamma_{n,h,1}/k_n \in L \subset K^c$, $T_n(\theta_0) \overset{d}{\longrightarrow} W_h^{(2)}$;

(iii) if $\lim_{n \to \infty} \gamma_{n,h,1}/k_n \in L^c \cap K^c$, the asymptotic distribution of $T_n(\theta_0)$ is stochastically dominated by $W_h^{(1)}$ or $W_h^{(2)}$.

Assumption M-B.1 allows for different h-dependent localized limit distributions that are relevant to the different possible limiting behaviors of $\gamma_{n,h,1}/k_n$. It collapses to Assumption S-B.1 when $k_n = n^{-r}$ and $K = H_1$ or $L = H_1$. The auxiliary sequence $\{k_n\}$ may depend upon the elements of $\{\gamma_{n,h}\}$ though this potential dependence is suppressed in the notation. Denote the limit distributions corresponding to (i) and (ii) as $J_h^{(1)}$ and $J_h^{(2)}$, which are the two localized limit distributions that obtain under the corresponding sequences of $\gamma_{n,h,1}/k_n$. Similarly, $c_h^{(1)}$ and $c_h^{(2)}$ denote the corresponding quantile functions. Finally, we denote the limit random variable under (iii) as $W_h^{(3)}$. This is a slight abuse of notation because there may be MLLDs that obtain under (iii) alone. Distinction between these distributions is not necessary here because of the imposed stochastic dominance. We will also make use of the following definition: $\zeta(\{\gamma_{n,h}\}) \equiv \lim_{n \to \infty} \gamma_{n,h,1}/k_n$.

The form that $\{k_n\}$, K and L take are specific to the testing problem at hand. However, we make a few general remarks in order to provide some intuition. The MLLDs frameowrk incorporates testing problems after a decision rule that compares some statistic to a sample-size-dependent quantity, say $\widetilde{c}_n$, decides the form $T_n(\theta_0)$ takes. When using such a decision rule, under the drifting sequence of parameters $\{\gamma_{n,h}\}$, the null limit distribution of $T_n(\theta_0)$ not only depends on the limit of $n^r \gamma_{n,h,1}$ (and $\gamma_{n,h,2}$), but it also depends on how fast $n^r \gamma_{n,h,1}$ grows relative to $\widetilde{c}_n$. The sequence $\{k_n\}$ is thus some (scaled) ratio of $\widetilde{c}_n$ to n^r and the sets K and L describe the limiting behavior of $n^r \gamma_{n,h,1}$ relative to $\widetilde{c}_n$.

This setting can clearly be further generalized to allow for other sequences like $\{\gamma_{n,h,1}/k_n\}$ to also determine the limiting behavior of $T_n(\theta_0)$ under H_0. For example, one additional sequence of this sort could allow for two additional localized limit distributions that are not necessarily stochastically dominated by any of the others. In this case, both the space containing the limit of $\gamma_{n,h,1}/k_n$ and the limit of this additional sequence could determine the

12

null limiting behavior of $T_n(\theta_0)$ under any given $\{\gamma_{n,h}\}$. We conjecture that a more general case like this obtains for hypothesis tests after more complicated consistent model selection procedures than those for which asymptotic results under drifting sequences of parameters are presently available. (The intuition for this is given in Section 2.2.1 below).

In this framework, we will also accommodate certain types of discontinuities in the localized limit distribution J_h. One form of these discontinuities occurs when h_1 is on the boundary of its parameter space H_1, entailing infinite values. In order to accommodate this type of discontinuity, define the following subset of H: $\bar{H} \equiv \text{int}(H_1) \times H_2$. Then the set of h corresponding to h_1 on the boundary of H_1 is equal to $\bar{H}^c$.

Assumption M-B.2. *Consider some fixed $\delta \in (0, 1)$.*
(i) As a function from $\bar{H}$ into $\mathbb{R}$, $c_h^{(i)}(1 - \delta)$ is continuous for $i = 1, 2$.
(ii) For $i = 1, 2$ and any $h \in \bar{H}$, there is some finite $\varepsilon_i \geq 0$ for which $J_h^{(i)}(\cdot)$ is continuous at $c_h^{(i)}(1 - \delta) + \varepsilon_i$.

Assumption M-B.2 is a continuity assumption that is a relaxed version of a direct adaptation of Assumption S-B.2 to the MLLDs framework. For the problems in this class, the localized quantiles can be infinite for $h \in \bar{H}^c$. This is why (i) is only required as a function from $\bar{H}$. Similarly, the localized limit distribution functions can be discontinuous at the their localized quantiles but continuous in a neighborhood near them. Part (ii) provides the flexibility to allow for this feature. A sufficient condition for Assumption M-B.2 to hold for $i = 1$ or 2 is that the $W_h^{(i)}$ is a continuous random variable with infinite support and a distribution function that is continuous in h. In this case $\varepsilon_i = 0$. As in the previous framework, we strengthen part (i) of this assumption to obtain stronger results.

Assumption M-BM.1. *For some fixed $\alpha \in (0, 1)$ and pairs $(\underline{\delta}^{(i)}, \bar{\delta}^{(i)}) \in [0, \alpha - \bar{\delta}^{(i)}] \times [0, \alpha - \underline{\delta}^{(i)}]$ for $i = 1, 2$, as a function of h and δ, $c_h^{(i)}(1 - \delta)$ is continuous over $\bar{H}$ and $[\underline{\delta}^{(i)}, \alpha - \bar{\delta}^{(i)}]$.*

Analogous remarks to those on Assumption S-BM.1 can be made here with the exception that continuity in h is no longer required at $h \in \bar{H}^c$.

2.2.1 Testing After Consistent Model Selection

Unlike hypothesis testing after conservative model selection, testing after *consistent* model selection entails substantially more complicated limiting behavior of a test statistic under the null hypothesis. The essential difference between conservative and consistent model selection in the context of our examples is that in consistent model selection, the comparison/critical

value used in the model selection criterion grows with the sample size. This is the case for example, with the popular Bayesian information criterion (BIC) and the Hannan-Quinn information criterion.

The simple post-consistent model selection testing framework provided by Leeb and Pötscher (2005) provides an illustrative example of a testing problem that fits the MLLDs framework. Hence, we shall consider it here. We now consider the regression model

$$y_i = \theta x_{1i} + \beta_2 x_{2i} + \epsilon_i \tag{5}$$

for $i = 1, \ldots, n$, where $\epsilon_i \overset{d}{\sim} i.i.d. N(0, \sigma^2)$ with $\sigma^2 > 0$, $X \equiv (x_1', \ldots, x_n')'$ with $x_i \equiv (x_{1i}, x_{2i})'$ is nonstochastic, full-rank and satisfies $X'X/n \to Q > 0$. For simplicity, assume that σ^2 is known though the unknown σ^2 case can be handled similarly. We are interested in testing $H_0 : \theta = \theta_0$ after deciding whether or not to include x_{2i} in the regression model (5) via a consistent model selection rule. As in the conservative model selection framework of Section 2.1.1, this decision is based on comparing the t-statistic for β_2 with some CV except now this CV c_n grows in the sample size such that $c_n \to \infty$ but $c_n/\sqrt{n} \to 0$. Formally, let

$$\sigma^2 (X'X/n)^{-1} = \begin{pmatrix} \sigma_{\theta,n}^2 & \sigma_{\theta,\beta_2,n} \\ \sigma_{\theta,\beta_2,n} & \sigma_{\beta_2,n}^2 \end{pmatrix}$$

and $\rho_n = \sigma_{\theta,\beta_2,n}/(\sigma_{\theta,n}\sigma_{\beta_2,n})$. Then the model selection procedure "chooses" to include x_{2i} in the regression if $|\sqrt{n}\hat{\beta}_2/\sigma_{\beta_2,n}| > c_n$, where $\hat{\beta}_2$ is the (unrestricted) OLS estimator of β_2 in the regression (5), and "chooses" to restrict $\beta_2 = 0$ otherwise. For this example, we will test H_0 by examining the non-studentized quantity $\sqrt{n}(\bar{\theta} - \theta_0)$, where $\bar{\theta}$ is equal to the unrestricted OLS estimator of θ in regression (5) when $|\sqrt{n}\hat{\beta}_2/\sigma_{\beta_2,n}| > c_n$ and the restricted OLS estimator of θ in (5) with β_2 restricted to equal zero when $|\sqrt{n}\hat{\beta}_2/\sigma_{\beta_2,n}| \leq c_n$.[4] That is for an upper one-sided test, the post-consistent model selection test statistic for testing H_0 is given by

$$T_n(\theta_0) = \sqrt{n}(\widetilde{\theta} - \theta_0)\mathbf{1}(|\sqrt{n}\hat{\beta}_2/\sigma_{\beta_2,n}| \leq c_n) + \sqrt{n}(\hat{\theta} - \theta_0)\mathbf{1}(|\sqrt{n}\hat{\beta}_2/\sigma_{\beta_2,n}| > c_n),$$

where $\widetilde{\theta}$ and $\hat{\theta}$ are the restricted and unrestricted estimators. Examining $-T_n(\theta_0)$ and $|T_n(\theta_0)|$ and their corresponding localized null limit distributions would allow us to perform the same analysis for lower one-sided and two-sided tests of H_0.

---

[4]Following Leeb and Pötscher (2005), we examine the non-studentized quantity rather than the t-statistic because use of the latter will not satisfy Assumption D and is therefore not amenable to the procedures put forth in this paper. Note that although the studentized quantity does not display a parameter-discontinuous null limit distribution, it suffers the same size-distortion problem when standard CVs are used.

Let the limits of all finite sample quantities be denoted by a ∞ subscript, e.g., $\sigma^2_{\theta,\infty}$. Then for this example, $\gamma_1 = \beta_2 \rho_\infty / \sigma_{\beta_2,\infty}$, $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2}) = (\sigma_{\theta,\infty}, \rho_\infty)'$ and $\gamma_3 = (\beta_2, \sigma^2, \sigma^2_{\theta,n}, \sigma^2_{\beta_2,n}, \rho_n)'$ and

$$T_n(\theta_0) \xrightarrow{d} \begin{cases} N(0, (1 - \gamma^2_{2,2})\gamma^2_{2,1}), & \text{if } \gamma_1 = 0 \\ N(0, \gamma^2_{2,1}), & \text{if } \gamma_1 \neq 0. \end{cases}$$

The parameter spaces in (1) are given by $\Gamma_1 = \mathbb{R}$, $\Gamma_2 = [\eta, M] \times [-1 + \omega, 1 - \omega]$ for some $\eta \in (0, M]$, $\omega \in (0, 1]$ and $M \in (0, \infty)$ and

$$\Gamma_3(\gamma_1, \gamma_2) = \{(\beta_2, \sigma^2, \sigma^2_{\theta,n}, \sigma^2_{\beta_2,n}, \rho_n) : \beta_2 \in \mathbb{R}, \sigma^2 \in (0, \infty), \gamma_1 = \beta_2 \rho_\infty / \sigma_{\beta_2,\infty}, \gamma_2 = (\sigma_{\theta,\infty}, \rho_\infty)'$$
$$\lim_{n \to \infty} \sigma^2_{\theta,n} = \gamma_{2,1}, \lim_{n \to \infty} \sigma^2_{\beta_2,n} = \beta_2 \gamma_{2,2}/\gamma_1, \lim_{n \to \infty} \rho_n = \gamma_{2,2}\}.$$

Clearly, Γ satisfies Assumption D. Using the results of Proposition A.2 in Leeb and Pötscher (2005), we can establish that Assumption M-B.1 is satisfied with $k_n = c_n \rho_n / \sqrt{n}$, $r = 1/2$, $K = (-1, 1)$, $J_h^{(1)}(x) = \Phi((1 - h^2_{2,2})^{-1/2}(x/h_{2,1} + h_1))$, $L = [-\infty, -1) \cup (1, \infty]$ and $J_h^{(2)}(x) = \Phi(x/h_{2,1})$. See Appendix II for details.

The intuition for why different null limit distributions for $T_n(\theta_0)$ obtain under $\{\gamma_{n,h}\}$ depending on how $\gamma_{n,h,1} = \beta_{2,n} \rho_n / \sigma_{\beta_2,n}$ behaves relative to $k_n = c_n \rho_n / \sqrt{n}$ as the sample size grows lies in the fact that when $|\gamma_{n,h,1}/k_n| = |\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n| < 1$, $T_n(\theta_0)$ equals $\sqrt{n}(\widetilde{\theta} - \theta_0)$ asymptotically. Conversely, when $|\gamma_{n,h,1}/k_n| = |\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n| > 1$, $T_n(\theta_0)$ equals $\sqrt{n}(\hat{\theta} - \theta_0)$ asymptotically. Under H_0 and the drifting sequence $\{\gamma_{n,h}\}$, the statistics $\sqrt{n}(\widetilde{\theta} - \theta_0)$ and $\sqrt{n}(\hat{\theta} - \theta_0)$ have different limit distributions, corresponding to $J_h^{(1)}$ and $J_h^{(2)}$, respectively. In the knife-edge case for which $\lim_{n\to\infty} |\gamma_{n,h,1}/k_n| = 1$, the limit of $T_n(\theta_0)$ then also depends on the limiting behavior of $c_n \pm \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}$ (see Leeb and Pötscher, 2005). However, no matter the limit of this latter quantity, the limit of $T_n(\theta_0)$ in this case is always stochastically dominated by the limit that pertains under either $\lim_{n\to\infty} |\gamma_{n,h,1}/k_n| < 1$ or $\lim_{n\to\infty} |\gamma_{n,h,1}/k_n| > 1$ (see Appendix II). Hence, under a given drifting sequence $\{\gamma_{n,h}\}$, the limiting behavior of $T_n(\theta_0)$ is not fully characterized by h, in violation of Assumption S-B.1.

By definition, $H_1 = \mathbb{R}_\infty$ and $H_2 = [\eta, M] \times [-1 + \omega, 1 - \omega]$. Turning now to Assumption M-BM.1, $\underline{\delta}^{(1)}$ can be set arbitrarily close to, but strictly greater than, zero. Since $c_h^{(1)}(1 - \delta)$ is the $(1-\delta)^{th}$ quantile of a normal distribution with mean $-h_1 h_{2,1}$ and variance $h^2_{2,1}(1 - h^2_{2,2})$ and $\bar{H} = \mathbb{R} \times [\eta, M] \times [-1 + \omega, 1 - \omega]$, $c_h^{(1)}(1 - \delta)$ is continuous in h over $\bar{H}$ for any $\delta \in (0, 1)$. Continuity in δ over $[\underline{\delta}^{(1)}, \alpha - \bar{\delta}^{(1)}]$ also follows for any $\bar{\delta}^{(1)} \in [0, \alpha - \underline{\delta}^{(1)}]$. Similar reasoning shows that, $\underline{\delta}^{(2)}$ can be set arbitrarily close to zero and $\bar{\delta}^{(2)}$ can be anywhere in its admissible range for Assumption M-BM.1 to hold (we discuss restrictions imposed on $\bar{\delta}^{(i)}$ for $i = 1, 2$

by later assumptions in Section 6). For $h \in \bar{H}$ and $i = 1, 2$, $J_h^{(i)}$ is the distribution function of a continuous random variable so that Assumption M-B.2(ii) holds with $\varepsilon_i = 0$.

As with the other examples studied in this paper, this simple illustrative example is not the most general of its kind to fall into this framework. Many of the assumptions in the above example can be relaxed. However, uniform asymptotic distributional results for more complex consistent model selection procedures are not readily available in the literature. This may be in part due to the very negative results put forth regarding attempts to conduct inference after even the simplest of such procedures (e.g., Leeb and Pötscher, 2005 and Andrews and Guggenberger, 2009b). As alluded to above, more complicated procedures, such as a consistent version of the sequential general-to-specific model selection approach of Leeb (2006) and Leeb and Pötscher (2008) or standard BIC approaches to more complicated models, likely require a straightforward extension of the MLLDs framework and the corresponding CVs introduced below. The intuition for this essentially follows from the same intuition as that used for the simple post-consistent model selection example provided above. As a simplification, suppose now that another regressor enters the potential model (5) with associated coefficient β_3. Using the obvious notation, we may also wish to determine whether β_3 should enter the model prior to testing H_0 by comparing $|\sqrt{n}\hat{\beta}_3/\sigma_{\beta_3}|$ to c_n. In this case $T_n(\theta_0)$ would take one of four, rather than two, possible values depending on both the value of $|\sqrt{n}\hat{\beta}_3/\sigma_{\beta_3}|$ and $|\sqrt{n}\hat{\beta}_2/\sigma_{\beta_2}|$ relative to c_n (ignoring knife-edge cases).

2.2.2 Other Examples

The class of super-efficient/hard-thresholding estimators studied by Andrews and Guggenberger (2009b) and Pötscher and Leeb (2009), including Hodges' estimator, also fit the MLLDs framework. A certain subclass of these estimators in fact requires $\varepsilon_i > 0$ in Assumption M-B.2(ii) for $i = 1$ or 2, unlike the problem considered immediately above. Related problems of testing after pretests using pretest CVs that grow in the sample size fit the MLLDs framework as well. Though some very recent work has explored the properties of uniformly valid confidence intervals for some of the problems that fall into this framework (e.g., Pötscher and Schneider, 2010), to the author's knowledge, this is the first time such a framework and corresponding uniformly valid inference procedure has been presented at this level of generality.

3 Bonferroni-Based Critical Values

For test statistics with parameter-discontinuous null limit distributions, the asymptotic NRP of the test, evaluated at a given parameter value permissible under H_0 can provide a very poor approximation to the true NRP and size of the test, even for large samples. In order to be more precise about this terminology, we introduce the following definitions for a test of $H_0 : \theta = \theta_0$, working under the framework described in Section 2. Let κ_n be the (possibly random or sample-size-dependent) CV being used. The NRP evaluated at $\gamma \in \Gamma$ is given by $P_{\theta_0,\gamma}(T_n(\theta_0) > \kappa_n)$, where $P_{\theta_0,\gamma}(E)$ denotes the probability of event E given that (θ_0, γ) are the true parameters describing the data-generating process (DGP). The asymptotic NRP of a test statistic $T_n(\theta_0)$ and cv κ_n evaluated at $\gamma \in \Gamma$ is given by $\limsup_{n\to\infty} P_{\theta_0,\gamma}(T_n(\theta_0) > \kappa_n)$. The exact and asymptotic sizes are defined as

$$\text{ExSZ}_n(\theta_0, \kappa_n) \equiv \sup_{\gamma \in \Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > \kappa_n)$$

$$\text{AsySz}(\theta_0, \kappa_n) \equiv \limsup_{n\to\infty} \text{ExSZ}_n(\theta_0, \kappa_n).$$

Note that the exact and asymptotic sizes of a test have the concept of uniformity built into their definitions in that $\text{ExSZ}_n(\theta_0, \kappa_n)$ is the largest NRP *uniformly* over the parameter space Γ and $\text{AsySz}(\theta_0, \kappa_n)$ is its limit. In order to have a test with approximately controlled exact size, and therefore controlled NRP at any $\gamma \in \Gamma$, we must control $\text{AsySz}(\theta_0, \kappa_n)$.

Under the frameworks of this paper, the primary theoretical step in controlling the asymptotic size of a test is to control the asymptotic NRP under all drifting sequences of parameters $\{\gamma_{n,h}\}$. That is, if we can find a (sequence of) CV(s) $\{\widetilde{\kappa}_n\}$ such that $\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > \widetilde{\kappa}_n) \leq \alpha$ for all $\{\gamma_{n,h}\}$ described in Section 2, we can construct a hypothesis test whose asymptotic size is bounded by α (see Andrews and Guggenberger, 2010b, Andrews et al., 2011 or the subsequencing arguments used in Appendix I for details). Since $c_h(1 - \alpha)$ (or $c_h^{(i)}(1 - \alpha)$, $i = 1, 2$) is the $(1 - \alpha)^{th}$ CV of the limit distribution of $T_n(\theta_0)$ under H_0 and the drifting sequence of parameters $\{\gamma_{n,h}\}$, we would ideally like to use a CV that is equal to $c_h(1 - \alpha)$ whenever $\{\gamma_{n,h}\}$ characterizes the true DGP in order to maximize the power of the resulting test while controlling its asymptotic size. Unfortunately, h cannot be consistently estimated under all drifting sequence DGPs. This has led to the construction of the so-called LF CV $\sup_{\mathfrak{h} \in H} c_{\mathfrak{h}}(1 - \alpha)$ and variants thereof (e.g., AG). Guarding against the worst-case drifting sequence DGP, this CV is often quite large, substantially reducing the power of the resulting test.

Though h cannot be consistently estimated under $\{\gamma_{n,h}\}$, in typical applications one can

find an estimator of h that converges in distribution to a random variable centered around the true value under H_0 and this drifting sequence DGP. This allows one to form asymptotically valid confidence sets for h and subsequently restrict attention to data-dependent regions inside of H relevant to the testing problem at hand, rather than guarding against the worst-case scenario, leading to smaller CVs and resulting tests with higher power. However, the additional uncertainty associated with the estimation of h must be taken into account for one to control the asymptotic NRP under all drifting sequences. This is where Bonferroni bounds become useful.

We now introduce two sets of three types of robust CVs based upon Bonferroni approaches. Each set corresponds to CVs to be used within either the SLLD or MLLDs framework. Within each set, the CVs are presented in increasing order of computational complexity. As the types of CV grow from least to most computationally complex, appropriately constructed tests using them tend to gain in power.

3.1 S-Bonf Robust Critical Values

We begin by examining Bonferroni-based size-corrected CVs for problems that are characterized by a SLLD, as described in Section 2.1. The first, most conservative but most computationally simple Bonferroni-based CV is defined as follows:

$$c_B^S(\alpha, \delta, \hat{h}_n) \equiv \sup_{\mathfrak{h} \in I_{\alpha-\delta}(\hat{h}_n)} c_{\mathfrak{h}}(1-\delta),$$

where $\delta \in [0, \alpha]$, $\hat{h}_n$ is some random vector taking value in an auxiliary space $\widetilde{H}$ and $I_{\alpha-\delta}(\cdot)$ is a correspondence from $\widetilde{H}$ into H. In applications, the space $\widetilde{H}$ will typically be equal to H or a space containing H but this is not necessary for the ensuing assumptions to hold. The random vector $\hat{h}_n$ is an estimator of h under H_0 and the DGP characterized by $\{\gamma_{n,h}\}$ and $I_{\alpha-\delta}(\hat{h}_n)$ serves as a $(\alpha - \delta)$-level confidence set for h. Construction of an estimator of h is typically apparent from the context of the testing problem, given that $h_1 = n^r \gamma_{n,h,1} + o(1)$ and $h_2 = \gamma_{n,h,2} + o(1)$. The S-Bonf CV generalizes the LF CV: when $I_0(x) = H$ for all $x \in \widetilde{H}$, $c_B^S(\alpha, \alpha, \hat{h}_n) = c_{LF}(\alpha) \equiv \sup_{\mathfrak{h} \in H} c_{\mathfrak{h}}(1-\alpha)$. The tuning parameter δ can be used to direct the power of the test towards different regions of the parameter space H.

Procedures using Bonferroni bounds in inference problems involving nuisance parameters and/or composite hypotheses have appeared in various contexts throughout the econometrics and statistics literature. Examples include Loh (1985), Stock (1991), Berger and Boos (1994), Silvapulle (1996), Staiger and Stock (1997), Romano and Wolf (2000), Hansen (2005b),

Moon and Schorfheide (2009), Chaudhuri and Zivot (2011) and Romano et al. (2012).[5] Romano et al. (2012), which applies a specific form of the S-Bonf CV to hypothesis testing in partially identified moment inequality models, was developed concurrently with this work. The methods of Loh (1985) and Hansen (2005b) are analogous to letting $\delta \to \alpha$ as $n \to \infty$ in the confidence set $I_{\alpha-\delta}(\hat{h}_n)$, but simply using the maximand $c_{\mathfrak{h}}(1-\alpha)$ in the construction of the CV. Though this approach is asymptotically size-correct since it leads to the use of $c_{LF}(\alpha)$ in the limit, it can have poor finite sample size control since it fails to fully account for the additional uncertainty involved in estimating h. We now impose further assumptions to ensure that tests utilizing c_B^S in the current context exhibit asymptotic size control.

Assumption S-B.3. *Consider some fixed $\beta \in [0,1]$. Under H_0 and when the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP for any fixed $h \in H$, there exists an estimator $\hat{h}_n$ taking values in some space $\widetilde{H}$ and a (nonrandom) continuous, compact-valued correspondence $I_\beta : \widetilde{H} \rightrightarrows H$ such that $\hat{h}_n \xrightarrow{d} \widetilde{h}$, a random vector taking values in $\widetilde{H}$ for which $P(h \in I_\beta(\widetilde{h})) \geq 1 - \beta$.*

Assumption S-B.3 assures that $\hat{h}_n$ is a well-behaved estimator of h under H_0 and $\{\gamma_{n,h}\}$ and imposes basic continuity assumptions on the correspondence $I_\beta(\cdot)$ used to construct the confidence set for h. It allows $I_\beta(\cdot)$ to take a variety of forms depending upon the context of the testing problem. For a given $\beta = \alpha - \delta$, this flexibility can be used to direct the power of the test towards different regions of H or to increase the computational tractability of constructing the S-Bonf-Min CVs (see, e.g., Romano et al., 2012). In a typical testing problem, under H_0 and $\{\gamma_{n,h}\}$, $\hat{h}_{n,1}$ converges weakly to a normally distributed random variable with mean h_1, the true localization parameter, making the construction of I_β very straightforward. Similarly, depending upon the testing problem, different choices of $\hat{h}_n$ may lead to tests with different power properties. For example, it may be advantageous to use $\hat{h}_n$ that imposes H_0 if this leads to smaller CVs.

Assumption S-B.4. *Under H_0 and when the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP, $(T_n(\theta_0), \hat{h}_n) \xrightarrow{d} (W_h, \widetilde{h})$ for all $h \in H$.*

Assumptions S-B.1 and S-B.3 already provide that $T_n(\theta_0) \xrightarrow{d} W_h$ and $\hat{h}_n \xrightarrow{d} \widetilde{h}$ under H_0 and $\{\gamma_{n,h}\}$ so that Assumption S-B.4 only ensures that this weak convergence occurs jointly. Assumptions S-B.3 and S-B.4 also allow for much flexibility in the estimation of h. Since

---

weak convergence is only required under H_0, $\hat{h}_n$ may be constructed to be an asymptotically biased estimator of h under some alternatives. We may now state the first result.

Theorem S-B. *Under Assumptions D and S-B.1 through S-B.4 for $\beta = \alpha - \delta$,*

$$\delta \leq \mathrm{AsySz}(\theta_0, c_B^S(\alpha, \delta, \hat{h}_n)) \leq \alpha.$$

Theorem S-B establishes the asymptotic size control of tests using S-Bonf CVs as well as providing a lower bound for the asymptotic size. If a consistent estimator of γ_2 is available, we can "plug" it into the CV to reduce its magnitude and increase the power of the test. Define the plug-in (PI) S-Bonf CV $c_{B-PI}^S(\alpha, \delta, \hat{h}_n)$ as above but with $I_{\alpha-\delta}(\hat{h}_n)_j = \{\hat{\gamma}_{n,2,j-p}\}$ for $j = p+1, \ldots, p+q$. This generalizes the PI LF CV (e.g., AG): when $I_0(x) = H_1 \times \{(x_{p+1}, \ldots, x_{p+q})\}$ for all $x \in \widetilde{H}$, $c_{B-PI}^S(\alpha, \alpha, \hat{h}_n) = c_{LF-PI}(\alpha, \hat{\gamma}_{n,2}) \equiv \sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, \hat{\gamma}_{n,2})}(1-\alpha)$. Similar comments to those regarding the choice of δ, $\hat{h}_n$ and $I_{\alpha-\delta}(\cdot)$ apply to this CV as well. We impose the following consistency assumption then state the analogous result for tests using PI S-Bonf CVs.

Assumption PI. $\hat{\gamma}_{n,2} - \gamma_{n,2} \xrightarrow{p} 0$ *for all sequences $\{\gamma_n\}$ with $\gamma_n \in \Gamma$ for all n.*

Corollary S-B-PI. *Under Assumptions D, S-B.1 through S-B.4 and PI for $\beta = \alpha - \delta$,*

$$\delta \leq \mathrm{AsySz}(\theta_0, c_{B-PI}^S(\alpha, \delta, \hat{h}_n) \leq \alpha.$$

3.2   S-Bonf-Adj Robust Critical Values

Though they are size-controlled, the S-Bonf CVs may be conservative, even in a uniform sense. This can be seen from the fact that we can only establish a lower bound of $\delta \leq \alpha$ on the asymptotic size of tests using these CVs. Instead of relying on Bonferroni bounds, for a given confidence set level, we can directly adjust the level of the localized quantile function according to the limit distribution of $(T_n(\theta_0, \hat{h}_n))$ to improve power as follows:

$$c_{B-A}^S(\alpha, \beta, \hat{h}_n) \equiv \sup_{\mathfrak{h} \in I_\beta(\hat{h}_n)} c_{\mathfrak{h}}(1 - \bar{\alpha}),$$

where $\beta \in [0, 1]$ and $\bar{\alpha} = \inf_{h \in H} \alpha(h)$ with $\alpha(h) \in [\alpha - \beta, \alpha]$ solving

$$P(W_h \geq \sup_{\mathfrak{h} \in I_\beta(\widetilde{h})} c_{\mathfrak{h}}(1 - \alpha(h))) = \alpha$$

or $\alpha(h) = \alpha$ if $P(W_h \geq \sup_{\mathfrak{h} \in I_\beta(\widetilde{h})} c_{\mathfrak{h}}(1 - \alpha)) < \alpha$. Unlike the S-Bonf CV, the level adjustment in this S-Bonf-Adj CV compensates for the limiting dependence between $T_n(\theta_0)$

and $\sup_{\mathfrak{h} \in I_\beta(\hat{h}_n)} c_{\mathfrak{h}}(\cdot)$. The level $\bar{\alpha}$ is an implicit function of the tuning parameter β, automatically adjusting to the user's choice. Clearly, for any given choice of $\delta \in [0, \alpha]$, $c_{B-A}^S(\alpha, \alpha - \delta, \hat{h}_n) \le c_B^S(\alpha, \delta, \hat{h}_n)$ so that a test using $c_{B-A}^S(\alpha, \alpha - \delta, \hat{h}_n)$ necessarily has higher power. Note that, like the S-Bonf CVs, the S-Bonf-Adj CVs also generalize the LF CV: when $I_0(x) = H$ for all $x \in \widetilde{H}$, $P(W_h \ge c_{LF}(\alpha)) = P(W_h \ge \sup_{\mathfrak{h} \in I_0(\tilde{h})} c_{\mathfrak{h}}(1 - \alpha)) \le \alpha$ for all $h \in H$ so that $\alpha(h) = \alpha$ for all $h \in H$ and $c_{B-A}^S(\alpha, \alpha, \hat{h}_n) = c_{LF}(\alpha)$.

To show correct asymptotic size of the S-Bonf-Adj CVs, we impose one additional condition which is essentially a continuity assumption.

Assumption S-BA. *(i)* $P(W_h = c_{B-A}^S(\alpha, \beta, \tilde{h})) = 0$ *for all* $h \in H$.
(ii) $P(W_{h^*} \ge c_{B-A}^S(\alpha, \beta, \tilde{h}^*)) = \alpha$ *for some* $h^* \in H$.

Assumption S-BA is analogous to Assumption Rob2 of Andrews and Cheng (2012) in the current general context. Part (ii) is not required for the asymptotic size to be bounded above by α; it is only used to show that the lower bound is also equal to α. Part (i) holds if $W_h - c_{B-A}^S(\alpha, \tilde{h})$ is a continuous random variable for all $h \in H$, as is typical in applications.

Theorem S-BA. *Under Assumptions D, S-B.1, S-B.2(i) evaluated at $\delta = \bar{\alpha}$, S-B.3, S-B.4 and S-BA,*

$$\text{AsySz}(\theta_0, c_{B-A}^S(\alpha, \beta, \hat{h}_n)) = \alpha.$$

We can also define a PI version of these CVs to further improve power as follows:

$$c_{B-A-PI}^S(\alpha, \beta, \hat{h}_n) \equiv \sup_{\mathfrak{h} \in I_\beta(\hat{h}_n)} c_{\mathfrak{h}}(1 - \bar{a}(\hat{\gamma}_{n,2})),$$

where $I_\beta(\hat{h}_n)_j = \{\hat{\gamma}_{n,2,j-p}\}$ for $j = p + 1, \ldots, p + q$, $\beta \in [0, \alpha]$ and $\bar{a} : H_2 \to [\underline{\delta}, \alpha - \bar{\delta}] \subset [\alpha - \beta, \alpha]$ is a level-adjustment function. For similar reasons to those given above, a test using $c_{B-A-PI}^S(\alpha, \alpha - \delta, \hat{h}_n)$ that satisfies the following assumption necessarily has higher power than one using $c_{B-PI}^S(\alpha, \delta, \hat{h}_n)$ and generalizes $c_{LF-PI}(\alpha, \hat{\gamma}_{n,2})$.

To show correct asymptotic size of tests using the PI S-Bonf-Adj CVs, we introduce a different continuity assumption.

Assumption S-BA-PI. *(i)* $\bar{a} : H_2 \to [\underline{\delta}, \alpha - \bar{\delta}]$ *is a continuous function.*
(ii) $P(W_h = c_{B-A-PI}^S(\alpha, \beta, \tilde{h})) = 0$ *for all* $h \in H$.
(iii) $P(W_h \ge \sup_{\mathfrak{h} \in I_\beta(\tilde{h})} c_{\mathfrak{h}}(1 - \bar{a}(h_2))) \le \alpha$ *for all* $h \in H$ *and there is some* $h^* \in H$ *such that* $P(W_{h^*} \ge c_{B-A-PI}^S(\alpha, \tilde{h}^*)) = \alpha$.

21

The second statement of part (iii) is not required for the asymptotic size to be bounded above by α; it is only used to show that the lower bound is also equal to α. As before, part (ii) holds if $W_h - c_{B-A-PI}(\alpha, \widetilde{h})$ is a continuous random variable, as is typical. Parts (i) and (iii) will hold if $\bar{a}(\cdot)$ is constructed appropriately. In practice, for a single given testing problem, one does not need to determine the entire function $\bar{a}(\cdot)$ since it will only be evaluated at a single point, $\hat{\gamma}_{n,2}$. Supposing $\bar{a}(\hat{\gamma}_{n,2})$ comes from evaluating a function satisfying Assumption S-BA-PI provides one with the asymptotic theoretical justification needed for correct asymptotic size and is consistent with practical implementation. For a given testing problem, practical determination of $\bar{a}(\cdot)$ that is consistent with part (iii) and yields the highest power proceeds by finding the largest $\bar{a} \in [\underline{\delta}, \alpha - \bar{\delta}]$ such that

$$\sup_{h_1 \in H_1} P(W_{(h_1, \gamma_{\hat{n},2})} \geq \sup_{\mathfrak{h} \in I_\beta((\widetilde{h}_1, \hat{\gamma}_{n,2}))} c_{\mathfrak{h}}(1 - \bar{a})) \leq \alpha. \tag{6}$$

Corollary S-BA-PI. *Under Assumptions D, S-B.1, S-B.3, S-B.4, PI, S-BM.1 and S-BA-PI,*

$$\mathrm{AsySz}(\theta_0, c^S_{B-A-PI}(\alpha, \beta, \hat{h}_n)) = \alpha.$$

The choice of β allows the practitioner the flexibility to direct power toward regions of the localization parameter space H. One heuristic way to direct power toward a given h (e.g., identification strength) is to use the PI S-Bonf-Adj CV with the β that minimizes the distance between $c_h(1 - \alpha)$ and $c^S_{B-A-PI}(\alpha, \beta, h)$ (assuming $E[\hat{h}_n] \approx h$). This will yield a S-Bonf-Adj CV that is "close" to the true localized quantile evaluated at h with high probability under drifting sequences $\{\gamma_{n,h}\}$. More generally, one may choose β to direct power towards regions of H according to a weighting scheme in analogy to maximizing weighted average power. For example, let F_H be a probability measure with support on H. One could select β to minimize $\int |c_h(1 - \alpha) - c^S_{B-A-PI}(\alpha, \beta, h)| dF_H(h).$[6]

3.3 S-Bonf-Min Robust Critical Values

Though β can be chosen to direct power in the construction of the S-Bonf-Adj CVs, different choices of β trade off power over different regions of the parameter space. (This is also true of δ in the S-Bonf CV.) Depending upon the choice of β, the S-Bonf-Adj CVs can lack power over large portions of the parameter space. In order to overcome this obstacle and produce tests that can simultaneously direct power while maintaining high power over most of the

---

[6]Recalling Jensen's inequality, one may prefer to choose β to minimize $|c_h(1 - \alpha) - E[c^S_{B-A-PI}(\alpha, \beta, \widetilde{h})]|$ or $\int |c_h(1 - \alpha) - E[c^S_{B-A-PI}(\alpha, \beta, \widetilde{h})]| dF_H(h)$.

parameter space, we introduce the following CV that minimizes over a set of S-Bonf and S-Bonf-Adj CVs:

$$c_{BM}^S(\alpha, \hat{h}_n) \equiv c_{min-B}(\alpha, \hat{h}_n) + \Delta,$$

where

$$c_{min-B}(\alpha, \hat{h}_n) = \min \left\{ c_{B-A}^S(\alpha, \beta_1, \hat{h}_n), \ldots, c_{B-A}^S(\alpha, \beta_r, \hat{h}_n), \inf_{\delta \in [\underline{\delta}, \alpha - \bar{\delta}]} c_B^S(\alpha, \delta, \hat{h}_n) \right\}$$

and $\Delta = \sup_{h \in H} \Delta(h)$ with $\Delta(h) \geq 0$ solving

$$P(W_h \geq c_{min-B}(\alpha, \widetilde{h}) + \Delta(h)) = \alpha$$

or $\Delta(h) = 0$ if $P(W_h \geq c_{min-B}(\alpha, \widetilde{h})) < \alpha$.

Here we refer to the input S-Bonf-Adj CVs $c_{B-A}^S(\alpha, \beta_i, \hat{h}_n)$ as the "anchors" and the corresponding values in the parameter space H for which the β_i are chosen to direct power toward as the "anchor points". Though computation of c_{BM}^S is more expensive than that of c_{B-A}^S, the anchors allow one to direct power while the minimum over S-Bonf-Min CVs component allows the test to achieve relatively high power over the entire parameter space H. The former is achieved by bringing the value of $c_{min-B-A}(\alpha, \hat{h}_n)$ "close" to that of $c_h(1-\alpha)$ with high probability at the anchor points. The latter is achieved by enabling $c_{min-B-A}(\alpha, h)$ to (conservatively) mimic the behavior of $c_h(1-\alpha)$ over the entire parameter space H. A simple choice of anchor is the LF CV.

The size-correction factor (SCF) Δ automatically adapts to the user's choices of confidence set correspondence $I_\beta(\cdot)$, localization parameter estimate $\hat{h}_n$ and the parameters $\underline{\delta}$, $\bar{\delta}$ and $\beta_1, \ldots, \beta_r$ since it is constructed so that the test has correct asymptotic size under H_0 for the given choice of these objects. Given the conservativeness already built into the Bonferroni approach, Δ will typically be quite small when the number of anchors is small, often being numerically indistinguishable from zero, so that $c_{BM}^S(\alpha, \hat{h}_n) \leq c_B^S(\alpha, \beta_i, \hat{h}_n), c_B^S(\alpha, \delta, \hat{h}_n)$ for $\delta \in [\underline{\delta}, \alpha - \bar{\delta}]$ and $i = 1, \ldots, r$, resulting in tests with higher power. In fact, in many contexts, simply setting $\Delta = 0$ will result in a test without perfect (asymptotic) size control, but very little size distortion and necessarily higher power. For some testing problems, size-control can be attained with Δ set exactly equal to zero. See Corollary SCF in Appendix III for a sufficient condition for this to occur.

Inherent in the construction of $c_{BM}^S(\alpha, \hat{h}_n)$ are the parameters $\underline{\delta}$ and $\bar{\delta}$. It should be emphasized that these parameters are *not* tuning parameters. They serve the role of restricting the localized quantile function $c_h(\cdot)$ and the confidence set $I_{\alpha-\delta}$ to regions of continuity (see

the discussion following Assumptions S-BM.1 and S-BM.2). Given that the aim of using S-Bonf-Min CVs is to maximize power, the minimization space $[\underline{\delta}, \alpha - \bar{\delta}]$ should be set as large as possible in order to obtain the smallest CV. This means that $\underline{\delta}$ and $\bar{\delta}$ should be set close to zero while satisfying Assumptions S-BM.1 and S-BM.2. Beyond computational complexity, there is a tradeoff involved in using too many anchor points (large r): the more anchor points one uses, the larger the size-correction factor Δ will be. This will lower power in portions of the parameter space that are far from the anchor points.

It is worth pointing out here the similarities that the use of S-Bonf-Min CVs shares with existing methods. In the specific contexts of inference in moment inequality models and inference robust to identification strength, Andrews and Soares (2010), Andrews and Barwick (2011) and Andrews and Cheng (2012) also use CVs that are functions of an inconsistent estimator of the localization parameter. Some of the size-correction techniques in these papers use a binary decision rule that uses as CV the localized quantile with $|h_1| = \infty$ once $|\hat{h}_{n,1}|$ crosses a sample-size dependent threshold and $c_{LF}(\alpha)$ otherwise. In contrast, the S-Bonf-Min CVs confine the range of h used to construct the test statistic to $I_{\alpha-\delta}(\hat{h}_n)$ and subsequently search for the smallest of these among admissible δ values (and anchors), without requiring an ad hoc threshold specification. In some sense this can be thought of as a "smoothed" version of these authors' methods. Cheng (2008) and Kabaila (1998) also employ CV selection methods based on binary decision rules for inference in the specific contexts of a weakly identified nonlinear regression model and conservative model selection.

Perhaps the most similar methodologies to those using S-Bonf-Min CVs are those that use a "transition function" to smooth between the LF CV and the one that obtains under infinite values of the localization parameter in a manner that depends upon an inconsistent estimate of the localization parameter. These methods are advocated in specific contexts by Andrews and Soares (2010) and Andrews and Barwick (2011) (using certain choices of their "ψ" function) and Andrews and Cheng (2012) (their type 2 robust CV). The transition function methods of Andrews and Barwick (2011) and Andrews and Cheng (2012) also necessitate the use of SCFs similar to Δ in the definition of c_{BM}^S. In contrast to these methods, the S-Bonf-Min CVs do not require an ad hoc choice of transition function. Rather, they adaptively use the full localized quantile function (rather than two points of it) and the data to "choose" which points along the localized quantile curves are relevant to the finite-sample testing problem at hand. This allows $c_{BM}^S(\alpha, \cdot)$ to closely mimic the true $(1-\alpha)^{th}$ localized quantile function. See Section 4.1 and Figures 1-2 for illustrations. This leads to asymptotic NRPs that are close to the asymptotic size of the test over wide ranges of h so that the tests are

nearly similar and consequently attain high power.

To show that tests utilizing $c_{BM}^S(\alpha, \hat{h}_n)$ have correct (non-conservative) asymptotic size, we strengthen Assumption S-B.3 as follows.

Assumption S-BM.2. *Consider some fixed $\alpha \in (0,1)$ and some pair $(\underline{\delta}, \bar{\delta}) \in [0, \alpha - \bar{\delta}] \times [0, \alpha - \underline{\delta}]$.*
(i) Assumption S-B.3 holds for all $\beta \in [\bar{\delta}, \alpha - \underline{\delta}]$.
(ii) $I_\beta(\mathfrak{h})$ is continuous in β over $[\bar{\delta}, \alpha - \underline{\delta}]$ for all $\mathfrak{h} \in \widetilde{H}$.

The quantity $\bar{\delta}$ serves as a lower bound and $\alpha - \underline{\delta}$ serves as an upper bound on the points β for which the correspondence $I_\beta(\mathfrak{h})$ must be continuous. In many applications, the confidence set I_β corresponds to a confidence set for a normal random variable. Hence, in order to satisfy Assumption S-B.3, we would need $I_0(x) = H$. However, this will typically involve a discontinuity of $I_\beta(\mathfrak{h})$ at $\beta = 0$ so that we must bound β from below by some value $\bar{\delta}$ greater than zero.

To show correct asymptotic size of tests using the S-Bonf-Min CVs, we introduce a new continuity assumption similar to S-BA.

Assumption S-BM.3. *(i) $P(W_h = c_{BM}^S(\alpha, \widetilde{h})) = 0$ for all $h \in H$.*
(ii) $P(W_{h^*} \geq c_{BM}^S(\alpha, \widetilde{h}^*)) = \alpha$ for some $h^* \in H$.

Similar comments to those following Assumption S-BA apply here. We provide an easy-to-verify sufficient condition for part (ii) to hold in Proposition S-BM of Appendix III.

Theorem S-BM. *Under Assumptions D, S-B.1, S-B.2(i) evaluated at the $\delta_i = \bar{\alpha}_i$ that corresponds to β_i, S-B.3 evaluated at β_i, S-B.4 and S-BM.1 through S-BM.3 for $i = 1, \ldots, r$,*

$$\text{AsySz}(\theta_0, c_{BM}^S(\alpha, \hat{h}_n)) = \alpha.$$

Depending upon the values of $\underline{\delta}$, $\bar{\delta}$ and β_i for $i = 1, \ldots, r$ used in the construction of $c_{BM}^S(\alpha, \hat{h}_n)$, some of the assumptions imposed in Theorem S-BM may be redundant. For example, Assumption S-B.2(i) evaluated at $\delta_i = \bar{\alpha}_i$ for $i = 1, \ldots, r$ is implied by Assumption S-BM.1 if $\bar{\alpha}_i \in [\underline{\delta}, \alpha - \bar{\delta}]$ for all $i = 1, \ldots, r$. Similarly, Assumption S-B.3 evaluated at β_i for $i = 1, \ldots, r$ is implied by Assumption S-BM.2 if $\beta_i \in [\bar{\delta}, \alpha - \underline{\delta}]$ for all $i = 1, \ldots, r$.

As with all of the CVs examined in this paper, if a consistent estimator of γ_2 is available, we may further improve power by using a PI version of the S-Bonf-Min CV:

$$c_{BM-PI}^S(\alpha, \hat{h}_n) \equiv c_{min-B-PI}(\alpha, \hat{h}_n) + \eta(\hat{\gamma}_{n,2}),$$

where

$$c_{min-B-PI}(\alpha, \hat{h}_n) = \min\left\{c_{B-A-PI}(\alpha, \beta_1, \hat{h}_n), \ldots, c_{B-A-PI}(\alpha, \beta_r, \hat{h}_n), \inf_{\delta \in [\underline{\delta}^m, \alpha - \bar{\delta}^m]} c^S_{B-PI}(\alpha, \delta, \hat{h}_n)\right\}$$

and $\eta : H_2 \to \mathbb{R}$. Note that, for each $i = 1, \ldots, r$ in the construction of anchors $c_{B-A-PI}(\alpha, \beta_i, \hat{h}_n)$, a different level-adjustment function $\bar{a}_i : H_2 \to [\underline{\delta}_i, \alpha - \bar{\delta}_i] \subset [\alpha - \beta_i, \alpha]$ may be used. Apart from using a PI estimator in the minimum-of-Bonferroni approach, the magnitude of the CV is further reduced and the power of the test is further increased by the use of a SCF that depends upon the PI estimator. This SCF function, satisfying the following assumption, is often numerically indistinguishable from zero in applications, making its practical relevance minimal as simply using $c_{min-B-PI}(\alpha, \hat{h}_n)$ often leads to minimal size-distortion. This again results from the conservativeness inherent in the Bonferroni approach. Similarly to the analogous SCF function used by Andrews and Barwick (2011), in order to maximize power one should select the smallest SCF function that satisfies the following assumption.

Assumption S-BM-PI. *(i)* $\eta : H_2 \to \mathbb{R}$ *is a continuous function.*
   (ii) $P(W_h = c_{min-B-PI}(\alpha, \widetilde{h}) + \eta(h_2)) = 0$ *for all* $h \in H$.
   (iii) $P(W_h \geq c_{min-B-PI}(\alpha, \widetilde{h}) + \eta(h_2)) \leq \alpha$ *for all* $h \in H$ *and there is some* $h^* \in H$ *such that* $P(W_{h^*} \geq c^S_{BM-PI}(\alpha, \widetilde{h}^*)) = \alpha$.

Part (i) is analogous to Assumption $\eta 1$ and parts (ii) and (iii) are analogous to Assumption $\eta 3$ of Andrews and Barwick (2011). As with the other PI CVs introduced in this paper, the SCF function η and level-adjustment functions $\bar{a}_i$ only need to be computed at the single value of $h_2 = \hat{\gamma}_{n,2}$ in practice. Similar comments to those following Assumption S-BA-PI apply.

We may now present the result establishing the correct asymptotic size of tests utilizing this PI CV.

Corollary S-BM-PI. *Under Assumptions D, PI, S-B.1, S-B.3 evaluated at β_i, S-B.4, PI, S-BA-PI(i) and (iii) evaluated at β_i and corresponding $\bar{a}_i(\cdot)$, S-BM.1 and S-BM.2 evaluated at pairs $(\underline{\delta}^m, \bar{\delta}^m)$ and $(\underline{\delta}_i, \bar{\delta}_i)$ and S-BM-PI for $i = 1, \ldots, r$,*

$$\text{AsySz}(\theta_0, c^S_{BM-PI}(\alpha, \hat{h}_n)) = \alpha.$$

Depending upon the values of $(\underline{\delta}^m, \bar{\delta}^m)$, $(\underline{\delta}_i, \bar{\delta}_i)$ and β_i for $i = 1, \ldots, r$ used in the construction of $c_{BM-A-PI}(\alpha, \hat{h}_n)$, (part of) Assumption S-B.3 evaluated at β_i for $i = 1, \ldots, r$ may be redundant. For example, it is implied by Assumption S-BM.2 evaluated at pairs

26

$(\underline{\delta}^m, \bar{\delta}^m)$ and $(\underline{\delta}_i, \bar{\delta}_i)$ for $i = 1, \dots, r$ if $\beta_i \in \bigcup_{j=1}^{r} [\underline{\delta}_j, \bar{\delta}_j] \cup [\underline{\delta}^m, \bar{\delta}^m]$ for $i = 1, \dots, r$. It is also worth noting that the full force of Assumption S-BA-PI(iii) is not necessary for the result of Corollary S-BM-A-PI to hold. The assumption is imposed as a guide for the anchoring procedure to direct power appropriately. As was the case for the SCF Δ above, the addition of $\eta(\hat{\gamma}_{n,2})$ is not always necessary to obtain correct asymptotic size. Corollary SCF in Appendix III provides a precise condition under which it may be dispensed with.

3.4 M-Bonf Robust Critical Values

The S-Bonf robust CVs that apply to problems with a SLLD can be adapted to handle those with MLLDs. In order to do this we need a method to choose which localized limit distribution $J_h^{(1)}$ or $J_h^{(2)}$ to construct the CVs from. For this, we utilize an estimator of $\zeta(\{\gamma_{n,h}\})$ that tells us with increasing precision which region of H_1 that $\zeta(\{\gamma_{n,h}\})$ lies within. Construction of this estimator can be deduced from the form of $\gamma_{n,h,1}/k_n$. In applications, $\hat{\zeta}$ is typically equal to $n^{-r}\hat{h}_{n,1}$ scaled by k_n^{-1} since $\gamma_{n,h,1} \approx n^{-r}h_1$ and $\zeta(\{\gamma_{n,h}\}) \approx \gamma_{n,h,1}/k_n$. The M-Bonf CV for the MLLDs framework is thus defined as follows:

$$
c_B^M(\alpha, \delta, \hat{h}_n) \equiv \begin{cases} c_B^{(1)}(\alpha, \delta, \hat{h}_n), & \text{if } \hat{\zeta} \in \bar{K} \\ c_B^{(2)}(\alpha, \delta, \hat{h}_n), & \text{if } \hat{\zeta} \in \bar{L} \\ \max\{c_B^{(1)}(\alpha, \delta, \hat{h}_n), c_B^{(2)}(\alpha, \delta, \hat{h}_n)\}, & \text{if } \hat{\zeta} \in \bar{K}^c \cap \bar{L}^c, \end{cases}
$$

where

$$
c_B^{(i)}(\alpha, \delta, \hat{h}_n) \equiv \sup_{\mathfrak{h} \in I_{\alpha-\delta}(\hat{h}_n)} c_{\mathfrak{h}}^{(i)}(1 - \delta) + \varepsilon_i
$$

for $i = 1, 2$, $\bar{K} \subset \text{int}(K)$ is closed and $\bar{L} \subset \text{int}(L)$ is closed. Asymptotic power is highest for sets $\bar{K}$ and $\bar{L}$ that are very "close" to K and L. This makes the region using the most conservative CV in the construction, $\bar{K}^c \cap \bar{L}^c$, relatively small. Similarly, ε_i for $i = 1, 2$ should be chosen to be the smallest feasible value that satisfies Assumption M-B.2(ii) in order to maximize power

Unlike existing approaches, adapting the Bonferroni approach to the MLLDs framework allows one to size correct the CVs that pertain to the typical type of testing problem that falls into this framework without killing the power of the test. This type of problem entails (at least) one of the localized distribution quantiles being equal to ∞ for some $h \in H$ so that approaches putting positive weight on LF CVs (such as those of AG, Andrews and Soares, 2010, Andrews and Cheng, 2012 and Andrews and Barwick, 2011) will lead to infinite CVs and zero power over much or all of the parameter space.

We now impose additional assumptions under which tests utilizing M-Bonf CVs have correct asymptotic size. Some of these are obvious counterparts to those imposed for S-Bonf CVs.

Assumption M-B.3. *Consider some fixed $\beta \in [0,1]$. Under H_0 and when the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP for any fixed $h \in H$, there exists an estimator $\hat{h}_n$ taking values in some space $\widetilde{H}$ and a (nonrandom) continuous, compact-valued correspondence $I_\beta : \widetilde{H} \rightrightarrows H$ such that $\hat{h}_n \overset{d}{\longrightarrow} \widetilde{h}$, a random vector taking values in $\widetilde{H}$ for which $P(h \in I_\beta(\widetilde{h})) \geq 1 - \beta$ and $I_\beta(\hat{h}_n), I_\beta(\widetilde{h}) \subset \bar{H}$ wp 1.*

The condition that $I_\beta(\hat{h}_n), I_\beta(\widetilde{h}) \subset \bar{H}$ wp 1 if $h \in \bar{H}$ is not restrictive and can be ensured by the proper construction of the confidence set for h. In the typical application, it states that if the entries of h are finite, the confidence set for h does not include infinite values. Similar remarks to those following Assumption S-B.3 apply here as well.

Assumption M-B.4. *Under H_0 and when the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP for any $h \in H$,*
(i) if $\zeta(\{\gamma_{n,h}\}) \in K$, $(T_n(\theta_0), \hat{h}_n) \overset{d}{\longrightarrow} (W_h^{(1)}, \widetilde{h})$
(ii) if $\zeta(\{\gamma_{n,h}\}) \in L$, $(T_n(\theta_0), \hat{h}_n) \overset{d}{\longrightarrow} (W_h^{(2)}, \widetilde{h})$.

Joint convergence does not need to be established for $\zeta(\{\gamma_{n,h}\}) \in L^c \cap K^c$. Similar remarks to those following Assumption S-B.4 apply here as well.

Assumption M-B.5. $\hat{\zeta} \overset{p}{\longrightarrow} \zeta(\{\gamma_{n,h}\})$ *under H_0 and the drifting sequence of parameters* $\{\gamma_{n,h}\}$.

Under Assumption M-B.5, the estimator $\hat{\zeta}$ yields the decision rule that "chooses" which Bonferroni CV is applicable under any given sequence $\{\gamma_{n,h}\}$ as the sample size increases. Recall that $\hat{\zeta}$ is typically equal to $k_n^{-1} n^{-r} \hat{h}_{n,1}$. Furthermore, $\hat{h}_{n,1}$ is typically asymptotically centered about h_1 under H_0 so that for finite h_1, $\hat{h}_{n,1} = h_1 + O_p(1) = n^r \gamma_{n,h,1} + O_p(1)$ and since typically $k_n n^r \to \infty$,

$$\hat{\zeta} = k_n^{-1} n^{-r}(n^r \gamma_{n,h,1} + O_p(1)) = k_n^{-1} \gamma_{n,h,1} + o_p(1) = \zeta(\{\gamma_{n,h}\}) + o_p(1).$$

This why $\zeta(\{\gamma_{n,h}\})$ consistently estimable even though h_1 is not.

As in the SLLD case, we can use a consistent estimator of γ_2 to increase the power of tests using M-Bonf robust CVs. Define the PI M-Bonf CV $c_{B-PI}^M(\alpha, \delta, \hat{h}_n)$ the same as $c_B^M(\alpha, \delta, \hat{h}_n)$

but with $I_{\alpha-\delta}(\hat{h}_n)_j = \{\hat{\gamma}_{n,2,j-p}\}$ for $j = p+1, \ldots, p+q$. For $i = 1, 2$, also define $c_{B-PI}^{(i)}(\alpha, \delta, \hat{h}_n)$ identically to $c_B^{(i)}(\alpha, \delta, \hat{h}_n)$ but with $I_{\alpha-\delta}(\hat{h}_n)_j = \{\hat{\gamma}_{n,2,j-p}\}$ for $j = p+1, \ldots, p+q$.

To conserve space, we relegate and additional high-level assumption used to provide an upper bound on the NRP under drifting sequences of parameters for which $h \in \bar{H}^c$ and assumptions used to establish a lower bound on the size of a test using M-Bonf and PI M-Bonf CVs to Appendix III. We furthermore refer the reader to Lemmas M-B and M-B-PI in Appendix I for the formal statements and proofs that

$$\mathrm{AsySz}(\theta_0, c_B^M(\alpha, \delta, \hat{h}_n)), \mathrm{AsySz}(\theta_0, c_{B-PI}^M(\alpha, \delta, \hat{h}_n)) \leq \alpha$$

under appropriate conditions, as well as a brief discussion of a sufficient condition.

The size-corrected CVs $c_B^M(\alpha, \delta, \hat{h}_n)$ and $c_{B-PI}^M(\alpha, \delta, \hat{h}_n)$ simultaneously accommodate problems of non-uniformity in the point-wise asymptotics of the distribution of the test statistic (Assumption D), localized null limit distributions that depend on the localization parameter discontinuously (Assumption M-B.1), localized null limit distributions that are discontinuous (Assumption M-B.2(ii)) and localized null limit distributions that "escape" to $\pm\infty$ (Assumption M-B.6 in Appendix III).

3.5   M-Bonf-Adj Robust Critical Values

Like their S-Bonf counterparts, the M-Bonf CVs may lead to conservative testing in a uniform sense and can be level-adjusted to improve power. Specifically, define the M-Bonf-Adj CV, $c_{B-A}^M(\alpha, \beta, \hat{h}_n)$, identically to the M-Bonf CV but replace "$c_B^{(i)}(\alpha, \delta, \hat{h}_n)$" in the definition with "$c_{B-A}^{(i)}(\alpha, \delta, \hat{h}_n)$" for $i = 1, 2$, where $\beta \in [0, 1]$,

$$c_{B-A}^{(i)}(\alpha, \beta, \hat{h}_n) = \sup_{\mathfrak{h} \in I_\beta(\hat{h}_n)} c_{\mathfrak{h}}^{(i)}(1 - \bar{\alpha}^{(i)})$$

and $\bar{\alpha}^{(i)} = \inf_{h \in \bar{H}} \alpha^{(i)}(h)$ with $\alpha^{(i)}(h) \in [0, \alpha]$ solving

$$P(W_h^{(i)} \geq \sup_{\mathfrak{h} \in I_\beta(\tilde{h})} c_{\mathfrak{h}}^{(i)}(1 - \alpha^{(i)}(h))) = \alpha$$

or $\alpha(h) = \alpha$ if $P(W_h^{(i)} \geq \sup_{\mathfrak{h} \in I_\beta(\tilde{h})} c_{\mathfrak{h}}^{(i)}(1 - \alpha)) < \alpha$ for $i = 1, 2$.

Analogous to the S-Bonf-Adj CV case, we can use a PI version of the M-Bonf-Adj CV to improve power when a consistent estimator of γ_2 is available. Define the PI M-Bonf-Adj CV, $c_{B-A-PI}^M(\alpha, \beta, \hat{h}_n)$, identically to the PI M-Bonf CV but replace "$c_{B-PI}^{(i)}(\alpha, \delta, \hat{h}_n)$" with "$c_{B-A-PI}^{(i)}(\alpha, \delta, \hat{h}_n)$" for $i = 1, 2$, where $\beta \in [0, 1]$, $I_\beta(\hat{h}_n)_j = \{\hat{\gamma}_{n,2,j-p}\}$ for $j = p+1, \ldots, p+q$,

$$c_{B-A-PI}^{(i)}(\alpha, \beta, \hat{h}_n) = \sup_{\mathfrak{h} \in I_\beta(\hat{h}_n)} c_{\mathfrak{h}}^{(i)}(1 - \bar{a}^{(i)}(\hat{\gamma}_{n,2}))$$

29

and $\bar{a}^{(i)} : H_2 \to [\underline{\delta}^{(i)}, \alpha - \bar{\delta}^{(i)}]$ for $i = 1, 2$.

Also in analogy with the S-Bonf-Adj CV, computation of $c^M_{B-A}(\alpha, \alpha - \delta, \hat{h}_n)$ ($c^M_{B-A-PI}(\alpha, \alpha - \delta, \hat{h}_n)$) is costlier than that of $c^M_B(\alpha, \delta, \hat{h}_n)$ ($c^M_{B-PI}(\alpha, \delta, \hat{h}_n)$) but results in tests with higher power. As is the case for (PI) S-Bonf-Adj CVs, the choice of β in the construction of the (PI) M-Bonf-Adj CVs provides the practitioner with the ability to direct the power of the resulting tests toward different regions of H. Straightforward generalization allows for different β's to be chosen to correspond to the different localized limit distributions used in the construction of the CVs, further enhancing this flexibility. The formal statements of the conditions under which

$$\text{AsySz}(\theta_0, c^M_{B-A}(\alpha, \beta, \hat{h}_n)), \text{AsySz}(\theta_0, c^M_{B-A-PI}(\alpha, \beta, \hat{h}_n)) = \alpha$$

may be found in Corollaries M-BA and M-BA-PI located in Appendix III. Additional assumptions imposed in these corollaries may also be found there.

3.6 M-Bonf-Min Robust Critical Values

As is the case for S-Bonf and S-Bonf-Adj CVs in the SLLD framework, the M-Bonf and M-Bonf-Adj CVs can be appropriately minimized to maintain high power over most of the parameter space while retaining correct asymptotic size and directing power. For this, consider the Bonf-Min CVs adapted to the MLLDs framework:

$$c^M_{BM}(\alpha, \hat{h}_n) \equiv \begin{cases} c^{(1)}_{min-B}(\alpha, \hat{h}_n) + \Delta_1, & \text{if } \hat{\zeta} \in \bar{K} \\ c^{(2)}_{min-B}(\alpha, \hat{h}_n) + \Delta_2, & \text{if } \hat{\zeta} \in \bar{L} \\ \max\{c^{(1)}_{min-B}(\alpha, \hat{h}_n) + \Delta_1, c^{(2)}_{min-B}(\alpha, \hat{h}_n) + \Delta_2\}, & \text{if } \hat{\zeta} \in \bar{K}^c \cap \bar{L}^c, \end{cases}$$

where for $i = 1, 2$,

$$c^{(i)}_{min-B}(\alpha, \hat{h}_n) = \min\left\{ c^{(i)}_{B-A}(\alpha, \beta_1, \hat{h}_n), \ldots, c^{(i)}_{B-A}(\alpha, \beta_r, \hat{h}_n), \inf_{\delta \in [\underline{\delta}^{(i)}, \alpha - \bar{\delta}^{(i)}]} c^{(i)}_B(\alpha, \delta, \hat{h}_n) \right\},$$

and $\Delta_i = \sup_{h \in \bar{H}} \Delta_i(h)$ with $\Delta_i(h) \geq 0$ solving $P(W^{(i)}_h \geq c^{(i)}_{min-B}(\alpha, \widetilde{h}) + \Delta_i(h)) = \alpha$ or $\Delta_i(h) = 0$ if $P(W^{(i)}_h \geq c^{(i)}_{min-B}(\alpha, \widetilde{h}) + \Delta_i(h)) < \alpha$ for $h \in \bar{H}$. Aside from a couple of generalizing adjustments, $c^M_{BM}(\alpha, \hat{h}_n)$ is essentially a version of $c^S_{BM}(\alpha, \hat{h}_n)$ that is selected by $\hat{\zeta}$. Hence, very similar comments to those made about $c^S_{BM}(\alpha, \hat{h}_n)$ and the objects used in its construction apply for $c^M_{BM}(\alpha, \hat{h}_n)$ (see Section 3.3). We also impose some analogous assumptions to those used in the SLLD framework.

Assumption M-BM.2. *Consider some fixed $\alpha \in (0,1)$ and some pairs $(\underline{\delta}^{(i)}, \bar{\delta}^{(i)}) \in [0, \alpha - \bar{\delta}^{(i)}] \times [0, \alpha - \underline{\delta}^{(i)}]$ for $i = 1, 2$.*

(i) Assumption M-B.3 holds for all $\beta \in [\bar{\delta}^{(1)}, \alpha - \underline{\delta}^{(1)}] \cup [\bar{\delta}^{(2)}, \alpha - \underline{\delta}^{(2)}]$.

(ii) $I_\beta(\mathfrak{h})$ is continuous in β over $[\bar{\delta}^{(1)}, \alpha - \underline{\delta}^{(1)}] \cup [\bar{\delta}^{(2)}, \alpha - \underline{\delta}^{(2)}]$ for all $\mathfrak{h} \in \widetilde{H}$.

Analogous comments to those following Assumption S-BM.2 apply here.

Assumption M-BM.3. *(i) $P(W_h^{(i)} = c_{min-B}^{(i)}(\alpha, \widetilde{h}) + \Delta_i) = 0$ for all $h \in \bar{H}$ and $i = 1, 2$.*

(ii) Either (a) there is some $h^{*(1)} \in H$ and $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \text{int}(\bar{K})$ such that $\liminf_{n \to \infty} P_{\theta_0, \gamma_{n,h^{*(1)}}}(T_n(\theta_0) > c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1) = \alpha$ or (b) there is some $h^{*(2)} \in H$ and $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \text{int}(\bar{L})$ such that $\liminf_{n \to \infty} P_{\theta_0, \gamma_{n,h^{*(2)}}}(T_n(\theta_0) > c_{min-B}^{(2)}(\alpha, \hat{h}_n) + \Delta_2) = \alpha$.

Similar comments to those following Assumption S-BM.3 apply here as well. Part (ii) is not required for the asymptotic size to be bounded above by α; it is only used to show that the lower bound is also equal to α. Proposition M-BM in Appendix III contains a sufficient condition for part (ii) to hold.

Assumption M-BM.4. *Consider some fixed $\alpha \in (0,1)$ and some pairs $(\underline{\delta}^{(i)}, \bar{\delta}^{(i)}) \in [0, \alpha - \bar{\delta}^{(i)}] \times [0, \alpha - \underline{\delta}^{(i)}]$ for $i = 1, 2$. Consider any $h \in \bar{H}^c$.*

(i) For any finite n and $i = 1, 2$, $P_{\theta_0, \gamma_{n,h}}(|c_{min-B}^{(i)}(\alpha, \hat{h}_n) + \Delta_i| < \infty) = 1$.

(ii) If $\zeta(\{\gamma_{n,h}\}) \in K$, then $\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1) \le \alpha$. If $\zeta(\{\gamma_{n,h}\}) \in L$, then $\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_{min-B}^{(2)}(\alpha, \hat{h}_n) + \Delta_2) \le \alpha$.

(iii) If $\zeta(\{\gamma_{n,h}\}) \in L^c \cap K^c$, there are some $\{\widetilde{\gamma}_{n,h}^{(1)}\}, \{\widetilde{\gamma}_{n,h}^{(2)}\} \subset \Gamma$ such that $\zeta(\{\widetilde{\gamma}_{n,h}^{(1)}\}) \in K$ and $\zeta(\{\widetilde{\gamma}_{n,h}^{(2)}\}) \in L$ and

$$\text{plim}_{n \to \infty} \frac{c_{min-B}^{(i)}(\alpha, \hat{h}_n(\widetilde{\gamma}_{n,h}^{(i)}))}{c_{min-B}^{(i)}(\alpha, \hat{h}_n(\gamma_{n,h}^{(i)}))} \le 1$$

wp 1 for $i = 1, 2$, where $\hat{h}_n(\gamma_{n,h})$ ($\hat{h}_n(\widetilde{\gamma}_{n,h}^{(i)})$) denotes the estimator of Assumption B.3 when H_0 and the drifting sequence of parameters $\{\gamma_{n,h}\}$ ($\{\widetilde{\gamma}_{n,h}^{(i)}\}$) characterize the true DGP.

Assumption M-BM.4 consists of high-level conditions that are not difficult to verify in the typical application. It is used to ensure an upper bound on the NRP under drifting sequences of parameters for which $h \in \bar{H}^c$, that is, cases for which entries of h_1 are infinite. By properly constructing the confidence set $I_{\alpha-\delta}(\cdot)$, part (i) can be shown to hold since in applications $\hat{h}_n$ has finite entries in finite samples and the localized quantiles are finite for $h \in \bar{H}$. Part (ii) must be verified directly. In the applications we have encountered,

this is not difficult since the M-Bonf-Min CVs have closed-form solutions. In applications, the sequences $\{\widetilde{\gamma}_{n,h}^{(i)}\}$ in part (iii) can be obtained by simply scaling $\{\gamma_{n,h}\}$ appropriately for $i = 1, 2$. The probability limit condition is typically not difficult to verify because the CVs have closed form solutions and/or they are invariant to differences in drifting sequences with the same localization parameter (leading to a ratio of exactly one in the limit). See the verification of the similar assumption, Assumption M-BM-PI.2 for the consistent model selection example in Appendix II for illustrative details.

We may now establish the correct size of tests using M-Bonf-Min CVs.

Theorem M-BM. *Under Assumptions D and M-B.1, M-B.2(i) evaluated at $\delta_i^{(1)} = \bar{\alpha}_i^{(1)}$ corresponding to β_i and $c_h^{(1)}$ and $\delta_i^{(2)} = \bar{\alpha}_i^{(2)}$ corresponding to β_i and $c_h^{(2)}$, either (a) M-B.3 evaluated at β_i or (b) $c_{B-A}^{(j)}(\alpha, \beta_i, h)$ is invariant to h for $j = 1, 2$, M-B.4, M-B.5 and M-BM.1 through M-BM.4 for $i = 1, \ldots, r$,*

$$\text{AsySz}(\theta_0, c_{BM}^M(\alpha, \hat{h}_n)) = \alpha.$$

A similar sufficient condition to the one discussed following Lemma M-B in Appendix I applies here as well.

We at last consider the PI version of the M-Bonf-Min robust CVs to further increase power while retaining correct asymptotic size in this context:

$$c_{BM-PI}^M(\alpha, \hat{h}_n) \equiv \begin{cases} c_{min-B-PI}^{(1)}(\alpha, \hat{h}_n) + \eta_1(\hat{\gamma}_{n,2}), & \text{if } \hat{\zeta} \in \bar{K} \\ c_{min-B-PI}^{(2)}(\alpha, \hat{h}_n) + \eta_2(\hat{\gamma}_{n,2}), & \text{if } \hat{\zeta} \in \bar{L} \\ \max\{c_{min-B-PI}^{(1)}(\alpha, \hat{h}_n) + \eta_1(\hat{\gamma}_{n,2}), c_{min-B-PI}^{(2)}(\alpha, \hat{h}_n) + \eta_2(\hat{\gamma}_{n,2})\}, & \text{if } \hat{\zeta} \in \bar{K}^c \cap \bar{L}^c, \end{cases}$$

where

$$c_{min-B-PI}^{(i)}(\alpha, \hat{h}_n) = \min \left\{ c_{B-A-PI}^{(i)}(\alpha, \beta_1, \hat{h}_n), \ldots, c_{B-A-PI}^{(i)}(\alpha, \beta_r, \hat{h}_n), \inf_{\delta \in [\underline{\delta}^{(i)}, \alpha - \bar{\delta}^{(i)}]} c_{B-PI}^{(i)}(\alpha, \delta, \hat{h}_n) \right\},$$

and $\eta_i : H_2 \to \mathbb{R}$ for $i = 1, 2$. Analogous remarks to those made about $c_{BM-PI}^S(\alpha, \hat{h}_n)$ apply. We now adapt some of the previous assumptions to the present framework and establish the correct asymptotic size of tests using this PI CV.

Assumption M-BM-PI.1. *(i) $\eta_i : H_2 \to \mathbb{R}$ is a continuous function for $i = 1, 2$.*
 (ii) $P(W_h^{(i)} = c_{min-B-PI}^{(i)}(\alpha, \widetilde{h}) + \eta_i(h_2)) = 0$ for all $h \in \bar{H}$ and $i = 1, 2$.
 (iii) $P(W_h^{(i)} \geq c_{min-B-PI}(\alpha, \widetilde{h}) + \eta_i(h_2)) \leq \alpha$ for all $h \in \bar{H}$ and $i = 1, 2$.

(iv) Either (a) there is some $h^{*(1)} \in H$ and $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \text{int}(\bar{K})$ such that $\liminf_{n\to\infty} P_{\theta_0,\gamma_{n,h^{*(1)}}}(T_n(\theta_0) > c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n) + \eta_1(\hat{\gamma}_{n,2})) = \alpha$ or (b) there is some $h^{*(2)} \in H$ and $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \text{int}(\bar{L})$ such that $\liminf_{n\to\infty} P_{\theta_0,\gamma_{n,h^{*(2)}}}(T_n(\theta_0) > c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n) + \eta_2(\hat{\gamma}_{n,2})) = \alpha$.

Parts (i)-(iii) are the MLLDs counterparts to Assumption S-BM-PI(i)-(iii) and thus may be considered analogously. As in the case of Assumption M-BM.3(ii), we can provide a sufficient condition for part (iv) to hold that may be easier to verify. See Proposition M-BM-PI in Appendix III for details.

Assumption M-BM-PI.2. *Consider some fixed $\alpha \in (0,1)$ and some pairs $(\underline{\delta}^{(i)}, \bar{\delta}^{(i)}) \in [0, \alpha - \bar{\delta}^{(i)}] \times [0, \alpha - \underline{\delta}^{(i)}]$ for $i = 1, 2$. Consider any $h \in \bar{H}^c$.*

(i) For any finite n and $i = 1, 2$, $P_{\theta_0,\gamma_{n,h}}(|c^{(i)}_{min-B-PI}(\alpha, \hat{h}_n)| < \infty) = 1$.

(ii) If $\zeta(\{\gamma_{n,h}\}) \in K$, then $\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n) + \eta_1(\hat{\gamma}_{n,2})) \leq \alpha$. If $\zeta(\{\gamma_{n,h}\}) \in L$, then $\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n) + \eta_2(\hat{\gamma}_{n,2})) \leq \alpha$.

(iii) If $\zeta(\{\gamma_{n,h}\}) \in L^c \cap K^c$, there are some $\{\tilde{\gamma}^{(1)}_{n,h}\}, \{\tilde{\gamma}^{(2)}_{n,h}\} \subset \Gamma$ such that $\zeta(\{\tilde{\gamma}^{(1)}_{n,h}\}) \in K$ and $\zeta(\{\tilde{\gamma}^{(2)}_{n,h}\}) \in L$ and

$$\text{plim}_{n\to\infty} \frac{c^{(i)}_{min-B-PI}(\alpha, \hat{h}_n(\tilde{\gamma}^{(i)}_{n,h}))}{c^{(i)}_{min-B-PI}(\alpha, \hat{h}_n(\gamma^{(i)}_{n,h}))} \leq 1$$

wp 1 for $i = 1, 2$, where $\hat{h}_n(\gamma_{n,h})$ $(\hat{h}_n(\tilde{\gamma}^{(i)}_{n,h}))$ denotes the estimator of Assumption B.3 when H_0 and the drifting sequence of parameters $\{\gamma_{n,h}\}$ $(\{\tilde{\gamma}^{(i)}_{n,h}\})$ characterize the true DGP.

This assumption is the direct adaption of Assumption M-BM.4 to the PI version of the M-Bonf-Min CVs. We may now establish the correct asymptotic size of the PI test.

Corollary M-BM-PI. *Under Assumptions D, PI, M-B.1, either (a) M-B.3 evaluated at β_i or (b) $c^{(j)}_{B-A}(\alpha, \beta_i, h)$ is invariant to h_1 for $j = 1, 2$ and $i = 1, \ldots, r$, M-B.4, M-B.5, M-BA-PI(i) and (iii) evaluated at β_i and corresponding $\bar{a}^{(1)}_i(\cdot)$, $\bar{a}^{(2)}_i(\cdot)$, M-BM.1, M-BM.2, M-BM-PI.1 and M-BM-PI.2,*

$$\text{AsySz}(\theta_0, c^M_{BM-PI}(\alpha, \hat{h}_n)) = \alpha.$$

Assumption M-BA-PI is given in Appendix III. A straightforward adaptation of the sufficiency conditions provided in Corollary SCF (in Appendix III) to the MLLDs context characterizes a class of problems for which setting Δ_i $(\eta_i(\hat{\gamma}_{n,2}))$ to zero for $i = 1$ and/or 2 yields a test with correct asymptotic size.

4 Testing After Conservative Model Selection

We now illustrate a way to construct the objects used in $c_{BM-PI}^S(\alpha, \hat{h}_n)$ for the hypothesis testing problem after conservative model selection introduced in Section 2.1.1. We also show how the remainder of the assumptions imposed in Corollary S-BM-PI are satisfied for this problem, enabling us to use a PI S-Bonf-Min CV to conduct a test with correct asymptotic size.

First, define $\hat{h}_n = (\hat{h}_{n,1}, \hat{\gamma}_{n,2})$, where

$$\hat{h}_{n,1} = \frac{\sqrt{n}\widetilde{\beta}_2}{\hat{\sigma}(n^{-1}X_2^{*\prime}M_{[X_1^*:X_3^*]}X_2^*)^{-1/2}} \text{ with } \widetilde{\beta}_2 \equiv (X_2^{*\prime}M_{X_3^*}X_2^*)^{-1}X_2^{*\prime}M_{X_3^*}(Y - X_1^*\theta_0)$$

($\widetilde{\beta}_2$ is the restricted least squares estimator of β_2, imposing H_0) and $\hat{\gamma}_{n,2}$ can be defined as in AG as

$$\hat{\gamma}_{n,2} = \frac{-n^{-1}\sum_{i=1}^n x_{1i}x_{2i}}{\left(n^{-1}\sum_{i=1}^n x_{1i}^2 n^{-1}\sum_{i=1}^n x_{2i}^2\right)^{1/2}}$$

with $\{(x_{1i}, x_{2i})\}$ being the residuals from the regressions of x_{ji}^* on x_{3i}^* for $j = 1, 2$.[7] Second, let $\widetilde{H} = H$ and define the correspondence $I_\beta : \widetilde{H} \rightrightarrows H$ as follows:

$$I_\beta(\mathfrak{h}) = \begin{cases} (\mathfrak{h}_1, \mathfrak{h}_2), & \text{if } \mathfrak{h}_1 = \pm\infty \\ ([\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_\xi, \mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_{1-\beta+\xi}], \mathfrak{h}_2), & \text{if } \mathfrak{h}_1 \in \mathbb{R}, \end{cases}$$

where $\xi = \xi(\beta) \in (0, \beta)$ is a continuous function of β, $\mathfrak{h}_1 \in H_1$, $\mathfrak{h}_2 \in H_2$ and z_b denotes the bth quantile of the standard normal distribution.[8] The function $\xi(\beta)$ can be chosen to direct power against certain regions of H_1 (at the expense of others). The agnostic, and perhaps intuitive, choice of $\xi(\beta)$ would be simply $\xi(\beta) = \beta/2$. This is not the only configuration of I_β that will satisfy Assumption S-BM.2. For example, if the practitioner has a priori knowledge on the sign that β_2 takes, this can be incorporated into the analysis to improve power (see the next example).

With these definitions in hand, we may verify the remaining assumptions for some chosen significance level $\alpha \in (0, 1)$. To see the proofs that Assumptions S-BM.2 and S-B.4 are satisfied for $\bar{\delta}$ that can be set arbitrarily close, but not equal to zero, check Appendix II. A byproduct of these proofs is that Assumption S-B.3 is also satisfied for any $\beta \in (0, 1)$.

---

[7]This is clearly not the only construction of $\hat{h}_n$ that satisfies the relevant assumptions.

[8]In this and the following examples, $I_\beta(\mathfrak{h})$ is defined at infinite values of $\mathfrak{h}_1$ for theoretical completeness only. For the estimator $\hat{h}_n$ we consider, it is in fact not possible to have $|\hat{h}_{n,1}| = \infty$ with positive probability for any finite n.

Assumption PI follows from the the weak law of large numbers for L^{1+d}-bounded independent random variables. Now, the fact that under H_0 and when the drifting sequence $\{\gamma_{n,h}\}$ with $|h_1| < \infty$ characterizes the true DGP,

$$
\begin{pmatrix} \widetilde{T}_{n,1}(\theta_0) \\ \hat{h}_{n,1} \\ T_{n,2} \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} -h_1 h_2 (1 - h_2^2)^{-1/2} \\ h_1 \\ h_1 \end{pmatrix}, \begin{pmatrix} 1 & -h_2(1 - h_2^2)^{1/2} & 0 \\ -h_2(1 - h_2^2)^{1/2} & 1 - h_2^2 & 1 - h_2^2 \\ 0 & 1 - h_2^2 & 1 \end{pmatrix} \right),
$$

$$
\begin{pmatrix} \hat{T}_{n,1}(\theta_0) \\ \hat{h}_{n,1} \\ T_{n,2} \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ h_1 \\ h_1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & h_2 \\ 0 & 1 - h_2^2 & 1 - h_2^2 \\ h_2 & 1 - h_2^2 & 1 \end{pmatrix} \right) \tag{7}
$$

(see Appendix II) makes it straightforward to calculate functions $\bar{a}_i(\cdot)$ and $\eta(\cdot)$ satisfying Assumptions S-BA-PI(i)+(iii) and S-BM-PI via computer simulation.

For example, working with the parameter space with $H_2 = [-0.99, 0.99]$ (limiting the absolute maximum asymptotic correlation between OLS estimators to 0.99) and setting $\alpha = 0.05$, $\underline{\delta}^m = \bar{\delta}^m = 0.005$, $c = 1.96$ (standard pretesting), $\xi(\beta) = \beta/2$, $r = 1$ and $\beta_1 = 0$, $\bar{a}_1(\cdot) = \alpha$, which clearly satisfies Assumption S-BA-PI(i) for $\bar{\delta}_1 = 0$. Furthermore,

$$
P(W_h \geq \sup_{\mathfrak{h} \in I_0(\tilde{h})} c_{\mathfrak{h}}(1 - \bar{a}_1(h_2))) = P(W_h \geq \sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, h_2)}(1 - \alpha)) \leq \alpha
$$

for all $h \in H$ and there is some $h^* \in H$ such that $P(W_{h^*} \geq \sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, h_2)}(1 - \alpha)) = \alpha$, by the continuity properties of W_h and $c_h(1-\alpha)$ discussed in Section 2.1.1, so that Assumption S-BA-PI(iii) is satisfied for $\beta_1 = 0$ and $\bar{a}_1(\cdot) = \alpha$. Finally, a function $\eta(\cdot)$ satisfying Assumption S-BM-PI for this choice of anchor is simply given by $\eta(\cdot) = 0$. Clearly $\eta(\cdot)$ is continuous over H_2 so that Assumption S-BM-PI(i) is satisfied. Examination of the random variables W_h and $\tilde{h}$ and their distribution functions reveals that $W_h - c_{min-B-PI}(\alpha, \tilde{h})$ is an absolutely continuous random variable so that Assumption S-BM-PI(ii) is satisfied. The function $\eta(\cdot)$ was constructed so that Assumption B3-PI(iii) holds and h^* can be set to $(7.3, 0)$, for example.

Alternatively, we can use the anchor to direct power toward a point in H. Using the same values as above, but finding β_1 to minimize the distance between $c_{0,0.6}(0.95)$ and $c^S_{B-A-PI}(0.05, \beta, (0, 0.6))$ (in order to direct power toward $h = (0, 0.6)$), we find $\beta_1 = 0.89$ with corresponding

$$
\begin{aligned}
\bar{a}_1(x) = \ &0.035 \cdot \mathbf{1}(0.99 \geq |x| \geq 0.91) + (0.49 - 0.5|x|)\mathbf{1}(0.91 > |x| > 0.9) \\
&+ 0.04 \cdot \mathbf{1}(0.9 \geq |x| \geq 0.61) + (0.345 - 0.5|x|)\mathbf{1}(0.61 > |x| > 0.6) \\
&+ 0.045 \cdot \mathbf{1}(0.6 \geq |x| \geq 0.21) + (0.15 - 0.5|x|)\mathbf{1}(0.21 \geq |x| \geq 0.2)
\end{aligned}
$$

$$+ 0.05\mathbf{1}(0.2 \geq |x| \geq 0).$$

This function was constructed to satisfy Assumption S-BA-PI(i)+(iii) while remaining as large as possible. Finally, a very small η function satisfying Assumption S-BM-PI for this choice of anchor is given by

$$\eta(x) = (0.61 - |x|)\mathbf{1}(0.61 > |x| > 0.58) + 0.03 \cdot \mathbf{1}(0.58 \geq |x| \geq 0.57)$$
$$+ (|x| - 0.54)\mathbf{1}(0.57 > |x| > 0.55) + 0.01 \cdot \mathbf{1}(0.55 \geq |x| \geq 0.48)$$
$$+ (|x| - 0.47)\mathbf{1}(0.48 > |x| > 0.47).$$

Assumption S-BM-PI holds for identical reasons to those given for the case of $\beta_1 = 0$ above. Here, h^* can be set to $(2.45, 0.6)$, for example.

Note that, for either choice of anchor, $\eta(\cdot)$ is either numerically indistinguishable from zero or $\sup_{\gamma_2 \in \Gamma_2} \eta(\gamma_2) = 0.03$.[9] At the same time, $c_{min-B-PI}(\alpha, h)$ ranges from about 2 to 23.7 so that, even when it is not equal to zero, the SCF is numerically dwarfed by $c_{min-B-PI}(\alpha, h)$. In fact, upon artificially setting $\eta(\cdot) = 0$ for the $\beta_1 = 0.89$ choice of anchor, the asymptotic size of the test is 5.1%, entailing minimal size distortion.

4.1 Critical Value Graphs: An Illustration

As an illustration of how the Bonf-Min CVs operate, we graph $c_{BM-PI}^S(\alpha, \cdot)$ for the two different anchor choices corresponding to $\beta_1 = 0$ and $\beta_1 = 0.89$, discussed above. In the construction of the graphs, α, $\underline{\delta}^m$, $\bar{\delta}^m$, c, $\xi(\beta)$, $\bar{a}_1(\cdot)$ and $\eta(\cdot)$ were set to the same values as above. Figure 1 graphs four CV functions for $h_2 = 0.9$ as a function of $h_1 \in H_1$: the true localized CV function (c_h), the PI LF CV (c_{LF-PI}), the PI S-Bonf-Min CV function (c_{BM-PI}^S) using the PI LF CV as its anchor ($\beta_1 = 0$) and the PI S-Bonf-Min CV function using $c_{B-A-PI}^S(0.05, 0.89, h)$ as its anchor. We examine the two different PI S-Bonf-Min CVs to compare the differences that the choice of anchor makes in the construction of the CVs. We can see that the PI S-Bonf-Min CV functions differentiate between different regions of h_1. Though both S-Bonf-Min CVs conservatively mimic the underlying localized quantile function $c_h(1 - \alpha)$, the CV using $c_{B-A-PI}^S(0.05, 0.89, h)$ as its anchor appears to more closely track it than does that using $c_{LF-PI}(0.05, h_2)$. For all but a small portion of the parameter space H_1, the CV using $c_{B-A-PI}^S(0.05, 0.89, h)$ as its anchor also lies below that

---

[9]Strictly speaking, for either choice of anchor, $\eta(\gamma_2) > 0$ for all $\gamma_2 \neq 0$ in this problem. However, for practical purposes, for values of $\eta(\cdot)$ less than 0.005, numerical approximation error dominates any miniscule size distortion that may arise from setting $\eta(\cdot)$ exactly equal to zero.

using $c_{LF-PI}(0.05, h_2)$ so that we can expect tests using the former to tend to have higher power than tests using the latter. Both the S-Bonf-Min CVs lie above the true localized quantile function in order to account for the asymptotic uncertainty in estimating h_1. At their peaks, the S-Bonf-Min CVs are either slightly larger or no larger than the PI LF CV, entailing corresponding tests with very minimal power loss over any portion of the parameter space. When h_1 falls outside of the range of roughly $[-6, 6]$, the PI S-Bonf-Min CVs collapse to approximate the standard normal asymptotic CV while the PI LF CV is constant and very high for all values of h_1. This enables tests using the PI S-Bonf-Min CVs to obtain substantially higher power when $\hat{h}_{n,1}$ lies outside of this range. This is also the case for large portions within $[-6, 6]$. Taking these facts together, we can expect large power gains and minimal power loss from using S-Bonf-Min CVs. We can also see from this graph how the S-Bonf-Min CVs smooth between their anchors and the standard normal CVs. In comparison, a technique using a binary decision rule (in the spirit of e.g., Andrews and Soares, 2010 and Andrews and Cheng, 2012) would choose the PI LF CV for $\hat{h}_{n,1}$ within some range and the standard normal CV for $\hat{h}_{n,1}$ outside of this range, rather than adaptively using the data to transition between them.

We also examine the above CV functions for $h_2 = 0.6$ to ascertain the generality of the above statements. We again see that the S-Bonf-Min CV using $c_{B-A-PI}^{S}(0.05, 0.89, h)$ as its anchor is smaller than that using $c_{LF-PI}(0.05, h_2)$ over most of H_1 and that when h_1 lies outside of a given range, the S-Bonf-Min CVs are substantially smaller than their LF counterparts. We also again see the smoothing described above. The general features of these two graphs are shared for all values of h_2 although the differences between the PI LF and S-Bonf-Min CVs shrink as h_2 approaches zero (compare the vertical axes in Figures 1 and 2).

4.2 Finite Sample Properties

We now analyze how the S-Bonf-Min CVs examined throughout this section behave in a realistic sample size. To this end, let us consider (an approximation to) the exact size and power of the two-sided post-conservative model selection t-test using these PI S-Bonf-Min CVs at a sample size of $n = 120$. The null hypothesis for the size analysis and the subsequent power analysis (to follow) is $H_0 : \theta = 0$. Let us consider the simplest model for which only the two regressors x_1^* and x_2^* enter. In the formation of the CVs, we used the exact same constructions as those used in the above CV graph illustration. In order to keep the scale of the t-statistic similar across values of γ_2, we fixed the variance of the OLS estimators of θ and

β_2 to unity. Additionally, we fixed $\sigma = 1$ and used normally distributed data so that x_{1i}^* and x_{2i}^* are normally distributed with mean zero, variance $1/(1-\gamma_2^2)$ and covariance $-\gamma_2/(1-\gamma_2^2)$ and ε_i is drawn from an independent standard normal distribution. For the size analysis, we follow AG and, for a fixed value of γ_2, report the maximum NRP over the parameter space of Γ_1. We do this by examining the point-wise NRP of the test for a wide range of values for β_2 and take the maximum over β_2 for each γ_2 under study.[10] The size values correspond to a 5% nominal level and are computed from 10,000 Monte Carlo replications, reported in Table 1. We examine the exact sizes of tests using $c_{PI-BM}^S(0.05, \hat{h}_n)$ with anchor $c_{LF-PI}(0.05, \hat{\gamma}_{n,2})$, $c_{PI-BM}^S(0.05, \hat{h}_n)$ with anchor $c_{B-A-PI}^S(0.05, 0.89, \hat{h}_n)$ and those using the latter CV but ignoring the SCF function, that is using $c_{min-B-PI}(0.05, \hat{h}_n)$, in order to assess the practical importance of the SCF function.[11] Table 1 indicates that in this finite-sample scenario, t-tests using the first type of CV have excellent size properties, with a maximal NRP of 0.052 over the entire grid of parameter values considered. Tests using the CV with anchor $c_{B-A-PI}^S(0.05, 0.89, \hat{h}_n)$ exhibit a 3.6% size distortion. The practical relevance of this size distortion may be questionable as it occurs at extreme values of the correlation parameter, $\gamma_2 = \pm 0.99$. If we were to restrict the correlation parameter space to, say $H_2 = [-0.95, 0.95]$, we would not see these size distortions. Moreover, they disappear as the sample size grows: for $n = 1200$, the maximal NRP for $|\gamma_2| = 0.99$ is equal to 4.8%. It is quite interesting to note that ignoring the SCF (or artificially setting $\eta(\cdot) = 0$) in this problem does not introduce any additional size distortion.

Moving on to the property the new CVs are designed to enhance, we examine the power of the two-sided post-conservative model selection t-testing procedure using $c_{BM-PI}^S(\alpha, \hat{h}_n)$ with the two different anchors for various values of β_2, θ and γ_2 and compare it to the power corresponding to $c_{LF-PI}(\alpha, \hat{h}_n)$. Figures 3-7 are the finite sample ($n = 120$) power curves for these three testing procedures at a 5% level. Each graph is plotted against the range -5 to 5 for β_2 at various fixed values of γ_2 and θ and based on 10,000 Monte Carlo replications (examining a range of β_2 is the finite sample counterpart to examining a range of h_1). Due to the symmetry properties of the localized null limit distribution, results for $-\gamma_2$ are quite similar: the corresponding power graphs are a reflection of the graphs for γ_2 across $\beta_2 = 0$.

The power graphs display some interesting features. First, the power of the tests using $c_{BM-PI}^S(\alpha, \hat{h}_n)$ with either anchor is *very* good over most of the parameter space. The

---

[10]Specifically, we search over a grid of values for $|\beta_2|$ in the interval $[0, 10]$ using step sizes .0025, .025, and .25 on the intervals $[0, 0.8]$, $[0.8, 3]$ and $[3, 10]$, respectively. We also examine the value of $|\beta_2| = 999,999$

[11]Note that the SCF function for the CV with anchor $c_{LF-PI}(0.05, \hat{\gamma}_{n,2})$ is already equal to zero.

tests using $c_{BM-PI}^S(\alpha, \hat{h}_n)$ achieve this, while retaining asymptotic validity, in part by using standard normal CVs when β_2 is "far enough" away from zero, which is determined by $\hat{h}_{n,1}$. This is a consequence of the minimum of Bonferroni CVs component of the S-Bonf-Min CVs. Second, it is evident that the testing procedures using PI S-Bonf-Min CVs have higher power than that using the PI LF procedure in all cases and often to a very large extent. For example, Figure 7 shows a power difference of nearly 100% between the testing procedures over almost the entire parameter space for β_2. For the S-Bonf-Min CV using the $c_{B-A-PI}^S(0.05, 0.89, \hat{h}_n)$ anchor, although there is a small region of the parameter space for which $c_{LF-PI}(\alpha, \hat{h}_n) < c_{BM-PI}^S(\alpha, \hat{h}_n)$, this region, and the difference between the CVs within it, are so small that any power dominance by the use of $c_{LF-PI}(\alpha, \hat{h}_n)$ is indiscernible. Third, the tests using S-Bonf-Min CVs tend to maintain power at or above the maximum power of the tests using PI LF CVs over most of the parameter space of β_2. Fourth, the ranges in β_2 of low power are very small for tests using S-Bonf-Min CVs, in contrast to those using $c_{LF-PI}(\alpha, \hat{h}_n)$. Fifth, the "envelope" where the power of the three tests coincide tends to be a very small portion of β_2's parameter space. Sixth, the differences in power values over β_2 for a particular test is increasing in γ_2, as should be expected. This brings us to a related feature: the differences in power between the testing procedures is larger for larger values of γ_2. Finally, we can see that the S-Bonf-Min CV test using anchor $c_{B-A-PI}^S(0.05, 0.89, h)$ has the best power performance. As displayed in Figures 1 and 2, this CV is not always the smallest. However, the differences between the two S-Bonf-Min CVs at points for which that using the $c_{LF-PI}(0.05, \hat{h}_n)$ anchor is smaller, are so small that they have no perceptible effect on power.

5 Hypothesis Testing when a Nuisance Parameter may be on a Boundary

In this section we show how to construct a PI S-Bonf-Min CV for the testing problem introduced in Section 2.1.2, testing when a nuisance parameter may be on the boundary of its parameter space. Our construction of the confidence set I_β is somewhat different in the context of this problem than the previous. It is instructive to notice this difference as it exemplifies the general feature that the confidence set I_β should be tailored to the testing problem at hand and provides some guidance on how to construct this object in different testing scenarios.

First, $\hat{h}_n \equiv (n^{1/2}\bar{X}_{n,2}/\hat{\sigma}_{n,2}, \hat{\gamma}_{n,2})$, where $\hat{\gamma}_{n,2} = \hat{\rho}_n$. Second, let $\widetilde{H} = \mathbb{R}_\infty \times [-1+\omega, 1-\omega]$

(note the difference with H in this context) and define $I_\beta : \widetilde{H} \rightrightarrows H$ as follows:

$$I_\beta(\mathfrak{h}) = I_\beta(\mathfrak{h}_1, \mathfrak{h}_2) = \begin{cases} (\mathfrak{h}_1, \mathfrak{h}_2), & \text{if } \mathfrak{h}_1 = \infty \\ ([\max\{0, \mathfrak{h}_1 + z_\xi\}, \max\{0, \mathfrak{h}_1 + z_{1-\beta+\xi}\}], \mathfrak{h}_2), & \text{if } \mathfrak{h}_1 < \infty, \end{cases}$$

using the same notation as in the previous example. As mentioned in the previous example, ξ can be chosen to direct power against particular regions of the alternative hypothesis. Also, in the context of this example, it makes sense to judiciously choose ξ. For example, Andrews and Guggenberger (2010b) illustrate that when examining lower one-sided tests, the limit of the test statistic W_h is stochastically decreasing (increasing) in h_1 for $h_2 < 0$ ($h_2 \geq 0$). When W_h is stochastically decreasing (increasing), large (small) values of ξ may increase the power of the test as the search in $\mathfrak{h}$ over the CVs of the test statistic will be limited to a set corresponding to smaller values.

Assumption PI follows immediately from the consistency of $\hat\rho_n$. Assumption S-B.4 follows from the definition of $T_n(\theta_0)$, the continuous mapping theorem and the facts that $\hat\rho_n$ is consistent and

$$\begin{pmatrix} n^{1/2}\bar{X}_{n,1}/\hat\sigma_{n,1} \\ n^{1/2}\bar{X}_{n,2}/\hat\sigma_{n,2} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 0 \\ h_1 \end{pmatrix} + Z_{h_2}$$

by the central limit theorem. Fixing a significance level $\alpha \in (0,1)$, it can be shown that Assumption S-B.3 is satisfied for any $\beta \in (0,1)$ and Assumption S-BM.2 is satisfied for $\bar\delta$ that can be set arbitrarily close to zero (see Appendix II). For the sake of brevity, we omit the details on appropriate $\bar{a}_i(\cdot)$ and $\eta(\cdot)$ functions that satisfy Assumptions S-BA-PI(i)+(iii) and S-BM-PI, simply noting that their general features are very similar to those in the previous example.

6   Testing After Consistent Model Selection

In this section we show how to construct the objects used in $c^M_{BM-PI}(\alpha, \hat{h}_n)$ as well as showing how the remainder of the assumptions of Corollary M-BM-PI are satisfied in the hypothesis testing after consistent model selection example introduced in Section 2.2.1.

First, let $\hat{h}_n = (\sqrt{n}\hat\beta_2\rho_n/\sigma_{\beta_2,n}, \hat\gamma_{n,2})$, where $\hat\gamma_{n,2} = (\sigma_{\theta,n}, \rho_n)$. Second, let $\widetilde{H} = H$ and, using the same notation as in previous examples,

$$I_\beta(\mathfrak{h}) = I_\beta(\mathfrak{h}_1, \mathfrak{h}_2, \mathfrak{h}_3) = \begin{cases} (\mathfrak{h}_1, \mathfrak{h}_{2,1}, \mathfrak{h}_{2,2}), & \text{if } \mathfrak{h}_1 = \pm\infty \\ ([\mathfrak{h}_1 + |\rho_n|z_\xi, \mathfrak{h}_1 + |\rho_n|z_{1-\beta+\xi}], \mathfrak{h}_{2,1}, \mathfrak{h}_{2,2}), & \text{if } \mathfrak{h}_1 \in \mathbb{R}, \end{cases}$$

where $\mathfrak{h}_1 \in H_1$ and $\mathfrak{h}_2 = (\mathfrak{h}_{2,1}, \mathfrak{h}_{2,2}) \in H_2$. The remarks made in Section 4 concerning the choice of ξ and configuration of I_β also apply here. For the additional estimator not present in the SLLD framework, let $\hat{\zeta} = \sqrt{n}\hat{\beta}_2/\sigma_{\beta_2,n}c_n$.

Fix some $\alpha \in (0,1)$. To see that Assumption M-BM.2 is satisfied, note that

$$\hat{h}_{n,1} \overset{d}{\sim} \sqrt{n}\gamma_1 + N(0, \rho_n^2) \tag{8}$$

so we have,

$$\widetilde{h} = \begin{cases} (h_1, h_2), & \text{if } h_1 = \pm\infty \\ (h_1 + Z_\rho, h_2), & \text{if } h_1 \in \mathbb{R}, \end{cases}$$

where $Z_\rho \overset{d}{\sim} N(0, \rho_\infty^2)$. To satisfy this assumption, $\bar{\delta}^{(1)}$ and $\bar{\delta}^{(2)}$ need only be set greater than zero. The proof that Assumption M-BM.2 is satisfied for $\bar{\delta}^{(i)} \in (0,1)$ is essentially the same as the proof of Assumption S-BM.2 for the post-conservative model selection example except for the additional condition of $I_\beta(\hat{h}_n), I_\beta(\widetilde{h}) \subset \bar{H}$ wp 1 if $h \in \bar{H}$. The same is true for showing that Assumption M-B.3 holds when evaluated at any $\beta \in (0,1)$. It is clear that this additional condition is satisfied from the definition of $I_\beta(\cdot)$ and the expressions for $\hat{h}_n$ and $\widetilde{h}$ above. To see how Assumption M-B.4 is satisfied, check Appendix II. Regarding Assumption M-B.5, under H_0 and $\{\gamma_{n,h}\}$, since $c_n \to \infty$,

$$\hat{\zeta} = \frac{\sqrt{n}\beta_{2,n}}{\sigma_{\beta_2,n}c_n} + \frac{1}{c_n}Z \overset{p}{\longrightarrow} \zeta,$$

where $Z \overset{d}{\sim} N(0,1)$. For Assumption PI, simply note that $\hat{\gamma}_{n,2} = (\sigma_{\theta,n}, \rho_n) = \gamma_{n,2}$ and the remainder of the assumption holds by construction. Now, the fact that under H_0 and when the drifting sequence $\{\gamma_{n,h}\}$ with $|h_1| < \infty$ characterizes the true DGP,

$$\begin{pmatrix} \sqrt{n}(\widetilde{\theta} - \theta_0) \\ \hat{h}_{n,1} \end{pmatrix} \overset{d}{\longrightarrow} N\left(\begin{pmatrix} -h_1 h_{2,1} \\ h_1 \end{pmatrix}, \begin{pmatrix} h_{2,1}^2(1 - h_{2,2}^2) & 0 \\ 0 & h_{2,2}^2 \end{pmatrix} \right),$$

$$\begin{pmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ \hat{h}_{n,1} \end{pmatrix} \overset{d}{\longrightarrow} N\left(\begin{pmatrix} 0 \\ h_1 \end{pmatrix}, \begin{pmatrix} h_{2,1}^2 & h_{2,1}h_{2,2}^2 \\ h_{2,1}h_{2,2}^2 & h_{2,2}^2 \end{pmatrix} \right) \tag{9}$$

(see Appendix II), makes it straightforward to calculate functions $\bar{a}_i^{(1)}(\cdot)$, $\bar{a}_i^{(2)}(\cdot)$ and $\eta_i(\cdot)$ satisfying Assumption M-BA-PI(i)+(iii) and M-BM-PI.1(i)-(iii).

For example, upon setting $\underline{\delta}^{(i)}, \bar{\delta}^{(i)} > 0$ for $i = 1, 2$, $r = 1$ and $\beta_1 = 0$, $\bar{a}_1^{(1)} = \bar{a}_1^{(2)} = \alpha$ satisfy Assumption M-BA-PI(i)+(iii) in analogy with this choice of anchor in the post-

41

conservative model selection example. Furthermore,

$$c^{(i)}_{B-A}(\alpha, 0, h) = c^{(i)}_{LF-PI}(\alpha, h_2) \equiv \sup_{\mathfrak{h}_1 \in H_1} c^{(i)}_{(\mathfrak{h}_1, h_2)}(1 - \alpha)$$

is invariant to h_1 for $i = 1, 2$, satisfying condition (b) of Corollary M-BM-PI. For this choice of anchor, we can simply set $\eta_i(\cdot) = 0$ for $i = 1, 2$ without the need to calculate via simulation because the MLLDs analog of the sufficient condition in Corollary SCF (in Appendix III) holds. To see this and how Assumptions M-BM-PI.1 and M-BM-PI.2 are satisfied in this example, check Appendix II.

6.1 Finite Sample Properties

As in the case for two-sided post-conservative model selection testing, we examine the finite sample behavior of the upper one-sided post-consistent model selection test but now based upon the PI M-Bonf-Min CVs using $c^{(i)}_{LF-PI}(\alpha, \hat{h}_n)$ as the anchor in each localized limit distribution corresponding to $i = 1, 2$. We use the exact same DGP as that used in the post-conservative example to conduct the analysis. Of course since the regressors are stochastic, this violates the assumptions of Leeb and Pötscher (2005). Nevertheless, as we shall see, size is well controlled for this DGP, consistent with the claim that Leeb and Pötscher's (2005) analysis can be extended to cases of stochastic regressors. With the same DGP, the only differences with the post-conservative example are (i) instead of the studentized t-statistic, we use the non-studentized version while providing a PI estimator for the variance of $\hat{\theta}$ and (ii) we examine a model selection CV that is growing in the sample size, namely, the BIC choice of $c_n = \sqrt{\log n}$. We also use the same choices for $\xi(\beta)$ as in the finite sample analysis of the post-conservative example as well as $\underline{\delta}^{(i)}$ and $\bar{\delta}^{(i)}$ values corresponding to the values of $\underline{\delta}$ and $\bar{\delta}$ in that example. For the size analysis, we approximate the exact size using the same procedure. Table 2 reports the maximum over β_2 finite sample null rejection frequencies at a 5% nominal level. The table shows that the new test again has excellent size properties with a maximal NRP of 5.2% over the entire grid of parameters considered. In this example, construction of the PI M-Bonf-Min CV involves very little computation since $\eta_i(\cdot) = 0$ and $c^M_{BM-PI}(\alpha, \hat{h}_n)$ is a closed-form a function of $\hat{h}_n$ and standard normal CVs.

Moving on to the power properties of the test, we computed power functions for the 5% test corresponding to four representative values of $\gamma_{2,2}$, the correlation between the $\hat{\theta}$ and $\hat{\beta}_2$: 0.9, 0.6, 0.3 and 0. The power functions are graphed against β_2 in Figures 8-11 and are based on 10,000 Monte Carlo replications. Each graph shows power curves as θ moves

away from the null value of zero for three different values: $\theta = 0.1$, 0.2 and 0.3. As in the post-conservative example, the results for $-\gamma_{2,2}$ are quite similar.

The power curves in this one-sided post-consistent model selection test are similar to those in the two-sided post-conservative model selection test with one important difference: they do not seem to exhibit as complex power behavior near $\beta_2 = 0$. As in the post-conservative example, power is very good over most of the parameter space. Power losses occur only over very small portions of the parameter space and, even at their lowest points, they are well above the nominal size of the text except when θ is very close to the null value of zero. The power losses occur at points surrounding $\beta_2 = 0$ since it is the area of the parameter space where $c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n)$ is more likely to be selected by $\hat{\zeta}$. In this particular problem, $c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n)$ is a more conservative CV than $c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n)$, as the former is constructed by taking a maximum over an interval centered about $\hat{h}_{n,1}$ while the latter is strictly a PI CV that does not depend upon $\hat{h}_{n,1}$ (see Appendix II for details).

7 Conclusion

This study provides new methods of size-correction for tests when the null limit distribution of a test statistic is discontinuous in a parameter. The CVs utilized by these size-corrections entail power gains in a variety of circumstances over existing size-correction methods. They also enable one to conduct tests with correct asymptotic size in a general class of testing problems for which uniformly valid methods were previously unavailable.

8   Appendix I: Proofs of Main Results

This appendix is composed of the proofs of the main results, followed by auxiliary lemmas used in these proofs, some of which are of independent interest.

Proof of Theorem S-B: By the Theorem of the Maximum, Assumptions S-B.2 and S-B.3, $c_B^S(\alpha, \delta, \cdot)$ is continuous over $\widetilde{H}$. Now, take any $h \in H$ and suppose the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP. Then, Assumptions S-B.2 through S-B.4 imply

$$\limsup_{n\to\infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^S(\alpha, \delta, \hat{h}_n)) \le P(W_h \ge c_B^S(\alpha, \delta, \widetilde{h}))$$

$$\begin{aligned}
&= P(W_h \ge c_B^S(\alpha, \delta, \widetilde{h}) \ge c_h(1-\delta)) \\
&\quad + P(W_h \ge c_h(1-\delta) > c_B^S(\alpha, \delta, \widetilde{h})) \\
&\quad + P(c_h(1-\delta) > W_h \ge c_B^S(\alpha, \delta, \widetilde{h})) \\
&\le P(W_h \ge c_h(1-\delta)) + P(c_h(1-\delta) > c_B^S(\alpha, \delta, \widetilde{h})) \le \alpha
\end{aligned}$$

$$\text{(A.1)}$$

since $c_h(1-\delta)$ is the $(1-\delta)^{th}$ quantile of W_h and $P(h \notin I_{\alpha-\delta}(\widetilde{h})) \le \alpha - \delta$. Hence, we have a uniform upper bound on the asymptotic NRP under all $\{\gamma_{n,h}\}$ sequences. To establish full uniformity over Γ, let $\{\widetilde{\gamma}_n\}$ be a sequence in Γ such that

$$\mathrm{AsySz}(\theta_0, c_B^S(\alpha, \delta, \hat{h}_n)) \equiv \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0, \gamma}(T_n(\theta_0) > c_B^S(\alpha, \delta, \hat{h}_n))$$

$$= \limsup_{n\to\infty} P_{\theta_0, \widetilde{\gamma}_n}(T_n(\theta_0) > c_B^S(\alpha, \delta, \hat{h}_n)).$$

Such a sequence always exists. Let $\{\widetilde{\gamma}_{k_n}\}$ be a subsequence of $\{\widetilde{\gamma}_n\}$ such that

$$\limsup_{n\to\infty} P_{\theta_0, \widetilde{\gamma}_n}(T_n(\theta_0) > c_B^S(\alpha, \delta, \hat{h}_n)) = \lim_{n\to\infty} P_{\theta_0, \widetilde{\gamma}_{k_n}}(T_{k_n}(\theta_0) > c_B^S(\alpha, \delta, \hat{h}_{k_n})).$$

Such a subsequence always exists. Since H is compact, there exists a subsequence of $\{\widetilde{\gamma}_{k_n}\}$, call it $\{\widetilde{\gamma}_{(k_n)_j}\}$, for which

$$\lim_{k_n\to\infty} ((k_n)_j)^r \widetilde{\gamma}_{(k_n)_j,1} \in H_1 \text{ and } \lim_{(k_n)_j\to\infty} \widetilde{\gamma}_{(k_n)_j,2} \in H_2.$$

That is, $\{\gamma_{(k_n)_j}\} = \{\gamma_{(k_n)_j,h}\}$ for some $h \in H$. Hence,

$$\mathrm{AsySz}(\theta_0, c_B^S(\alpha, \delta, \hat{h}_n)) = \lim_{n\to\infty} P_{\theta_0, \widetilde{\gamma}_{(k_n)_j},h}(T_{(k_n)_j}(\theta_0) > c_B^S(\alpha, \delta, \hat{h}_{(k_n)_j})) \le \alpha,$$

where the inequality follows from (A.1). Finally,

$$\delta = \mathrm{AsySz}(\theta_0, \sup_{\mathfrak{h}\in H} c_\mathfrak{h}(1-\delta)) \le \mathrm{AsySz}(\theta_0, c_B^S(\alpha, \delta, \hat{h}_n)),$$

A.1

where the equality is the direct result of Theorem 2 in AG (Assumptions D, S-B.1 and S-B.2 imply Assumptions A, B, L and M(a) of their paper hold with "α" replaced by "δ") and the inequality follows from the fact that

$$c_B^S(\alpha, \delta, \hat{h}_n) \equiv \sup_{\mathfrak{h} \in I_\beta(\hat{h}_n)} c_{\mathfrak{h}}(1 - \delta) \leq \sup_{\mathfrak{h} \in H} c_{\mathfrak{h}}(1 - \delta)$$

for all $\gamma \in \Gamma$ and n. ■

Proof of Corollary S-B-PI: The upper bound follows from identical arguments to those made in the proof of Theorem S-B. For the lower bound, simply note that for each $h_2 \in H_2$, there is some $h_1^* \in H_1$ such that $\sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, h_2)}(1 - \delta) = c_{(h_1^*, h_2)}(1 - \delta)$ by virtue of the extreme value theorem and Assumption S-B.2(i). Hence, for some $h^* \in H$ with $\sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, h_2^*)}(1 - \delta) = c_{h^*}(1 - \delta)$,

$$\text{AsySz}(\theta_0, c_{B-PI}^S(\alpha, \delta, \hat{h}_n)) \geq \text{AsySz}(\theta_0, c_{LF-PI}(\delta, \hat{\gamma}_{n,2}))$$
$$\geq \liminf_{n \to \infty} P_{\theta_0, \gamma_{n,h^*}}(T_n(\theta_0) > \sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, \hat{\gamma}_{n,2})}(1 - \delta)) \geq P(W_{h^*} > c_{h^*}(1 - \delta)) = \delta,$$

where the third inequality follows from Assumptions S-B.1, PI and the continuity of $\sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, \cdot)}(1 - \delta)$, the latter of which is implied by Assumption S-B.2(i). The equality follows from Assumption S-B.2(ii). The sequence $\{\gamma_{n,h^*}\} \subset \Gamma$ exists by, cf., Lemma 7 of Andrews and Guggenberger (2010b). ■

Proof of Theorem S-BA: By the same arguments as those used in the proof of Theorem S-B, for any $h \in H$,

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_{B-A}^S(\alpha, \beta, \hat{h}_n)) \leq P(W_h \geq \sup_{\mathfrak{h} \in I_\beta(\tilde{h})} c_{\mathfrak{h}}(1 - \bar{\alpha})).$$

Thus, the same subsequencing argument as that used in the proof of Theorem S-B provides that there is some $h \in H$ such that

$$\text{AsySz}(\theta_0, c_{B-A}^S(\alpha, \beta, \hat{h}_n)) \leq P(W_h \geq \sup_{\mathfrak{h} \in I_\beta(\tilde{h})} c_{\mathfrak{h}}(1 - \bar{\alpha})) \leq \alpha,$$

where the latter inequality holds by S-BA(i) and the definition of $\bar{\alpha}$. On the other hand,

$$\text{AsySz}(\theta_0, c_{B-A}^S(\alpha, \beta, \hat{h}_n)) \geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h^*}}(T_n(\theta_0) > c_{B-A}^S(\alpha, \beta, \hat{h}_n))$$
$$\geq P(W_{h^*} \geq c_{B-A}^S(\alpha, \beta, \tilde{h}^*)) = \alpha. ■$$

Proof of Corollary S-BA-PI: By the Theorem of the Maximum, Assumptions S-BM.1, S-BA-PI(i) and S-B.3 imply that $c_{B-A-PI}^S(\alpha, \beta, \cdot)$ is continuous. Hence, by the same arguments as those used in the proof of Theorem S-B, for any $h \in H$,

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_{B-A-PI}^S(\alpha, \beta, \hat{h}_n)) \leq P(W_h \geq \sup_{\mathfrak{h} \in I_\beta(\tilde{h})} c_{\mathfrak{h}}(1 - \bar{a}(h_2))).$$

A.2

The same subsequencing argument provides that there is some $h \in H$ such that

$$\text{AsySz}(\theta_0, c_{B-A-PI}^S(\alpha, \beta, \hat{h}_n)) \leq P(W_h \geq \sup_{\mathfrak{h} \in I_\beta(\widetilde{h})} (1 - \bar{a}(h_2))) \leq \alpha$$

by Assumptions PI and S-BA-PI(ii)-(iii). On the other hand,

$$\text{AsySz}(\theta_0, c_{B-A-PI}^S(\alpha, \beta, \hat{h}_n)) \geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}^*}(T_n(\theta_0 > c_{B-A-PI}^S(\alpha, \beta, \hat{h}_n)))$$
$$\geq P(W_{h^*} \geq c_{B-A-PI}^S(\alpha, \beta, \widetilde{h}^*)) = \alpha. \ \blacksquare$$

Proof of Theorem S-BM: Lemma BM provides that $c_{min-B}(\alpha, \cdot)$ is continuous over $\widetilde{H}$ by Assumptions S-B.2(i) evaluated at $\delta_i = \bar{\alpha}_i$, S-B.3 evaluated at β_i, S-BM.1 and S-BM.2 for $i = 1, \ldots, r$. Hence, using Assumptions S-B.1 and S-B.4, for any $h \in H$,

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_{BM}^S(\alpha, \hat{h}_n)) \leq P(W_h \geq c_{min-B}(\alpha, \widetilde{h}) + \Delta).$$

Thus, the same subsequencing argument as that used in the proof of Theorem S-B provides that there is some $h \in H$ such that

$$\text{AsySz}(\theta_0, c_{BM}^S(\alpha, \hat{h}_n)) \leq P(W_h \geq c_{min-B}(\alpha, \widetilde{h}) + \Delta) \leq \alpha,$$

where the latter inequality holds by Assumption S-BM.3(i) and the definition of Δ. On the other hand, from Assumption S-BM.3(ii),

$$\text{AsySz}(\theta_0, c_{BM}^S(\alpha, \hat{h}_n)) \geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h^*}}(T_n(\theta_0) > c_{BM}^S(\alpha, \hat{h}_n)) \geq P(W_{h^*} \geq c_{BM}^S(\alpha, \widetilde{h}^*)) = \alpha.$$

$\blacksquare$

Proof of Corollary S-BM-PI: Lemma BM provides that $c_{min-B-PI}(\alpha, \cdot)$ is continuous over $\widetilde{H}$ by Assumptions S-B.3 evaluated at β_i and S-BM.1 and S-BM.2 at pairs $(\underline{\delta}^m, \bar{\delta}^m)$ and $(\underline{\delta}_i, \bar{\delta}_i)$ for $i = 1, \ldots, r$. Hence, using Assumptions S-B.1, S-B.4, PI and S-BM-PI(i),

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_{BM-PI}^S(\alpha, \hat{h}_n)) \leq P(W_h \geq c_{min-B-PI}(\alpha, \widetilde{h}) + \eta(h_2))$$

for any $h \in H$. Thus, the same subsequencing argument used in the proof of Theorem S-B provides that there is some $h \in H$ such that

$$\text{AsySz}(\theta_0, c_{BM-PI}^S(\alpha, \hat{h}_n)) \leq P(W_h \geq c_{min-B-PI}(\alpha, \widetilde{h}) + \eta(h_2)) \leq \alpha,$$

which follows from Assumption S-BM-PI(ii)-(iii). On the other hand, using Assumption S-BM-PI(iii), the same argument used in the proof of Theorem S-BM provides

$$\text{AsySz}(\theta_0, c_{BM-PI}^S(\alpha, \hat{h}_n)) \geq \alpha. \ \blacksquare$$

A.3

Proof of Theorem M-BM: Many parts of the proof are quite similar to the corresponding parts in the proof of Lemma M-B, so details are omitted. We now study $\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{BM}^M(\alpha, \hat{h}_n))$ for the same three cases corresponding to Assumption M-B.1.

Case 1: $\zeta(\{\gamma_{n,h}\}) \in K$. Using very similar arguments to those used in the proof of Lemma M-B, we can show that for any $h \in H$,

$$\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{BM}^M(\alpha, \hat{h}_n)) \leq \limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) \geq c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1), \quad \text{(A.2)}$$

here using Assumptions M-B.2(i) evaluated at $\delta_i^{(1)} = \bar{\alpha}_i^{(1)}$ corresponding to β_i and $c_h^{(1)}$ and $\delta_i^{(2)} = \bar{\alpha}_i^{(2)}$ corresponding to β_i and $c_h^{(2)}$, M-B.3 evaluated at β_i, M-BM.1, M-BM.2(i) and M-BM.4(i) for $i = 1, \ldots, r$ to establish the finiteness of $c_{min-B}^{(2)}(\alpha, \hat{h}_n) + \Delta_2$ for any finite n and $h \in H$. Now, for $h \in \bar{H}^c$, Assumption M-BM.4(ii) immediately provides that (A.2) is bounded above by α. For $h \in \bar{H}$, Assumptions M-B.2(i) evaluated at $\delta_i^{(1)} = \bar{\alpha}_i^{(1)}$ corresponding to β_i and $c_h^{(1)}$ and $\delta_i^{(2)} = \bar{\alpha}_i^{(2)}$ corresponding to β_i and $c_h^{(2)}$, M-B.3 evaluated at β_i, for $i = 1, \ldots, r$, M-BM.1 and M-BM.2, in conjunction with Lemma BM, imply that $c_{min-B}^{(1)}(\alpha, \cdot) + \Delta_1$ is continuous so that by Assumption M-B.4(i), (A.2) is bounded above by

$$P(W_h^{(1)} \geq c_{min-B}^{(1)}(\alpha, \widetilde{h}) + \Delta_1) \leq \alpha,$$

where the inequality results from Assumption M-BM.3(i) and the definition of Δ_1.

Case 2: $\zeta(\{\gamma_{n,h}\}) \in L$. The proof that $\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{BM}^M(\alpha, \hat{h}_n)) \leq \alpha$ for all $h \in H$ that fall into this category is again symmetric to the proof for Case 1.

Case 3: $\zeta(\{\gamma_{n,h}\}) \in K^c \cap L^c$. Assumptions M-B.2(i) evaluated at $\delta_i^{(1)} = \bar{\alpha}_i^{(1)}$ corresponding to β_i and $c_h^{(1)}$ and $\delta_i^{(2)} = \bar{\alpha}_i^{(2)}$ corresponding to β_i and $c_h^{(2)}$, M-B.3 evaluated at β_i, for $i = 1, \ldots, r$, M-BM.1, M-BM.2(i) and M-BM.4(i) and similar arguments to those in the proof of Lemma M-B provide that for any $h \in H$,

$$\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{BM}^M(\alpha, \hat{h}_n))$$
$$\leq \limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) \geq \max\{c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1, c_{min-B}^{(2)}(\alpha, \hat{h}_n) + \Delta_2\}), \quad \text{(A.3)}$$

in this case. Assumptions M-B.1(iii) and M-BM.4(ii)-(iii) and a similar change of measure to that used in the proof of Theorem B-M provide that (A.3) is bounded above by α for $h \in \bar{H}^c$. For $h \in \bar{H}$, if $W_h^{(3)}$ is stochastically dominated by $W_h^{(1)}$, (A.3) is bounded above by

$$\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(W_h^{(1)} \geq c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1) \leq P(W_h^{(1)} \geq c_{min-B}^{(1)}(\alpha, \widetilde{h}) + \Delta_1) \leq \alpha,$$

where we have again use the continuity of $c_{min-B}^{(1)}(\alpha, \cdot) + \Delta_1$ and the final inequality was established above. If $W_h^{(3)}$ is stochastically dominated by $W_h^{(2)}$, the argument is symmetric.

We can now establish that $\text{AsySz}(\theta_0, c_{BM}^M(\alpha, \hat{h}_n)) \leq \alpha$ using the same type of subsequencing argument as that used in the proofs of all the main results.

A.4

Finally, if (a) of Assumption M-BM.3(ii) holds, similar arguments to those used to establish (A.4) yield

$$\text{AsySz}(\theta_0, c_{BM}^M(\alpha, \hat{h}_n)) \geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h^*(1)}}(T_n(\theta_0) > c_{BM}^M(\alpha, \hat{h}_n))$$

$$\geq \liminf_{n \to \infty} P_{\theta_0, \gamma_{n,h^*(1)}}(T_n(\theta_0) > c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1) = \alpha.$$

If (b) of Assumption M-BM.3(ii) holds, the argument is symmetric. ∎

Proof of Corollary M-BM-PI: The proof is very similar to the proof of Theorem M-BM so most details are omitted. Assumptions M-B.3 evaluated at β_i for $i = 1, \ldots, r$, M-BM.1, M-BM.2(i), M-BM-PI.1(i) and M-BM-PI.2(i) can be used to establish the finiteness of $c_{min-B-PI}^{(i)}(\alpha, \hat{h}_n) + \eta_i(\hat{\gamma}_{n,2})$ for $i = 1, 2$. In this case, Assumption M-BM-PI.1(iii) is used directly to establish the upper bound on limiting null rejection probabilities when $h \in \bar{H}$. For the lower bound, Assumption M-BM-PI.1(iv) plays the role of M-BM.3(ii) in Theorem M-BM. ∎

The following is an auxiliary Lemma used in many of the proofs.

Lemma BM. *Let $\mathfrak{H}$, $\mathfrak{D}$, $\widetilde{\mathfrak{H}}$ and $\mathfrak{B}$ be metric spaces, $f : \mathfrak{H} \times \mathfrak{D} \to \mathbb{R}$ be a function that is continuous in both of its arguments and $G : \widetilde{\mathfrak{H}} \times \mathfrak{B} \rightrightarrows \mathfrak{H}$ be a compact-valued correspondence that is continuous in both of its arguments. If the set $\{(d, b) \in \mathfrak{D} \times \mathfrak{B} : d + b = a\}$ is nonempty and compact, then*

$$\inf_{\{(d,b) \in \mathfrak{D} \times \mathfrak{B}: d+b=a\}} \sup_{x \in G(y,b)} f(x, d)$$

is continuous in $y \in \widetilde{\mathfrak{H}}$.

Proof : The maximum theorem provides that $\sup_{x \in G(y,b)} f(x, d)$ is continuous in $(y, b, d) \in \widetilde{\mathfrak{H}} \times \mathfrak{B} \times \mathfrak{D}$. A second application of the maximum theorem yields the lemma's claim since, as a correspondence from $\mathfrak{H}$ into $\mathfrak{D} \times \mathfrak{B}$, $\{(d, b) \in \mathfrak{D} \times \mathfrak{B} : d + b = a\}$ is compact-valued and trivially continuous. ∎

The following lemma establishes the correct asymptotic size of tests using M-Bonf CVs. Assumptions M-B.6 and M-B-LB can be found in Appendix III.

Lemma M-B. *Under Assumptions D and M-B.1 through M-B.6 for $\beta = \alpha - \delta$,*

$$\text{AsySz}(\theta_0, c_B^M(\alpha, \delta, \hat{h}_n)) \leq \alpha.$$

If Assumption M-B-LB also holds,

$$P(W_{h^*(i)}^{(i)} > c_{h^*(i)}^{(i)}(1 - \delta) + \varepsilon_i) \leq \text{AsySz}(\theta_0, c_B^M(\alpha, \delta, \hat{h}_n)),$$

where $i = 1$ if (a) holds and $i = 2$ if (b) holds.

A.5

Proof: Take any $h \in H$. We break down the proof that $\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^M(\alpha, \delta, \hat{h}_n)) \leq \alpha$ for any $h \in H$ into the three cases corresponding to Assumption M-B.1.

Case 1: $\zeta(\{\gamma_{n,h}\}) \in K$. Assumptions M-B.2(i), M-B.3 and M-B.6(i) provide that for any $h \in H$ and finite n,

$$P_{\theta_0, \gamma_{n,h}}(|\sup_{\mathfrak{h} \in I_{\alpha - \delta}(\hat{h}_n)} c_{\mathfrak{h}}^{(2)}(1 - \delta)| < \infty) = 1.$$

Hence, for any $h \in H$ and $\epsilon > 0$,

$$P_{\theta_0, \gamma_{n,h}}(|\mathbf{1}(\hat{\zeta} \in \bar{L})c_B^{(2)}(\alpha, \delta, \hat{h}_n)| \geq \epsilon) \leq P_{\theta_0, \gamma_{n,h}}(\hat{\zeta} \in \bar{L}) \to 0$$

by Assumption M-B.5 and the fact that $\bar{L} \subset \text{int}(K^c)$. Hence, using the fact that $\bar{L}^c = \bar{K} \cup (\bar{K}^c \cap \bar{L}^c)$,

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^M(\alpha, \delta, \hat{h}_n)) = \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > \mathbf{1}(\hat{\zeta} \in \bar{K})c_B^{(1)}(\alpha, \delta, \hat{h}_n)$$

$$+ \mathbf{1}(\hat{\zeta} \in \bar{L})c_B^{(2)}(\alpha, \delta, \hat{h}_n) + \mathbf{1}(\hat{\zeta} \in \bar{K}^c \cap \bar{L}^c) \max\{c_B^{(1)}(\alpha, \delta, \hat{h}_n), c_B^{(2)}(\alpha, \delta, \hat{h}_n)\})$$

$$\leq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > \mathbf{1}(\hat{\zeta} \in \bar{L}^c)c_B^{(1)}(\alpha, \delta, \hat{h}_n) + \mathbf{1}(\hat{\zeta} \in \bar{L})c_B^{(2)}(\alpha, \delta, \hat{h}_n))$$

$$= \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) \geq c_B^{(1)}(\alpha, \delta, \hat{h}_n)), \tag{A.4}$$

where the final equality follows from Assumption M-B.5 and the fact that $\mathbf{1}(x \in \bar{L}^c)$ is continuous at any $x \in K$. If $h \in \bar{H}^c$, Assumption M-B.6(ii) immediately implies that (A.4) is bounded above by α. Suppose then that $h \in \bar{H}$. Given Assumptions M-B.2(i) and M-B.3, the Maximum Theorem implies that $c_B^{(1)}(\alpha, \delta, \cdot)$ is continuous when $I_\beta(\cdot) \subset \bar{H}$. Hence, Assumptions M-B.1(i), M-B.3 and M-B.4(i) imply that (A.4) is bounded above by

$$P(W_h^{(1)} \geq c_B^{(1)}(\alpha, \delta, \widetilde{h})) \leq P(W_h^{(1)} \geq c_h^{(1)}(1 - \delta) + \varepsilon_1) + P(h \notin I_{\alpha - \delta}(\widetilde{h})) \leq \alpha, \tag{A.5}$$

where the first inequality is the result of the same type of Bonferroni arguments made in the proof of Theorem S-B and the second inequality follows from Assumptions M-B.2(ii) and M-B.3.

Case 2: $\zeta(\{\gamma_{n,h}\}) \in K$. The proof that $\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^M(\alpha, \delta, \hat{h}_n)) \leq \alpha$ for all $h \in H$ in this case is symmetric to the proof for Case 1.

Case 3: $\zeta(\{\gamma_{n,h}\}) \in K^c \cap L^c$. Assumptions M-B.2(i), M-B.3 and M-B.6(i) provide that for $i = 1, 2$, any $h \in H$ and finite n,

$$P_{\theta_0, \gamma_{n,h}}(|\sup_{\mathfrak{h} \in I_{\alpha - \delta}(\hat{h}_n)} c_{\mathfrak{h}}^{(i)}(1 - \delta)| < \infty) = 1$$

so that for any $\epsilon > 0$, we have

$$P_{\theta_0, \gamma_{n,h}}(|\mathbf{1}(\hat{\zeta} \in \bar{K})c_B^{(1)}(\alpha, \delta, \hat{h}_n)| \geq \epsilon), P_{\theta_0, \gamma_{n,h}}(|\mathbf{1}(\hat{\zeta} \in \bar{L})c_B^{(2)}(\alpha, \delta, \hat{h}_n)| \geq \epsilon) \to 0$$

by Assumption M-B.5. Hence, similar expressions to those leading to (A.4) yield

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^M(\alpha, \delta, \hat{h}_n))$$

A.6

$$\leq \limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) \geq \max\{c_B^{(1)}(\alpha,\delta,\hat{h}_n), c_B^{(2)}(\alpha,\delta,\hat{h}_n)\}). \tag{A.6}$$

Suppose $h \in \bar{H}^c$. If $W_h^{(3)}$ is stochastically dominated by $W_h^{(1)}$, (A.6) is bounded above by

$$\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_B^{(1)}(\alpha,\delta,\hat{h}_n)) \leq \limsup_{n\to\infty} P_{\theta_0,\tilde{\gamma}_{n,h}^{(1)}}(T_n(\theta_0) > c_B^{(1)}(\alpha,\delta,\hat{h}_n)) \leq \alpha,$$

for some $\{\tilde{\gamma}_{n,h}^{(1)}\} \subset \Gamma$ with $\zeta(\{\tilde{\gamma}_{n,h}^{(1)}\}) \in K$ satisfying Assumption M-B.6(iii), where the first inequality follows from Assumption M-B.6(iii) and the second from Assumption M-B.6(ii). If $W_h^{(3)}$ is stochastically dominated by $W_h^{(2)}$, the argument is symmetric. Now suppose $h \in \bar{H}$. If $W_h^{(3)}$ is stochastically dominated by $W_h^{(1)}$, (A.6) is bounded above by

$$\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(W_h^{(1)} \geq c_B^{(1)}(\alpha,\delta,\hat{h}_n)) \leq P(W_h^{(1)} \geq c_B^{(1)}(\alpha,\delta,\tilde{h})) \leq \alpha,$$

where the first inequality results from the continuity of $c_B^{(1)}(\alpha,\delta,\cdot)$ when $I_\beta \subset \bar{H}$ and Assumption B.3 and the second inequality has been established in (A.5). If $W_h^{(3)}$ is stochastically dominated by $W_h^{(2)}$, the argument is symmetric.

Now, since we have established that $\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_B^M(\alpha,\delta,\hat{h}_n)) \leq \alpha$ for any $h \in H$, $\text{AsySz}(\theta_0, c_B^M(\alpha,\delta,\hat{h}_n)) \leq \alpha$ follows from the same type of subsequencing argument as that used in the proof of Theorem S-B.

Finally, if (a) of Assumption M-B-LB holds,

$$\begin{aligned}
\text{AsySz}(\theta_0, c_B^M(\alpha,\delta,\hat{h}_n)) &\geq \limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h^*(1)}}(T_n(\theta_0) > c_B^M(\alpha,\delta,\hat{h}_n)) \\
&\geq \liminf_{n\to\infty} P_{\theta_0,\gamma_{n,h^*(1)}}(T_n(\theta_0) > c_B^{(1)}(\alpha,\delta,\hat{h}_n)) \\
&\geq \liminf_{n\to\infty} P_{\theta_0,\gamma_{n,h^*(1)}}(T_n(\theta_0) > \sup_{\mathfrak{h}\in H} c_{\mathfrak{h}}^{(1)}(1-\delta) + \varepsilon_1) \\
&\geq P(W_{h^*(1)}^{(1)} > c_{h^*(1)}^{(1)}(1-\delta) + \varepsilon_1),
\end{aligned}$$

where the second inequality follows from very similar arguments to those used to establish (A.4) and the final inequality follows from Assumption M-B.1(i). If (b) of Assumption M-B-LB holds, the argument is symmetric. ∎

Clearly, if $\varepsilon_i = 0$ and $J_{h^*(i)}^{(i)}(\cdot)$ is continuous at $c_{h^*(i)}(1-\delta)$, then the lower bound is equal to δ. We now remark on an important, easy to verify, sufficient condition that implies some of the imposed assumptions hold: as a function from H in $\mathbb{R}$, $c_h^{(i)}(1-\delta)$ is continuous for $i = 1$ or 2. Clearly this a strengthening of Assumption M-B.2(i). In conjunction with the other assumptions of Lemma M-B, it also implies that one of the h^*'s in Assumption M-B-LB exists (but it does not imply the existence of corresponding γ sequences) and M-B.6(i)-(ii) holds. To see why Assumptions M-B.6(ii) holds in this case, notice that application of the Maximum Theorem provides that $c_B^{(i)}(\alpha,\delta,\cdot)$ is continuous and Bonferroni arguments yield $P(W_h^{(i)} \geq c_B^{(i)}(\alpha,\delta,\tilde{h})) \leq \alpha$. See the arguments surrounding (A.5) in the proof of Lemma M-B for details. Finally, if this condition holds, the statements made in Assumption M-B.6(iii)

A.7

with regard to $\{\widetilde{\gamma}_{n,h}^{(j)}\}$ may be dropped since then the same arguments made in Case 3 of the proof for $h \in \bar{H}$ may be generalized to all $h \in H$.

The following lemma establishes the correct asymptotic size of tests using PI M-Bonf CVs. Assumption M-B-PI-LB can be found in Appendix III.

Lemma M-B-PI. *Under Assumptions D, PI and M-B.1 through M-B.6 for $\beta = \alpha - \delta$,*

$$\text{AsySz}(\theta_0, c_{B-PI}^M(\alpha, \delta, \hat{h}_n)) \leq \alpha.$$

If Assumption M-B-PI-LB also holds,

$$P(W_{(h_1^{*(i)}, h_2^*)}^{(i)} > c_{(h_1^{*(i)}, h_2^*)}^{(i)} (1 - \delta) + \varepsilon_i) \leq \text{AsySz}(\theta_0, c_{B-PI}^M(\alpha, \delta, \hat{h}_n)),$$

where $i = 1$ if (a) holds and $i = 2$ if (b) holds.

Proof: The upper bound follows from identical arguments to those made in Lemma M-B. For the lower bound, if (a) of Assumption M-B-PI-LB holds,

$$\text{AsySz}(\theta_0, c_{B-PI}^M(\alpha, \delta, \hat{h}_n)) \geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,(h_1^{*(1)}, h_2^*)}}(T_n(\theta_0) > c_{B-PI}^M(\alpha, \delta, \hat{h}_n))$$

$$\geq \liminf_{n \to \infty} P_{\theta_0, \gamma_{n,(h_1^{*(1)}, h_2^*)}}(T_n(\theta_0) > c_{B1-PI}^{(1)}(\alpha, \delta, \hat{h}_n))$$

$$\geq \liminf_{n \to \infty} P_{\theta_0, \gamma_{n,(h_1^{*(1)}, h_2^*)}}(T_n(\theta_0) > \sup_{\mathfrak{h}_1 \in H_1} c_{(\mathfrak{h}_1, \hat{\gamma}_{n,2})}^{(1)}(1 - \delta) + \varepsilon_1)$$

$$\geq P(W_{(h_1^{*(1)}, h_2^*)}^{(1)} > c_{(h_1^{*(1)}, h_2^*)}^{(1)}(1 - \delta) + \varepsilon_1),$$

where the second inequality follows from very similar arguments to those used to establish (A.4) and the final inequality follows from Assumptions M-B.1(i), PI and M-B-PI-LB. If (b) of Assumption M-B-PI-LB holds, the argument is symmetric. ∎

9 Appendix II: Derivation of Results for Examples

9.1 Testing After Conservative Model Selection

The following is a verification that Assumption S-BM.2 is satisfied in the context of this example for $\bar{\delta}$ that can be set arbitrarily close to zero. Assume that H_0 holds and the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP.

(i) First, the weak law of large numbers for L^{1+d}-bounded independent random variables provides that $\hat{\gamma}_{n,2} \xrightarrow{p} h_2$. Results used to show (S11.10) in Andrews and Guggenberger (2009c) provide that for $|h_1| < \infty$, $\hat{h}_{n,1} \xrightarrow{d} h_1 + \sqrt{1 - h_2^2} Z$, where $Z \overset{d}{\sim} N(0,1)$. Furthermore, when $|h_1| = \infty$, $\hat{h}_{n,1} \xrightarrow{p} h_1$ since $\hat{h}_{n,1} = n^{1/2} \widetilde{\beta}_2 / (\sigma Q^{22} + o_p(1))$ and

$$n^{1/2} \widetilde{\beta}_2 = \sigma(Q^{22})^{1/2} n^{1/2} \gamma_{n,h,1} + \sigma n^{1/2} (X_2' X_2)^{-1} X_2' \varepsilon = \sigma(Q^{22})^{1/2} n^{1/2} \gamma_{n,h,1} + O_p(1).$$

Hence,

$$\widetilde{h} = \begin{cases} (h_1, h_2), & \text{if } h_1 = \pm\infty \\ (h_1 + \sqrt{1 - h_2^2} Z, h_2), & \text{if } h_1 \in \mathbb{R}. \end{cases}$$

A.8

Second, take any $\beta \in (0,1)$. Using the definition of I_β, if $h_1 = \pm\infty$, then

$$P_{\theta_0, \gamma_{n,h}}(h \in I_\beta(\widetilde{h})) = P(h = h) = 1.$$

If $h_1 \in \mathbb{R}$,

$$\begin{aligned}
&P_{\theta_0, \gamma_{n,h}}(h \in I_\beta(\widetilde{h})) \\
&= P((h_1, h_2) \in ([h_1 + \sqrt{1 - h_2^2}Z + \sqrt{1 - h_2^2}z_\xi, h_1 + \sqrt{1 - h_2^2}Z + \sqrt{1 - h_2^2}z_{1-\beta+\xi}], h_2)) \\
&= P(Z \in [z_\xi, z_{1-\beta+\xi}]) = 1 - \beta.
\end{aligned}$$

Third, as defined, I_β is clearly compact-valued for any $\beta \in (0,1)$. Finally, to show continuity of I_β over $\widetilde{H}$, first note that for $\beta > 0$, $I_\beta(\mathfrak{h}) = ([\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_\xi, \mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_{1-\beta+\xi}], \mathfrak{h}_2)$. Now, take any $\mathfrak{h} \in \widetilde{H} = \mathbb{R}_\infty \times [-1+\omega, 1-\omega]$ and let $\{\mathfrak{h}_k\}$ be a sequence in $\widetilde{H}$ such that $\mathfrak{h}_k \to \mathfrak{h}$. For each k, take any $b_k \in I_\beta(\mathfrak{h}_k)$ such that $b_k \to b$. Then, $b_k = (a_k \circledast (\mathfrak{h}_{1,k} + \sqrt{1 - \mathfrak{h}_{2,k}^2}z_\xi) + (1 - a_k) \circledast (\mathfrak{h}_{1,k} + \sqrt{1 - \mathfrak{h}_{2,k}^2}z_{1-\beta+\xi}), \mathfrak{h}_{2,k})$, for some $a_k \in [0,1]$, where $\circledast$ is a binary operator defined for any pair $(\mathfrak{a}, \mathfrak{b}) \in [0,1] \times \mathbb{R}_\infty$ as follows:

$$\mathfrak{a} \circledast \mathfrak{b} = \begin{cases} \mathfrak{a}\mathfrak{b}, & \text{if } \mathfrak{a} \neq 0 \\ 0, & \text{if } \mathfrak{a} = 0. \end{cases}$$

Now note that for $(a, b, c) \in [0,1] \times \mathbb{R}_\infty \times [-1+\omega, 1-\omega]$,

$$\begin{aligned}
\mathfrak{f}(a, b, c) &\equiv a \circledast (b + \sqrt{1 - c^2}z_\xi) + (1 - a) \circledast (b + \sqrt{1 - c^2}z_{1-\beta+\xi}) \\
&= \begin{cases} a(b + \sqrt{1 - c^2}z_\xi) + (1 - a)(b + \sqrt{1 - c^2}z_{1-\beta+\xi}), & \text{if } a \in (0,1) \\ b + \sqrt{1 - c^2}z_{1-\beta+\xi}, & \text{if } a = 0 \\ b + \sqrt{1 - c^2}z_\xi, & \text{if } a = 1 \end{cases}
\end{aligned}$$

is continuous in all three arguments. Then, since both b_k and $\mathfrak{h}_k$ converge and $[0,1]$ is compact, $a_k \to a \in [0,1]$. Hence, $b = (a \circledast (\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_\xi) + (1-a) \circledast (\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_{1-\beta+\xi}), \mathfrak{h}_2) \in I_\beta(\mathfrak{h})$ so that I_β is upper hemicontinuous. Now, take any $\mathfrak{h} \in \widetilde{H}$ and let $\{\mathfrak{h}_m\}$ be a sequence in $\widetilde{H}$ such that $\mathfrak{h}_m \to \mathfrak{h}$. For each $b \in I_\beta(\mathfrak{h})$, $b = (a \circledast (\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_{1+\xi}) + (1 - a) \circledast (\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_{1-\beta+\xi}), \mathfrak{h}_2)$ for some $a \in [0,1]$. Then, for $b_m = (a \circledast (\mathfrak{h}_{1,m} + \sqrt{1 - \mathfrak{h}_{2,m}^2}z_\xi) + (1 - a) \circledast (\mathfrak{h}_{1,m} + \sqrt{1 - \mathfrak{h}_{2,m}^2}z_{1-\beta+\xi}), \mathfrak{h}_{2,m}) \in I_\beta(\mathfrak{h}_m)$, $b_m \to b$ so that I_β is also lower hemicontinuous and therefore continuous for any $\beta \in (0,1)$.

(ii) First note that $\mathfrak{h}_2$ is trivially continuous in β. Similarly for $\beta > 0$, when $|\mathfrak{h}_1| = \infty$, $[\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_\xi, \mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_{1-\beta+\xi}] = [\mathfrak{h}_1, \mathfrak{h}_1]$ is trivially continuous in β. Very similar arguments to those used in (i) above show that $[\mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_1, \mathfrak{h}_1 + \sqrt{1 - \mathfrak{h}_2^2}z_2]$ is continuous in $(z_1, z_2) \in \{(\bar{z}_1, \bar{z}_2) \in \mathbb{R}^2 : \bar{z}_1 \leq \bar{z}_2\}$ when $|\mathfrak{h}_1| < \infty$. Since $z_\xi = z_{\xi(\beta)}$ and $z_{1-\beta+\xi} = z_{1-\beta+\xi(\beta)}$ are continuous in $\beta \in (0,1)$, we have the continuity of $I_\beta(\mathfrak{h})$ in β over $(0,1)$ for all $\mathfrak{h} \in \widetilde{H}$. ∎

A.9

To see that Assumption S-B.4 is satisfied, assume that H_0 holds and the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP. Assume $|h_1| < \infty$. Using Lemma S3 of Andrews and Guggenberger (2009c),

$$\hat{h}_{n,1} = \frac{n^{1/2}\beta_2/\sigma_n + (n^{-1}X_2'X_2)^{-1}n^{-1/2}X_2'\varepsilon}{(\hat{\sigma}/\sigma_n)(n^{-1}X_2'M_{X_1}X_2)^{-1/2}} = h_1 + (Q_n^{22})^{-1/2}(Q_{n,22})^{-1}e_2'n^{-1/2}X^{\perp\prime}\varepsilon + o_p(1)$$

where $X^\perp = [x_1^\perp, \ldots, x_n^\perp]'$ and the second equality follows analogously to (S11.18) in Andrews and Guggenberger (2009c). Combining this with expressions (S11.18) and (S11.25) of Andrews and Guggenberger (2009c),

$$\begin{pmatrix} \widetilde{T}_{n,1}(\theta_0) \\ \hat{h}_{n,1} \\ T_{n,2} \end{pmatrix} = \begin{pmatrix} -h_1h_2(1-h_2)^{-1/2} + Q_{n,11}^{-1/2}n^{-1/2}e_1'X^{\perp\prime}\varepsilon \\ h_1 + (Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'n^{-1/2}X^{\perp\prime}\varepsilon \\ h_1 + (Q_n^{22})^{1/2}(e_2 - Q_{n,12}Q_{n,11}^{-1}e_1)'n^{-1/2}X^{\perp\prime}\varepsilon \end{pmatrix} + o_p(1) \xrightarrow{d} \begin{pmatrix} \widetilde{Z}_{h,1} \\ \widetilde{h}_1 \\ Z_{h,2} \end{pmatrix},$$

by the Lindberg Central Limit Theorem and the Cramér-Wold device, where the limiting random vector is a multivariate normal. Similarly, $(\hat{T}_{n,1}(\theta_0), \hat{h}_{n,1}, T_{n,2})' \xrightarrow{d} (\hat{Z}_{h,1}, \widetilde{h}_1, Z_{h,2})'$, a multivariate normal random vector. If $|h_1| = \infty$, $\hat{h}_{n,1} \xrightarrow{p} h_1$ so that we again obtain joint convergence of the above random three-vectors. Now, take any $(t_1, t_2)' \in \mathbb{R}^2$ and fixed $x \in \mathbb{R}$,

$$\begin{aligned} P_{\theta_0,\gamma_{n,h}}(t_1 T_n(\theta_0) + t_2\hat{h}_{n,1} \leq x) &= P_{\theta_0,\gamma_{n,h}}(t_1|\widetilde{T}_{n,1}(\theta_0)| + t_2\hat{h}_{n,1} \leq x, |T_{n,2}| \leq c) \\ &\quad + P_{\theta_0,\gamma_{n,h}}(t_1|\hat{T}_{n,1}(\theta_0)| + t_2\hat{h}_{n,1} \leq x, |T_{n,2}| > c) \\ &\rightarrow P(t_1|\widetilde{Z}_{h,1}| + t_2\widetilde{h}_1 \leq x, |Z_{h,2}| \leq c) \\ &\quad + P(t_1|\hat{Z}_{h,1}| + t_2\widetilde{h}_1 \leq x, |Z_{h,2}| > c) \\ &= P(t_1 W_h + t_2\widetilde{h}_1 \leq x), \end{aligned}$$

since $W_h \overset{d}{\sim} |\widetilde{Z}_{h,1}|\mathbb{I}(|Z_{h,2}| \leq c) + |\hat{Z}_{h,1}|\mathbb{I}(|Z_{h,2}| > c)$ by (S11.10) of Andrews and Guggenberger (2009c). Hence, the Cramér-Wold device yields the joint convergence of $T_n(\theta_0)$ and $\hat{h}_{n,1}$. Finally, since $\hat{\gamma}_{n,2} \xrightarrow{p} h_2$, we have the joint convergence of $T_n(\theta_0)$ and $\hat{h}_n = (\hat{h}_{n,1}, \hat{\gamma}_{n,2})$. ∎

Here we show that (7) holds. Assume H_0 holds and the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP with $|h_1| < \infty$. We have already shown the joint convergence of the two random vectors to multivariate normal random vectors above. The asymptotic covariance matrix is all that remains to be shown. Andrews and Guggenberger (2009c) show $\text{Var}(\widetilde{Z}_{h,1}) = \text{Var}(\hat{Z}_{h,1}) = \text{Var}(Z_{h,2}) = 1$, $\text{Cov}(\widetilde{Z}_{h,1}, Z_{h,2}) = 0$ and $\text{Cov}(\hat{Z}_{h,1}, Z_{h,2}) = h_2$. Using the representation above, we obtain the following results in a similar fashion:

$$\begin{aligned} \text{Cov}(\widetilde{Z}_{h,1}, \widetilde{h}_1) &= \lim_{n\to\infty} E_{G_n}(Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'n^{-1}X^{\perp\prime}X^\perp e_1 Q_{n,11}^{-1/2} \\ &= \lim_{n\to\infty} (Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'Q_n e_1 Q_{n,11}^{-1/2} = \lim_{n\to\infty} (Q_n^{22})^{-1/2}Q_{n,22}^{-1}Q_{n,21}Q_{n,11}^{-1/2} \end{aligned}$$

A.10

$$= \lim_{n \to \infty} -\gamma_{n,h,2}(1 - \gamma_{n,h,2}^2)^{1/2} = -h_2(1 - h_2^2)^{1/2},$$

$$\text{Cov}(Z_{h,2}, \widetilde{h}_1) = \lim_{n \to \infty} E_{G_n}(Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'n^{-1}X^{\perp\prime}X^{\perp}(e_2 - Q_{n,12}Q_{n,11}^{-1}e_1)(Q_n^{22})^{1/2}$$

$$= \lim_{n \to \infty} (Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'Q_n(e_2 - Q_{n,12}Q_{n,11}^{-1}e_1)(Q_n^{22})^{1/2}$$

$$= \lim_{n \to \infty} (Q_n^{22})^{-1/2}Q_{n,22}^{-1}(Q_{n,22} - Q_{n,12}^2Q_{n,11}^{-1})(Q_n^{22})^{1/2}$$

$$= \lim_{n \to \infty} 1 - \gamma_{n,h,2}^2 = 1 - h_2^2,$$

$$\text{Cov}(\hat{Z}_{h,1}, \widetilde{h}_1) = \lim_{n \to \infty} E_{G_n}(Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'n^{-1}X^{\perp\prime}X^{\perp}(e_1 - Q_{n,12}Q_{n,22}^{-1}e_2)(Q_n^{11})^{1/2}$$

$$= \lim_{n \to \infty} (Q_n^{22})^{-1/2}Q_{n,22}^{-1}e_2'Q_n(e_1 - Q_{n,12}Q_{n,22}^{-1}e_2)(Q_n^{11})^{1/2}$$

$$= \lim_{n \to \infty} (Q_n^{22})^{-1/2}Q_{n,22}^{-1}(Q_{n,12} - Q_{n,12}Q_{n,22}^{-1}Q_{n,22})(Q_n^{11})^{1/2} = 0. \ \blacksquare$$

9.2 Testing when a Nuisance Parameter may be on a Boundary

To verify that the distribution function of W_h is given by (4), note that by (3),

$$W_h \overset{d}{\sim} \begin{cases} -Z_{h_2,1}, & \text{if } Z_{h_2,2} \geq -h_1 \\ -Z_{h_2,1} + h_2(Z_{h_2,2} + h_1), & \text{if } Z_{h_2,2} < -h_1 \end{cases}$$

so that

$$P(W_h \leq x) = P(-Z_{h_2,1} \leq x, Z_{h_2,2} \geq -h_1) \tag{A.7}$$
$$+ P(-Z_{h_2,1} + h_2(Z_{h_2,2} + h_1) \leq x, Z_{h_2,2} < -h_1). \tag{A.8}$$

Since $Z_{h_2,1} \overset{d}{\sim} h_2 Z_{h_2,2} + \sqrt{1 - h_2^2}\widetilde{Z}$, where $\widetilde{Z} \overset{d}{\sim} N(0,1)$ and $E[Z_{h_2,2}\widetilde{Z}] = 0$, we have that (A.8) is equal to

$$P\left(-\sqrt{1 - h_2^2}\widetilde{Z} + h_2 h_1 \leq x, Z_{h_2,2} < -h_1\right) = \Phi\left(\frac{x - h_2 h_1}{\sqrt{1 - h_2^2}}\right)\Phi(-h_1).$$

For similar reasons, $Z_{h_2,2}$ conditional on $Z_{h_2,1} = z_1$ is distributed $N(h_2 z_1, 1 - h_2^2)$ so that, letting $f(z_2|z_1)$ denote the conditional density, (A.7) is equal to

$$\int_{-x}^{\infty} \int_{-h_1}^{\infty} f(z_2|z_1)\phi(z_1)dz_2 dz_1 = \int_{-x}^{\infty} \left(1 - \int_{-\infty}^{-h_1} (1 - h_2^2)^{-1/2}\phi\left(\frac{z_2 - h_2 z_1}{(1 - h_2)^{1/2}}\right)dz_2\right)\phi(z_1)dz_1$$

$$= \int_{-x}^{\infty} \left(1 - \int_{-\infty}^{-h_1(1-h_2^2)^{-1/2}} \phi\left(\bar{z}_2 - \frac{h_2 z_1}{(1 - h_2^2)^{1/2}}\right)d\bar{z}_2\right)\phi(z_1)dz_1$$

$$= \int_{-x}^{\infty} \left(1 - \Phi\left(\frac{-h_1 - h_2 z}{(1 - h_2^2)^{1/2}}\right)\right)\phi(z)dz,$$

yielding the desired result. $\blacksquare$

Here we verify Assumption S-BM.2 and the claim that $\bar{\delta}$ can be set arbitrarily close to zero in this example. Assume that H_0 holds and the drifting sequence of parameters $\{\gamma_{n,h}\}$ characterizes the true DGP.

(i) First, $\hat{\gamma}_{n,2} = \hat{\rho}_n \xrightarrow{p} h_2$ and the central limit theorem provides that for $h_1 < \infty$, $n^{1/2}\bar{X}_{n,2}/\hat{\sigma}_{n,2} \xrightarrow{d} h_1 + Z_{h_2,2}$, where $Z_{h_2,2} \overset{d}{\sim} N(0,1)$. When $h_1 = \infty$, $n^{1/2}\bar{X}_{n,2}/\hat{\sigma}_{n,2} \xrightarrow{p} h_1$ since $\hat{\sigma}_{n,2} \xrightarrow{p} \sigma_2$, $n^{1/2}\gamma_{n,h,1} \to \infty$ and

$$n^{1/2}\bar{X}_{n,2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\sigma_2\gamma_{n,h,1} + e_{i2}] = \frac{\sigma_2}{n}\sum_{i=1}^{n}n^{1/2}\gamma_{n,h,1} + O_p(1)$$

where e_{i2} is iid with finite variance σ_2^2. Hence,

$$\widetilde{h} = \begin{cases} (h_1, h_2) & \text{if } h_1 = \infty \\ (h_1 + Z_{h_2,2}, h_2) & \text{if } h_1 \in \mathbb{R}_+. \end{cases}$$

Second, take any $\beta \in (0,1)$. As in the previous example, $P_{\theta_0,\gamma_{n,h}}(h \in I_\beta(\widetilde{h})) = 1$ when $h_1 = \infty$. If $h_1 \in \mathbb{R}_+$,

$$\begin{aligned} P(h \in I_\beta(\widetilde{h})) &= P(h_1 \in [\max\{0, h_1 + Z_{h_2,2} + z_\xi\}, \max\{0, h_1 + Z_{h_2,2} + z_{1-\beta+\xi}\}]) \\ &\geq P(h_1 \in [\max\{0, h_1 + Z_{h_2,2} + z_\xi\}, h_1 + Z_{h_2,2} + z_{1-\beta+\xi}]) \\ &= P(h_1 \in [h_1 + Z_{h_2,2} + z_\xi, h_1 + Z_{h_2,2} + z_{1-\beta+\xi}]) \\ &= P(Z_{h_2,2} \in [z_\xi, z_{1-\beta+\xi}]) = 1 - \beta \end{aligned}$$

since $P(h_1 < 0) = 0$. Third, as defined, $I_\beta(\mathfrak{h})$ is clearly compact-valued. Finally, to show continuity of $I_\beta(\mathfrak{h})$ over $\widetilde{H}$, note that for any $\beta \in (0,1)$, $I_\beta(\mathfrak{h}) = ([\max\{0, \mathfrak{h}_1 + z_\xi\}, \max\{0, \mathfrak{h}_1 + z_{1-\beta+\xi}\}], \mathfrak{h}_2)$. Now, take any $\mathfrak{h} \in \widetilde{H} = \mathbb{R}_\infty$ and let $\{\mathfrak{h}_k\}$ be a sequence in $\widetilde{H}$ such that $\mathfrak{h}_k \to \mathfrak{h}$. For each k, take any $b_k \in I_\beta(\mathfrak{h}_k)$ such that $b_k \to b$. Then we have four cases: (i) if $\mathfrak{h}_{1,k} = \infty$, $b_k = \mathfrak{h}_k = (\mathfrak{h}_{1,k} + z_\xi, \mathfrak{h}_{2,k}) = (\mathfrak{h}_{1,k} + z_{1-\beta+\xi}, \mathfrak{h}_{2,k})$; (ii) if $\mathfrak{h}_{1,k} + z_{1-\beta+\xi} \leq 0$, $b_k = (0, \mathfrak{h}_{2,k})$; (iii) if $\mathfrak{h}_{1,k} + z_\xi \leq 0$ and $\mathfrak{h}_{1,k} + z_{1-\beta+\xi} > 0$, $b_k = ((1 - a_k)(\mathfrak{h}_{1,k} + z_{1-\beta+\xi}), \mathfrak{h}_{2,k})$ for some $a_k \in [0,1]$; (iv) if $(\mathfrak{h}_{1,k} + z_\xi) \in (0,\infty)$, then $b_k = (a_k(\mathfrak{h}_{1,k} + z_\xi) + (1 - a_k)(\mathfrak{h}_{1,k} + z_{1-\beta+\xi}), \mathfrak{h}_{2,k})$ for some $a_k \in [0,1]$. In other terms, $b_k = (a_k \odot (\mathfrak{h}_{1,k} + z_\xi) + (1 - a_k) \odot (\mathfrak{h}_{1,k} + z_{1-\beta+\xi}), \mathfrak{h}_{2,k})$ for some $a_k \in [0,1]$, where $\odot$ is a binary operator defined for any pair $(\mathfrak{a}, \mathfrak{b}) \in [0,1] \times \widetilde{H}_1$ as follows:

$$\mathfrak{a} \odot \mathfrak{b} = \begin{cases} \mathfrak{a}\mathfrak{b}, & \text{if } \mathfrak{b} \in \mathbb{R}_+ \\ 0, & \text{if } \mathfrak{b} < 0 \text{ or both } \mathfrak{b} = \infty \text{ and } \mathfrak{a} = 0 \\ \mathfrak{b}, & \text{if } \mathfrak{b} = \infty \text{ and } \mathfrak{a} > 0. \end{cases}$$

Now note that for $(a, b) \in [0,1] \times \widetilde{H}_1$,

$$\mathfrak{f}(a, b) \equiv a \odot (b + z_\xi) + (1 - a) \odot (b + z_{1-\beta+\xi})$$

A.12

$$
= \begin{cases}
a(b + z_\xi) + (1 - a)(b + z_{1-\beta+\xi}), & \text{if } (b + z_\xi) \in \mathbb{R}_+ \\
(1 - a)(b + z_{1-\beta+\xi}), & \text{if } (b + z_{1-\beta+\xi}) \in \mathbb{R}_+ \text{ and } (b + z_\xi) < 0 \\
0, & \text{if } (b + z_{1-\beta+\xi}) < 0 \\
b + z_{1-\beta+\xi}, & \text{if } (b + z_\xi) = \infty
\end{cases}
$$

is clearly continuous in a. It is also continuous in b since $\lim_{b \to -z_\xi} \mathfrak{f}(a, b) = (1-a)(z_{1-\beta+\xi} - z_\xi)$, $\lim_{b \to -z_{1-\beta+\xi}} \mathfrak{f}(a, b) = 0$ and $\lim_{b \to \infty} \mathfrak{f}(a, b) = b + z_{1-\beta+\xi}$. Then, the same arguments used for the previous example provide that I_β is upper hemicontinuous. Similar reasoning yields lower hemicontinuity so that I_β is continuous for all $\beta \in (0, 1)$.

Very similar arguments to those used in the previous example establish the continuity of $I_\beta(\mathfrak{h})$ in β over $(0, 1)$ for all $\mathfrak{h} \in \widetilde{H}$. $\blacksquare$

9.3 Testing After Consistent Model Selection

To prove that Assumption M-B.1 holds for this example with the quantities as described in Section 2.2.1, note the following. As defined, part (i) of Assumption M-B.1 is implied by 1.(i) of Proposition A.2 of Leeb and Pötscher (2005) when we treat the domain of $\Phi(\cdot)$ as $\mathbb{R}_\infty$. Part (ii) of Assumption M-B.1 follows similarly from 2.(i) of Proposition A.2 in Leeb and Pötscher (2005). To show that part (iii) of Assumption M-B.1 holds, we must examine the cases for which $\gamma_{n,h,1}/k_n = \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to \pm 1$. First, if $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to 1$ and $c_n - \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to \infty$ or $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to -1$ and $c_n + \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to \infty$, then, by 1.(ii)-(iii) of Proposition A.2 in Leeb and Pötscher (2005), the null limit distribution of $T_n(\theta_0)$ is given by $J_h^{(1)}$. Second, if $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to 1$ and $c_n - \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to -\infty$ or $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to -1$ and $c_n - \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to -\infty$, then, by 2.(ii)-(iii) of Proposition A.2 in Leeb and Pötscher (2005), the null limit distribution of $T_n(\theta_0)$ is given by $J_h^{(2)}$. Third, suppose $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to 1$ and $c_n - \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to r$ for some $r \in \mathbb{R}$. If $h_1 \equiv \lim_{n \to \infty} \sqrt{n}\gamma_{n,h,1} = \lim_{n \to \infty} \sqrt{n}\beta_{2,n}\rho_n/\sigma_{\beta_2,n} = \infty$, the null limit distribution of $T_n(\theta_0)$, evaluated at x is given by

$$
\Phi(r) + \int_{-\infty}^{x} \frac{1}{h_{2,1}} \phi\left(\frac{u}{h_{2,1}}\right) \Phi\left(\frac{-r + h_{2,2}h_{2,1}^{-1}u}{(1 - h_{2,2}^2)^{1/2}}\right) du
$$

(see 3. of Proposition A.2 in Leeb and Pötscher, 2005). This distribution function is increasing in r and its limit as $r \to -\infty$ is equal to $J_h^{(2)}(x)$ so that the null limit distribution of $T_n(\theta_0)$ is stochastically dominated by $J_h^{(2)}$. If $h_1 = -\infty$, the null limit distribution of $T_n(\theta_0)$, evaluated at x is given by

$$
\int_{-\infty}^{x} \frac{1}{h_{2,1}} \phi\left(\frac{u}{h_{2,1}}\right) \Phi\left(\frac{-r + h_{2,2}h_{2,1}^{-1}u}{(1 - h_{2,2}^2)^{1/2}}\right) du
$$

(again see 3. of Proposition A.2 in Leeb and Pötscher, 2005), while $J_h^{(1)}(x) = 0$ for all $x \in \mathbb{R}$, i.e., it is the distribution function of a pointmass at ∞, and hence stochastically dominates. If $h_1 \in \mathbb{R}$, $h_{2,2} \equiv \lim_{n \to \infty} \rho_n = 0$ since $c_n \to \infty$ so that $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to \infty$. Hence, by 3. of

A.13

Proposition A.2 in Leeb and Pötscher (2005), the null limit distribution of $T_n(\theta_0)$, evaluated at x is given by $\Phi(r)\Phi(x/h_{2,1} + h_1) + \Phi(-r)\Phi(x/h_{2,1})$, which is stochastically dominated by $J_h^{(1)}(x)$ if $h_1 \le 0$ and by $J_h^{(2)}(x)$ if $h_1 > 0$. The proof of the fourth and final case for which $\sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n}c_n \to -1$ and $c_n + \sqrt{n}\beta_{2,n}/\sigma_{\beta_2,n} \to s$ for some $s \in \mathbb{R}$ is quite similar to the previous case and is hence omitted. ∎

To see that Assumption M-B.4 is satisfied, begin by noting that under H_0 and $\{\gamma_{n,h}\}$,

$$\sqrt{n}(\widetilde{\theta} - \theta_0) = \sqrt{n}(X_1'X_1)^{-1}X_1'X_2\beta_{2,n} + \sqrt{n}(X_1'X_1)^{-1}X_1'\epsilon$$
$$\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}\epsilon$$
$$\hat{h}_{n,1} = \sqrt{n}\rho_n\beta_{2,n}/\sigma_{\beta_2,n} + \sqrt{n}(\rho_n/\sigma_{\beta_2,n})(X_2'M_{X_1}X_2)^{-1}X_2'M_{X_1}\epsilon,$$

where X_i is the ith column of X for $i = 1, 2$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$, so that

$$\begin{pmatrix} \sqrt{n}(\widetilde{\theta} - \theta_0) \\ \hat{h}_{n,1} \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} -\sigma_{\theta,n}\sqrt{n}\gamma_{n,h,1} \\ \sqrt{n}\gamma_{n,h,1} \end{pmatrix}, \begin{pmatrix} \sigma_{\theta,n}^2(1-\rho_n^2) & 0 \\ 0 & \rho_n^2 \end{pmatrix}\right) \overset{d}{\longrightarrow} \begin{pmatrix} W_h^{(1)} \\ \widetilde{h}_1 \end{pmatrix},$$
$$\begin{pmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ \hat{h}_{n,1} \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} 0 \\ \sqrt{n}\gamma_{n,h,1} \end{pmatrix}, \begin{pmatrix} \sigma_{\theta,n}^2 & \sigma_{\theta,\beta_2,n}\rho_n/\sigma_{\beta_2,n} \\ \sigma_{\theta,\beta_2,n}\rho_n/\sigma_{\beta_2,n} & \rho_n^2 \end{pmatrix}\right) \overset{d}{\longrightarrow} \begin{pmatrix} W_h^{(2)} \\ \widetilde{h}_1 \end{pmatrix}.$$

Then, for any fixed $(t_1, t_2)' \in \mathbb{R}^2$ and $x \in \mathbb{R}$ and $\{\gamma_{n,h}\}$ such that $\zeta(\{\gamma_{n,h}\}) \notin \{-1, 1\}$,

$$P_{\theta_0,\gamma_{n,h}}(t_1 T_n(\theta_0) + t_2\hat{h}_{n,1} \le x) = P_{\theta_0,\gamma_{n,h}}(t_1\sqrt{n}(\widetilde{\theta} - \theta_0) + t_2\hat{h}_{n,1} \le x, |\hat{\zeta}| \le 1)$$
$$+ P_{\theta_0,\gamma_{n,h}}(t_1\sqrt{n}(\hat{\theta} - \theta_0) + t_2\hat{h}_{n,1} \le x, |\hat{\zeta}| > 1)$$
$$\longrightarrow \begin{cases} P(t_1 W_h^{(1)} + t_2\widetilde{h}_1 \le x), & \text{if } |\zeta(\{\gamma_{n,h}\})| < 1 \\ P(t_1 W_h^{(2)} + t_2\widetilde{h}_1 \le x), & \text{if } |\zeta(\{\gamma_{n,h}\})| > 1. \end{cases}$$

Hence, the Cramér-Wold device, along with the facts that $\gamma_{n,h,2,1} = \sigma_{\theta,n} \to \sigma_{\theta,\infty} = h_{2,1}$ and $\gamma_{n,h,2,2} = \rho_n \to \rho_\infty = h_{2,2}$, yields the desired result. ∎

To see that (9) holds, simply note that the limiting random vectors in (9) are the limits of $(\sqrt{n}(\widetilde{\theta} - \theta_0), \hat{h}_{n,1})'$ and $(\sqrt{n}(\hat{\theta} - \theta_0), \hat{h}_{n,1})'$ above. ∎

Assumption M-BM-PI.1(i) is trivially satisfied since $\eta_i(\cdot) = 0$ for $i = 1, 2$. Examining the distribution functions of $\widetilde{h}$ and $W_h^{(i)}$ for $i = 1, 2$ (see Sections 6 and 2.2.1, respectively), we can see that for $h \in \bar{H}$:

$$\sup_{\mathfrak{h} \in I_{\alpha-\delta}(\widetilde{h})} c_{\mathfrak{h}}^{(1)}(1-\delta) = -\widetilde{h}_1 h_{2,1} - h_{2,1}|h_{2,2}|z_{\xi(\alpha-\delta)} + h_{2,1}(1 - h_{2,2}^2)^{1/2}z_{1-\delta},$$

$$c_{B-A-PI}^{(1)}(\alpha, 0, \widetilde{h}) = c_{LF-PI}^{(1)}(\alpha, h_2) = \sup_{\mathfrak{h}_1 \in H_1}\{-\mathfrak{h}_1 h_{2,1} + h_{2,1}(1 - h_{2,2}^2)^{1/2}z_{1-\alpha}\} = \infty,$$

A.14

$$c_{min-B-PI}^{(1)}(\alpha,\widetilde{h}) = -\widetilde{h}_1 h_{2,1} + \inf_{\delta\in[\underline{\delta}^{(1)},\alpha-\bar{\delta}^{(1)}]}\{-h_{2,1}|h_{2,2}|z_{\xi(\alpha-\delta)} + h_{2,1}(1-h_{2,2}^2)^{1/2}z_{1-\delta}\},$$

$$\begin{pmatrix} W_h^{(1)} \\ \widetilde{h}_1 \end{pmatrix} \overset{d}{\sim} \left(\begin{pmatrix} -h_1 h_{2,1} \\ h_1 \end{pmatrix}, \begin{pmatrix} h_{2,1}^2(1-h_{2,2}^2) & 0 \\ 0 & h_{2,2}^2 \end{pmatrix} \right)$$

so that

$$W_h^{(1)} - c_{min-B-PI}^{(1)}(\alpha,\widetilde{h}) \overset{d}{\sim} N(0,h_{2,1}^2) + \inf_{\delta\in[\underline{\delta}^{(1)},\alpha-\bar{\delta}^{(1)}]}\{-h_{2,1}|h_{2,2}|z_{\xi(\alpha-\delta)} + h_{2,1}(1-h_{2,2}^2)^{1/2}z_{1-\delta}\},$$

which is an absolutely continuous random variable. Similarly, for $h\in\bar{H}$,

$$W_h^{(2)} - c_{min-B-PI}^{(2)}(\alpha,\widetilde{h}) \overset{d}{\sim} N(0,h_{2,1}^2) + \min\{h_{2,1}z_{1-\alpha}, \inf_{\delta\in[\underline{\delta}^{(2)},\alpha-\bar{\delta}^{(2)}]}h_{2,1}z_{1-\delta}\} \overset{d}{\sim} N(0,h_{2,1}^2) + h_{2,1}z_{1-\alpha}$$

so that Assumption M-BM-PI.1(ii) is satisfied. We can see from the above expressions that for $h\in\bar{H}$,

$$\widetilde{\delta}^{(1)} \equiv \mathrm{argmin}_{\delta\in[\underline{\delta}^{(1)},\alpha-\bar{\delta}^{(1)}]}\, c_{B-PI}^{(1)}(\alpha,\delta,\widetilde{h})$$
$$= \mathrm{argmin}_{\delta\in[\underline{\delta}^{(1)},\alpha-\bar{\delta}^{(1)}]}\{-h_{2,1}|h_{2,2}|z_{\xi(\alpha-\delta)} + h_{2,1}(1-h_{2,2}^2)^{1/2}z_{1-\delta}\},$$

which is nonrandom. Similarly, for $h\in\bar{H}$,

$$\widetilde{\delta}^{(2)} \equiv \mathrm{argmin}_{\delta\in[\underline{\delta}^{(2)},\alpha-\bar{\delta}^{(2)}]}\, c_{B-PI}^{(2)}(\alpha,\delta,\widetilde{h}) = \mathrm{argmin}_{\delta\in[\underline{\delta}^{(2)},\alpha-\bar{\delta}^{(2)}]}\, h_{2,1}z_{1-\delta} = h_{2,1}z_{1-\alpha+\bar{\delta}^{(2)}},$$

which is nonrandom, so that the MLLDs analog of Corollary SCF holds. This implies that Assumption M-BM-PI.1(iii) is satisfied with $\eta_i(\cdot) = 0$ for $i = 1,2$. This latter result also guarantees that part (b) of Assumption M-BM-PI.1(iv) is satisfied since

$$\liminf_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{min-B-PI}^{(2)}(\alpha,\hat{h}_n)) = P(W_h^{(2)} > h_{2,1}z_{1-\alpha}) = \alpha$$

for all $h\in\bar{H}$ with $\zeta(\{\gamma_{n,h}\})\in L$ by Assumption M-B.4 and the continuity given by Lemma BM. To find the sequence $\{\gamma_{n,h^{*(2)}}\}$, take any sequence $\{\gamma_n\}$ with $\lim_{n\to\infty}\sqrt{n}\gamma_{n,1}\in \mathrm{int}(\bar{L})\cap\bar{H}_1 = (-\infty,-1)\cup(1,\infty)$ and $\gamma_{n,2,2} = \rho_n = 1/c_n$ ($\gamma_{n,2,1}$ need only converge to some $h_{2,1}\in[\eta,M]$) so that $\lim_{n\to\infty}\gamma_{n,1}/k_n = \lim_{n\to\infty}\sqrt{n}\gamma_{n,1}\in\mathrm{int}(\bar{L})$. For this sequence, $h^{*(2)} = (\lim_{n\to\infty}\sqrt{n}\gamma_{n,1}, h_{2,1}, 0)$. $\blacksquare$

Moving on to Assumption M-BM-PI.2, note that $|c_{min-B-PI}^{(1)}(\alpha,\hat{h}_n)| \le |\sup_{\mathfrak{h}\in I_{\alpha-\delta}(\hat{h}_n)} c_\mathfrak{h}^{(1)}(1-\delta)|$ for any $\delta\in[\underline{\delta}^{(1)},\alpha-\bar{\delta}^{(1)}]$. Suppose $P_{\theta_0,\gamma_{n,h}}(|\sup_{\mathfrak{h}\in I_{\alpha-\delta}(\hat{h}_n)} c_\mathfrak{h}^{(1)}(1-\delta)| = \infty) > 0$ for some such δ and $h\in\bar{H}^c$. This means

$$P_{\theta_0,\gamma_{n,h}}\left(|\sup_{\mathfrak{h}_1\in[\hat{h}_{n,1}+\hat{\gamma}_{n,2,2}z_\xi,\hat{h}_{n,1}+\hat{\gamma}_{n,2,2}z_{1-\alpha+\delta+\xi}]}\{-\mathfrak{h}_1\hat{\gamma}_{n,2,1} + z_{1-\delta}\hat{\gamma}_{n,2,1}(1-\hat{\gamma}_{n,2,2}^2)^{1/2}\}| = \infty\right) > 0$$

A.15

and consequently

$$P_{\theta_0,\gamma_{n,h}}(-\infty \text{ or } \infty \in [\hat{h}_{n,1} + \hat{\gamma}_{n,2,2}z_\xi, \hat{h}_{n,1} + \hat{\gamma}_{n,2,2}z_{1-\beta+\xi}]) > 0.$$

But this is at odds with the fact that $\hat{h}_n \overset{d}{\sim} \sqrt{n}\gamma_{n,h,1} + N(0,\hat{\gamma}_{n,2,2}^2)$ under $P_{\theta_0,\gamma_{n,h}}$ and the definition of Γ for this problem, and we arrive at a contradiction. Also, we clearly have $P_{\theta_0,\gamma_{n,h}}(|c_{min-B-PI}^{(2)}(\alpha,\hat{h}_n)| = \infty) = 0$ since $c_{min-B-PI}^{(2)}(\alpha,\hat{h}_n) = \hat{\gamma}_{n,2,1}z_{1-\alpha}$. Hence, Assumption M-BM-PI.2(i) is satisfied. Now, consider any $h \in \bar{H}^c$ and $\{\gamma_{n,h}\}$ with $\zeta(\{\gamma_{n,h}\}) \in K$. Using the facts that, under $\gamma_{n,h}$,

$$c_{B-A-PI}^{(1)}(\alpha,0,\hat{h}_n) = c_{LF-PI}^{(1)}(\alpha,\hat{h}_n)) = \sup_{\mathfrak{h}_1 \in H_1}\{-\mathfrak{h}_1\hat{\gamma}_{n,2,1} + \hat{\gamma}_{n,2,1}(1-\hat{\gamma}_{n,2,2})^{1/2}\} = \infty,$$

$$\sup_{\mathfrak{h} \in I_{\alpha-\delta}(\hat{h}_n)} c_{\mathfrak{h}}^{(1)}(1-\delta) = -\sqrt{n}\gamma_{n,h,1} - \sqrt{n}\frac{\hat{\gamma}_{n,2,1}\hat{\gamma}_{n,2,2}X_2'M_{X_1}\epsilon}{\sigma_{\beta_2,n}X_2'M_{X_1}X_2}$$
$$- \hat{\gamma}_{n,2,1}\hat{\gamma}_{n,2,2}z_\xi + \hat{\gamma}_{n,2,1}(1-\hat{\gamma}_{n,2,2}^2)^{1/2}z_{1-\delta}$$

for any $\delta \in [\underline{\delta}^{(1)}, \alpha - \bar{\delta}^{(1)}]$, we have

$$c_{min-B-PI}^{(1)}(\alpha,\hat{h}_n) = -\sqrt{n}\gamma_{n,h,1} - \sqrt{n}\frac{\hat{\gamma}_{n,2,1}\hat{\gamma}_{n,2,2}X_2'M_{X_1}\epsilon}{\sigma_{\beta_2,n}X_2'M_{X_1}X_2}$$
$$+ \inf_{\delta \in [\underline{\delta}^{(1)}, \alpha-\bar{\delta}^{(1)}]}\{-\hat{\gamma}_{n,2,1}\hat{\gamma}_{n,2,2}z_\xi + \hat{\gamma}_{n,2,1}(1-\hat{\gamma}_{n,2,2}^2)^{1/2}z_{1-\delta}\}.$$

We also have

$$T_n(\theta_0) = \left(-\sqrt{n}\gamma_{n,h,1} + \sqrt{n}\frac{X_1'\epsilon}{X_1'X_1}\right)\mathbf{1}(\zeta(\{\gamma_{n,h}\}) + o_p(1) \leq 1)$$
$$+ \sqrt{n}\frac{X_1'M_{X_2}\epsilon}{X_1'M_{X_2}X_1}\mathbf{1}(\zeta(\{\gamma_{n,h}\}) + o_p(1) > 1)$$
$$= -\sqrt{n}\gamma_{n,h,1} + \sqrt{n}\frac{X_1'\epsilon}{X_1'X_1} + o_p(1)$$

so that

$$\limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(T_n(\theta_0) > c_{min-B-PI}^{(1)}(\alpha,\hat{h}_n))$$

$$= \limsup_{n\to\infty} P_{\theta_0,\gamma_{n,h}}(\sqrt{n}\frac{X_1'\epsilon}{X_1'X_1} + \sqrt{n}\frac{\hat{\gamma}_{n,2,1}\hat{\gamma}_{n,2,2}X_2'M_{X_1}\epsilon}{\sigma_{\beta_2,n}X_2'M_{X_1}X_2} \geq$$
$$\inf_{\delta \in [\underline{\delta}^{(1)}, \alpha-\bar{\delta}^{(1)}]}\{-\hat{\gamma}_{n,2,1}\hat{\gamma}_{n,2,2}z_\xi + \hat{\gamma}_{n,2,1}(1-\hat{\gamma}_{n,2,2}^2)^{1/2}z_{1-\delta}\})$$
$$= P(h_{2,1}(1-h_{2,2}^2)^{1/2}Z_1 + h_{2,1}h_{2,2}Z_2 \geq \inf_{\delta \in [\underline{\delta}^{(1)}, \alpha-\bar{\delta}^{(1)}]}\{-h_{2,1}h_{2,2}z_{\xi(\alpha-\delta)} + h_{2,1}(1-h_{2,2}^2)^{1/2}z_{1-\delta}\})$$
$$= P(h_{2,1}(1-h_{2,2}^2)^{1/2}Z_1 + h_{2,1}h_{2,2}Z_2 \geq -h_{2,1}h_{2,2}z_{\xi(\alpha-\tilde{\delta})} + h_{2,1}(1-h_{2,2}^2)^{1/2}z_{1-\tilde{\delta}})$$
$$\leq P(h_{2,1}(1-h_{2,2}^2)^{1/2}Z_1 \geq h_{2,1}(1-h_{2,2}^2)^{1/2}z_{1-\tilde{\delta}}) + P(h_{2,1}h_{2,2}Z_2 \geq h_{2,1}h_{2,2}z_{1-\xi(\alpha-\tilde{\delta})})$$

A.16

$$= \widetilde{\delta} + \xi(\alpha - \widetilde{\delta}) < \widetilde{\delta} + \alpha - \widetilde{\delta} = \alpha,$$

where $Z_1, Z_2 \overset{d}{\sim} N(0,1)$ and $\widetilde{\delta} = \mathrm{argmin}_{\delta \in [\underline{\delta}^{(1)}, \alpha - \bar{\delta}^{(1)}]} \{-h_{2,1}h_{2,2}z_{\xi(\alpha-\delta)} + h_{2,1}(1 - h_{2,2}^2)^{1/2}z_{1-\delta}\}$. Now consider any $h \in \bar{H}^c$ and $\{\gamma_{n,h}\}$ with $\zeta(\{\gamma_{n,h}\}) \in L$. Then,

$$P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n)) = P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > \hat{\gamma}_{n,2,1}z_{1-\alpha}) \to P(W_h^{(2)} > h_{2,1}z_{1-\alpha}) = \alpha.$$

Hence, Assumption M-BM-PI.2(ii) is satisfied. Finally, suppose $\zeta(\{\gamma_{n,h}\}) \in L^c \cap K^c$ for some $h \in \bar{H}^c$. Let $\widetilde{\gamma}_{n,h,1}^{(1)} = \chi_{(1)}\gamma_{n,h,1}$ and $\widetilde{\gamma}_{n,h,2}^{(1)} = \gamma_{n,h,2}$ for all n with $\chi_{(1)} \in (0,1)$ so that $\lim_{n\to\infty} \sqrt{n}\widetilde{\gamma}_{n,h,1}^{(1)} = \chi_{(1)} \lim_{n\to\infty} \sqrt{n}\gamma_{n,h,1} = h_1 = \pm\infty$ and, recalling that $L^c \cap K^c = \{-1,1\}$,

$$\lim_{n\to\infty} \sqrt{n}\widetilde{\gamma}_{n,h,1}^{(1)}/k_n = \chi_{(1)} \lim_{n\to\infty} \sqrt{n}\gamma_{n,h,1}/k_n \in \{-\chi_{(1)}, \chi_{(1)}\} \subset (-1,1) = K.$$

Under $\{\gamma_{n,h}\}$, by results derived above, $c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n) = -\sqrt{n}\gamma_{n,h,1}\gamma_{n,h,2,1} + O_p(1)$ so that

$$\mathrm{plim}_{n\to\infty} \frac{c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n(\widetilde{\gamma}_{n,h}^{(1)}))}{c^{(1)}_{min-B-PI}(\alpha, \hat{h}_n(\gamma_{n,h}))} = \mathrm{plim}_{n\to\infty} \frac{-\sqrt{n}\chi_{(1)}\gamma_{n,h,1}\gamma_{n,h,2} + O_p(1)}{-\sqrt{n}\gamma_{n,h,1}\gamma_{n,h,2} + O_p(1)} = \chi_{(1)} \leq 1.$$

Now let $\widetilde{\gamma}_{n,h,1}^{(2)} = \chi_{(2)}\gamma_{n,h,1}$ and $\widetilde{\gamma}_{n,h,2}^{(1)} = \gamma_{n,h,2}$ for all n with $\chi_{(2)} \in (1,\infty)$ so that $\lim_{n\to\infty} \sqrt{n}\widetilde{\gamma}_{n,h,1}^{(2)} = \chi_{(2)} \lim_{n\to\infty} \sqrt{n}\gamma_{n,h,1} = h_1$ and

$$\lim_{n\to\infty} \sqrt{n}\widetilde{\gamma}_{n,h,1}^{(2)}/k_n = \chi_{(2)} \lim_{n\to\infty} \sqrt{n}\gamma_{n,h,1}/k_n \in \{-\chi_{(2)}, \chi_{(2)}\} \subset [-\infty, -1) \cup (1, \infty] = L.$$

Under $\{\gamma_{n,h}\}$, $c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n) = \hat{\gamma}_{n,2,1}z_{1-\alpha}$ so that

$$\mathrm{plim}_{n\to\infty} \frac{c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n(\widetilde{\gamma}_{n,h}^{(2)}))}{c^{(2)}_{min-B-PI}(\alpha, \hat{h}_n(\gamma_{n,h}))} = \mathrm{plim}_{n\to\infty} 1 = 1.$$

Hence, Assumption M-BM-PI.2(iii) holds. ∎

10 Appendix III: Auxiliary Assumptions and Results

In this appendix, we provide some auxiliary results mentioned in the main text as well as some auxiliary assumptions enforced in the lemmas of Appendix I.

Corollary SCF. *Suppose Assumptions D, S-B.1, S-B.2(i) evaluated at the $\delta_i = \bar{\alpha}_i$ that corresponds to β_i, S-B.3 evaluated at β_i, S-B.4, S-BM.1 and S-BM.2 hold for $i = 1, \ldots, r$. If there are some $h^* \in H$ and $\tilde{i} \in \{1, \ldots, r\}$ such that $P(W_{h^*} \geq \sup_{\mathfrak{h} \in I_{\beta_{\tilde{i}}}(\tilde{h}^*)} c_{\mathfrak{h}}(1 - \bar{\alpha}_{\tilde{i}})) = \alpha$ and, for every $h \in H$, either (a) $P(c^S_{BM}(\alpha, \tilde{h}) = c^S_B(\alpha, \tilde{\delta}, \tilde{h})) = 1$ for some nonrandom $\tilde{\delta} \in [\underline{\delta}, \alpha - \bar{\delta}]$ or (b) $P(c^S_{BM}(\alpha, \tilde{h}) = c_{B-A}(\alpha, \beta_{\hat{i}}, \tilde{h})) = 1$ for some nonrandom $\hat{i} \in \{1, \ldots, r\}$, then*

$$\mathrm{AsySz}(\theta_0, c^S_{BM}(\alpha, \hat{h}_n)) = \alpha$$

when $\Delta = 0$.

Suppose instead that Assumptions D, PI, S-B.1, S-B.3 evaluated at β_i, S-B.4, PI, S-BA-PI(i) and (iii) evaluated at β_i and corresponding $\bar{a}_i(\cdot)$, S-BM.1 and S-BM.2 evaluated at pairs $(\underline{\delta}^m, \bar{\delta}^m)$ and $(\underline{\delta}_i, \bar{\delta}_i)$ hold for $i = 1, \ldots, r$. If there are some $h^* \in H$ and $\widetilde{i} \in \{1, \ldots, r\}$ such that $P(W_{h^*} \geq \sup_{\mathfrak{h} \in I_{\beta_{\widetilde{i}}}(\widetilde{h}^*)} c_{\mathfrak{h}}(1 - \bar{a}_{\widetilde{i}}(h_2^*))) = \alpha$ and, for every $h \in H$ either

(a) $P(c^S_{BM-PI}(\alpha, \widetilde{h}) = c^S_{B-PI}(\alpha, \widetilde{\delta}, \widetilde{h})) = 1$ for some nonrandom $\widetilde{\delta} \in [\underline{\delta}^m, \alpha - \bar{\delta}^m]$ or (b) $P(c^S_{BM-PI}(\alpha, \widetilde{h}) = c_{B-A-PI}(\alpha, \beta_{\hat{i}}, \widetilde{h})) = 1$ for some nonrandom $\hat{i} \in \{1, \ldots, r\}$, then

$$\mathrm{AsySz}(\theta_0, c^S_{BM-PI}(\alpha, \hat{h}_n)) = \alpha$$

when $\eta(\cdot) = 0$.

Proof: We provide the proof for the non-PI case as that for the PI case is quite similar. By the same arguments as those used in the proof of Theorem S-BM, for any $h \in H$,

$$\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c^S_{min-B}(\alpha, \hat{h}_n)) \leq P(W_h \geq c_{min-B}(\alpha, \widetilde{h})).$$

If (a) holds,

$$P(W_h \geq c_{min-B}(\alpha, \widetilde{h})) = P(W_h \geq c^S_B(\alpha, \widetilde{\delta}, \widetilde{h})) \leq \alpha,$$

where the inequality follows from (A.1). If (b) holds,

$$P(W_h \geq c_{min-B}(\alpha, \widetilde{h})) = P(W_h \geq c_{B-A}(\alpha, \beta_{\hat{i}}, \widetilde{h})) \leq \alpha.$$

We then obtain $\mathrm{AsySz}(\theta_0, c^S_{BM}(\alpha, \hat{h}_n)) \leq \alpha$ by identical arguments to those used in the proof of Theorem S-B. Finally, since $c^S_{BM}(\alpha, \hat{h}_n) \leq c_{B-A}(\alpha, \beta_{\widetilde{i}}, \hat{h}_n)$ for all $\gamma \in \Gamma$ and n,

$$\alpha = P(W_{h^*} \geq \sup_{\mathfrak{h} \in I_{\beta_{\widetilde{i}}}(\widetilde{h}^*)} c_{\mathfrak{h}}(1 - \bar{\alpha}_{\widetilde{i}})) \leq \mathrm{AsySz}(\theta_0, c_{B-A}(\alpha, \beta_{\widetilde{i}}, \hat{h}_n)) \leq \mathrm{AsySz}(\theta_0, c^S_{BM}(\alpha, \hat{h}_n)),$$

similarly to the proof of Theorem S-BA. ∎

The enforcement of Assumption S-B.2 for $\delta = \alpha$ in the above corrolary is only used to establish the lower bound of α on the asymptotic size. The new condition that ensures size control in the absence of a SCF holds, for example, in the non-PI context when $\widetilde{\delta}$ is invariant to $\widetilde{h}$ and in the PI context when $\widetilde{\delta}$ is invariant to $\widetilde{h}_1$.

The following proposition provides a sufficient condition for part (ii) of Assumption S-BM.3 to hold.

Proposition S-BM. *(i) For $\Delta > 0$, Assumption S-BM.3(ii) holds if the distribution function of $\widetilde{W}_h \equiv W_h - c_{min-B}(\alpha, \widetilde{h})$, $\widetilde{J}_h(\cdot)$ say, is continuous over $h \in H$ and strictly increasing at $\Delta(h)$ for all $h \in \hat{H} \equiv \{h \in H : \widetilde{J}_h(0) \leq 1 - \alpha\}$.*

(ii) For $\Delta = 0$, Assumption S-BM.3(ii) holds if there are some $h^* \in H$ and $\widetilde{i} \in \{1, \ldots, r\}$ such that $P(W_{h^*} \geq \sup_{\mathfrak{h} \in I_{\beta_{\widetilde{i}}}(\widetilde{h}^*)} c_{\mathfrak{h}}(1 - \bar{\alpha}_{\widetilde{i}})) = \alpha$.

Proof: (i) Note that for $h \in \hat{H}$, $\Delta(h) = \widetilde{J}_h^{-1}(1-\alpha)$ is continuous over $h \in \hat{H} \subset H$ (by the continuity of $\widetilde{J}_h$ in h) and that $\hat{H}$ is compact (being the preimage of a closed set of a continuous function). Hence, the extreme value theorem provides that $\Delta = \sup_{h \in \hat{H}} \Delta(h) = \Delta(h^*)$ for some $h^* \in \hat{H}$ so that $P(W_{h^*} \geq c_{BM}^S(\alpha, \widetilde{h}^*)) = P(W_{h^*} \geq c_{min-B}(\alpha, \widetilde{h}^*) + \Delta(h^*)) = \alpha$ since $h^* \in \hat{H}$.

(ii) $P(W_{h^*} \geq c_{BM}^S(\alpha, \widetilde{h}^*)) \leq \alpha$ by construction and $P(W_{h^*} \geq c_{BM}^S(\alpha, \widetilde{h}^*)) \geq P(W_{h^*} \geq \sup_{\mathfrak{h} \in I_{\beta_{\widetilde{i}}}(\widetilde{h}^*)} c_{\mathfrak{h}}(1-\bar{\alpha}_{\widetilde{i}})) = \alpha$. ∎

The following two assumptions are used in Lemma M-B in Appendix I.

Assumption M-B.6. *Consider some fixed $(\alpha, \delta) \in (0,1) \times [0, \alpha]$ and any $h \in \bar{H}^c$.*
(i) For any finite n and $i = 1, 2$,

$$P_{\theta_0, \gamma_{n,h}}(|\sup_{\mathfrak{h} \in I_{\alpha-\delta}(\hat{h}_n)} c_{\mathfrak{h}}^{(i)}(1-\delta)| < \infty) = 1.$$

(ii) If $\zeta(\{\gamma_{n,h}\}) \in K$, then $\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^{(1)}(\alpha, \delta, \hat{h}_n)) \leq \alpha$. If $\zeta(\{\gamma_{n,h}\}) \in L$, then $\limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h}}(T_n(\theta_0) > c_B^{(2)}(\alpha, \delta, \hat{h}_n)) \leq \alpha$.
(iii) If $\zeta(\{\gamma_{n,h}\}) \in L^c \cap K^c$, there are some $\{\widetilde{\gamma}_{n,h}^{(1)}\}, \{\widetilde{\gamma}_{n,h}^{(2)}\} \subset \Gamma$ such that $\zeta(\{\widetilde{\gamma}_{n,h,1}^{(1)}\}) \in K$ and $\zeta(\{\widetilde{\gamma}_{n,h,1}^{(2)}\}) \in L$ and

$$\text{plim}_{n \to \infty} \frac{\sup_{\mathfrak{h} \in I_{\alpha-\delta}(\hat{h}_n(\widetilde{\gamma}_{n,h}^{(i)}))} c_{\mathfrak{h}}^{(i)}(1-\delta)}{\sup_{\mathfrak{h} \in I_{\alpha-\delta}(\hat{h}_n(\gamma_{n,h}))} c_{\mathfrak{h}}^{(i)}(1-\delta)} \leq 1$$

wp 1 for $i = 1, 2$, where $\hat{h}_n(\gamma_{n,h})$ $(\hat{h}_n(\widetilde{\gamma}_{n,h}^{(i)}))$ denotes the estimator of Assumption B.3 when H_0 and the drifting sequence of parameters $\{\gamma_{n,h}\}$ $(\{\widetilde{\gamma}_{n,h}^{(i)}\})$ characterize the true DGP.

Similar comments to those following Assumption M-BM.4 apply to the above assumption. The following assumption is used only to establish a lower bound on the size of a test using M-Bonf CVs. It is not required to show size control.

Assumption M-B-LB. *Either (a) $\sup_{\mathfrak{h} \in H} c_{\mathfrak{h}}^{(1)}(1-\delta) = c_{h^{*(1)}}^{(1)}(1-\delta)$ for some $h^{*(1)} \in H$ and there is some $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \text{int}(\bar{K})$ or (b) $\sup_{\mathfrak{h} \in H} c_{\mathfrak{h}}^{(2)}(1-\delta) = c_{h^{*(2)}}^{(2)}(1-\delta)$ for some $h^{*(2)} \in H$ and there is some $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \text{int}(\bar{L})$.*

Like Assumption M-B.2, this assumption is also essentially a continuity condition. An appeal to the continuity in h of one of the localized quantiles and the extreme value theorem can be used to verify it so long as the corresponding condition on the normalized drifting sequence is shown to hold.

We replace Assumption M-B-LB by the following to establish the lower bound on the asymptotic size of the corresponding test using PI M-Bonf CVs.

Assumption M-B-PI-LB. *Consider some fixed $\delta \in (0,1)$. For some $h_2^* \in H_2$, either (a)* $\sup_{\mathfrak{h}_1 \in H_1} c^{(1)}_{(\mathfrak{h}_1, h_2^*)}(1-\delta) = c^{(1)}_{(h_1^{*(1)}, h_2^*)}(1-\delta)$ *for some* $h_1^{*(1)} \in H_1$, $\sup_{\mathfrak{h}_1 \in H_1} c^{(1)}_{(\mathfrak{h}_1, \cdot)}(1-\delta)$ *is continuous at h_2^* as a function into $\mathbb{R}_\infty$ and there is some $\{\gamma_{n,(h_1^{*(1)}, h_2^*)}\} \in \Gamma$ with $\zeta(\{\gamma_{n,(h_1^{*(1)}, h_2^*)}\}) \in$ $\mathrm{int}(\bar{K})$ or (b) $\sup_{\mathfrak{h}_1 \in H_1} c^{(2)}_{(\mathfrak{h}_1, h_2^*)}(1-\delta) = c^{(2)}_{(h_1^{*(2)}, h_2^*)}(1-\delta)$ for some $h_1^{*(2)} \in H_1$, $\sup_{\mathfrak{h}_1 \in H_1} c^{(2)}_{(\mathfrak{h}_1, \cdot)}(1-\delta)$ is continuous at h_2^* as a function into $\mathbb{R}_\infty$ and there is some $\{\gamma_{n,(h_1^{*(2)}, h_2^*)}\} \in \Gamma$ with $\zeta(\{\gamma_{n,(h_1^{*(2)}, h_2^*)}\}) \in \mathrm{int}(\bar{L})$.*

Similar comments with regard to continuity to those following Assumption M-B-LB apply here as well.

We now present formal statements regarding the asymptotic size of tests using (PI) M-Bonf-Adj CVs as well as additional imposed assumptions. The proofs of the following two corollaries are very similar to those found in Appendix I and are hence omitted.

Assumption M-BA. *(i) $P(W_h^{(i)} = c^{(i)}_{B-A}(\alpha, \beta, \widetilde{h})) = 0$ for all $h \in \bar{H}$ and $i = 1, 2$.*

(ii) Either (a) $P(W_{h^{*(1)}} \geq c^{(1)}_{B-A}(\alpha, \beta, \widetilde{h}^{*(1)})) = \alpha$ for some $h^{*(1)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \mathrm{int}(\bar{K})$ or (b) $P(W_{h^{*(2)}} \geq c^{(2)}_{B-A}(\alpha, \beta, \widetilde{h}^{*(2)})) = \alpha$ for some $h^{*(2)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \mathrm{int}(\bar{L})$.

Similar remarks to those following Assumption S-BA apply here.

Corollary M-BA. *Under Assumptions D, M-B.1, M-B.2(i) evaluated at $\delta = \bar{\alpha}^{(1)}$ for $i = 1$ and $\delta = \bar{\alpha}^{(2)}$ for $i = 2$, M-B.3 through M-B.6, replacing "$\alpha - \delta$" by β and "δ" by $\bar{\alpha}^{(1)}$ for $i = 1$ and $\bar{\alpha}^{(2)}$ for $i = 2$ in parts (i) and (iii) of Assumption M-B.6 and "$c_B^{(i)}$" by "$c_{B-A}^{(i)}$" for $i = 1, 2$ in part (ii) of Assumption M-B.6, and M-BA,*

$$\mathrm{AsySz}(\theta_0, c^M_{B-A}(\alpha, \beta, \hat{h}_n)) = \alpha.$$

Assumption M-BA-PI. *(i) For $i = 1, 2$, $\bar{a}^{(i)} : H_2 \to [\underline{\delta}^{(i)}, \alpha - \bar{\delta}^{(i)}]$ is a continuous function.*

(ii) For $i = 1, 2$, $P(W_h^{(i)} = c^{(i)}_{B-A-PI}(\alpha, \beta, \widetilde{h})) = 0$ for all $h \in \bar{H}$.

(iii) For $i = 1, 2$, $P(W_h^{(i)} \geq \sup_{\mathfrak{h} \in I_\beta(\widetilde{h})} c^{(i)}_{\mathfrak{h}}(1 - \bar{a}^{(i)}(h_2))) \leq \alpha$ for all $h \in \bar{H}$.

(iv) Either (a) $P(W_{h^{*(1)}} \geq c^{(1)}_{B-A-PI}(\alpha, \beta, \widetilde{h}^{*(1)})) = \alpha$ for some $h^{*(1)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \mathrm{int}(\bar{K})$ or (b) $P(W_{h^{*(2)}} \geq c^{(2)}_{B-A-PI}(\alpha, \beta, \widetilde{h}^{*(2)})) = \alpha$ for some $h^{*(2)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \mathrm{int}(\bar{L})$.

Similar remarks to those following Assumption S-BA-PI apply here.

Corollary M-BA-PI. *Under Assumptions D, PI, M-B.1, M-BM.1, M-B.3 through M-B.6, replacing "$\alpha - \delta$" by β and "δ" by $\bar{a}^{(1)}(\hat{\gamma}_{n,2})$ for $i = 1$ and $\bar{a}^{(2)}(\hat{\gamma}_{n,2})$ for $i = 2$ in parts (i) and (iii) of Assumption M-B.6 and "$c_B^{(i)}$" by "$c_{B-A-PI}^{(i)}$" for $i = 1, 2$ in part (ii) of Assumption M-B.6, and M-BA-PI,*

$$\mathrm{AsySz}(\theta_0, c^M_{B-A-PI}(\alpha, \beta, \hat{h}_n)) = \alpha.$$

We now present the final two sufficient conditions, useful for verifying Assumptions imposed in Theorem M-BM and Corollary M-BM-PI.

A.20

Proposition M-BM. *Suppose Assumptions D, M-B.1, M-B.2(i) evaluated at $\delta_i^{(1)} = \bar{\alpha}_i^{(1)}$ corresponding to β_i and $c_h^{(1)}$ and $\delta_i^{(2)} = \bar{\alpha}_i^{(2)}$ corresponding to β_i and $c_h^{(2)}$, either (a) M-B.3 evaluated at β_i or (b) $c_{B-A}^{(1)}(\alpha, \beta_i, h)$ and $c_{B-A}^{(2)}(\alpha, \beta_i, h)$ are invariant to h, M-B.4, M-B.5, M-BM.1, M-BM.2 and M-BM.3(i) hold for $i = 1, \ldots, r$. If either (a) $P(W_{h^{*(1)}}^{(1)} \geq c_{min-B}^{(1)}(\alpha, \widetilde{h}^{*(1)}) + \Delta_1) = \alpha$ for some $h^{*(1)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \text{int}(\bar{K})$ or (b) $P(W_{h^{*(2)}}^{(2)} \geq c_{min-B}^{(2)}(\alpha, \widetilde{h}^{*(2)}) + \Delta_2) = \alpha$ for some $h^{*(2)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \text{int}(\bar{L})$, then Assumption M-BM.3(ii) holds.*

Proof: Suppose (a) holds. Then,

$$
\begin{aligned}
\alpha &\geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h^{*(1)}}}(T_n(\theta_0) > c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1) \\
&\geq \liminf_{n \to \infty} P_{\theta_0, \gamma_{n,h^{*(1)}}}(T_n(\theta_0) > c_{min-B}^{(1)}(\alpha, \hat{h}_n) + \Delta_1) \\
&\geq P(W_{h^{*(1)}}^{(1)} > c_{min-B}^{(1)}(\alpha, \widetilde{h}^{*(1)}) + \Delta_1) = \alpha,
\end{aligned}
$$

where the first inequality is shown in the proof of Theorem M-BM. The proof when (b) holds is very similar. ∎

Proposition M-BM-PI. *Suppose Assumptions D, PI, M-B.1, either (a) M-B.3 evaluated at β_i or (b) $c_{B-A}^{(1)}(\alpha, \beta_i, h)$ and $c_{B-A}^{(2)}(\alpha, \beta_i, h)$ are invariant to h_1 for $i = 1, \ldots, r$, M-B.4, M-B.5, M-BM.1, M-BM.2 and M-BM-PI.1(i)-(iii) hold. If either (a) $P(W_{h^{*(1)}}^{(1)} \geq c_{min-B-PI}^{(1)}(\alpha, \widetilde{h}^{*(1)}) + \eta_1(h_2^{*(1)})) = \alpha$ for some $h^{*(1)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(1)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(1)}}\}) \in \text{int}(\bar{K})$ or (b) $P(W_{h^{*(2)}}^{(2)} \geq c_{min-B-PI}^{(2)}(\alpha, \widetilde{h}^{*(2)}) + \eta_2(h_2^{*(2)})) = \alpha$ for some $h^{*(2)} \in \bar{H}$ and there is some $\{\gamma_{n,h^{*(2)}}\} \subset \Gamma$ with $\zeta(\{\gamma_{n,h^{*(2)}}\}) \in \text{int}(\bar{L})$, then Assumption M-BM-PI.1(iv) holds.*

Proof: Suppose (a) holds. Then,

$$
\begin{aligned}
\alpha &\geq \limsup_{n \to \infty} P_{\theta_0, \gamma_{n,h^{*(1)}}}(T_n(\theta_0) > c_{min-B-PI}^{(1)}(\alpha, \hat{h}_n) + \eta_1(\hat{\gamma}_{n,2})) \\
&\geq \liminf_{n \to \infty} P_{\theta_0, \gamma_{n,h^{*(1)}}}(T_n(\theta_0) > c_{min-B-PI}^{(1)}(\alpha, \hat{h}_n) + \eta_1(\hat{\gamma}_{n,2})) \\
&\geq P(W_{h^{*(1)}}^{(1)} > c_{min-B-PI}^{(1)}(\alpha, \widetilde{h}^{*(1)}) + \eta_1(h_2^{*(2)})) = \alpha,
\end{aligned}
$$

where the first inequality is shown in the proof of Corollary M-BM-PI. The proof when (b) holds is very similar. ∎

These sufficient conditions are useful because they are typically the direct byproducts of the construction of the SCFs Δ_i (or SCF functions $\eta_i(\cdot)$) so that the only part that needs to be verified is the existence of the corresponding parameter sequences.

References

Andrews, D. W. K., 1999. Estimation when a parameter is on a boundary. Econometrica 67, 1341–1383.

Andrews, D. W. K., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683–734.

Andrews, D. W. K., Barwick, P. J., 2011. Inference for parameters defined by moment inequalities: A recommended moment selection procedure, forthcoming in Econometrica.

Andrews, D. W. K., Cheng, X., 2012. Estimation and inference with weak, semi-strong and strong identification. Econometrica 80, 2153–2211.

Andrews, D. W. K., Cheng, X., Guggenberger, P., 2011. Generic results for establishing the asymptotic size of confidence sets and tests, Cowles Foundation Discussion Paper No. 1813.

Andrews, D. W. K., Guggenberger, P., 2009a. Hybrid and size-corrected subsampling methods. Econometrica 77, 721–762.

Andrews, D. W. K., Guggenberger, P., 2009b. Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. Journal of Econometrics 152, 19–27.

Andrews, D. W. K., Guggenberger, P., 2009c. Supplement to 'hybrid and size-corrected subsampling methods'. Econometrica Supplementary Material.

Andrews, D. W. K., Guggenberger, P., 2010a. Applications of subsampling, hybrid, and size-correction methods. Journal of Econometrics 158, 285–305.

Andrews, D. W. K., Guggenberger, P., 2010b. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory 26, 426–468.

Andrews, D. W. K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–157.

Berger, R. L., Boos, D. D., 1994. P values maximized over a confidence set for the nuisance parameter. Journal of the American Statistical Association 89, 1012–1016.

Chaudhuri, S., Zivot, E., 2011. A new method of projection-based inference in GMM with weakly identified nuisance parameters. Journal of Econometrics 164, 239–251.

Cheng, X., 2008. Robust confidence intervals in nonlinear regression under weak identification, Unpublished Manuscript, Department of Economics, University of Pennsylvania.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65, 1365–1387.

Elliott, G., Müller, U. K., 2012. Pre and post break parameter inference, Unpublished Manuscript, Department of Economics, Princeton University.

Elliott, G., Müller, U. K., Watson, M. W., 2012. Nearly optimal tests when a nuisance parameter is present under the null hypothesis, Unpublished Manuscript, Department of Economics, Princeton University.

Guggenberger, P., 2010. The impact of a Hausman pretest on the asymptotic size of a hypothesis test. Econometric Theory 26, 369–382.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., Chen, L., 2012. On the asymptotic sizes of subset Anderson-Rubin and Lagrange multiplier tests in linear instrumental variables regression, forthcoming in Econometrica.

Hansen, B. E., 2005a. Challenges for econometric model selection. Econometric Theory 21, 60–68.

Hansen, B. E., 2007. Least squares model averaging. Econometrica 75, 1175–1189.

Hansen, B. E., 2012. Efficient shrinkage in parametric models, Unpublished Manuscript, Department of Economics, University of Wisconsin.

Hansen, P. R., 2005b. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.

Kabaila, P., 1998. Valid confidence intervals in regression after variable selection. Econometric Theory 14, 463–482.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental varaibles regression. Econometrica 70, 1781–1803.

Leeb, H., 2006. The distribution of a linear predictor after model selection: unconditional finite-sample distributions and asymptotic approximations. In: 2nd Lehmann Symposium-Optimality. Vol. 49 of Institute of Mathematical Statistics Lecture Notes-Monograph Series. Institute of Mathematical Statistics, Beachwood, OH, pp. 291–311.

Leeb, H., Pötscher, B. M., 2005. Model selection and inference: facts and fiction. Econometric Theory 21, 21–59.

Leeb, H., Pötscher, B. M., 2006. Can one estimate the conditional distribution of post-model-selection estimators? The Annals of Statistics 34, 2554–2591.

Leeb, H., Pötscher, B. M., 2008. Can one estimate the unconditional distribution of post-model-selection estimators? Econometric Theory 24, 338–376.

Leeb, H., Pötscher, B. M., 2012. Testing in the presence of nuisance parameters: some comments on tests post-model-selection and random critical values, Unpublished Manuscript, Department of Statistics, University of Vienna.

Loh, W.-Y., 1985. A new method for testing separate families of hypotheses. Journal of the American Statistical Association 80, 362–368.

Mikusheva, A., 2007. Uniform inference in autoregressive models. Econometrica 75, 1411–1452.

Moon, H. R., Schorfheide, F., 2009. Estimation with overidentifying inequality moment conditions. Journal of Econometrics 153, 136–154.

Pötscher, B. M., Leeb, H., 2009. On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. Journal of Multivariate Analysis 100, 2065–2082.

Pötscher, B. M., Schneider, U., 2010. Confidence sets based on penalized maximum likelihood estimators. Electronic Journal of Statistics 4, 334–360.

Romano, J. P., Shaikh, A. M., Wolf, M., 2012. A simple two-step method for testing moment inequalities with an application to inference in partially identified models, Unpublished Manuscript, Stanford University, Department of Economics.

Romano, J. P., Wolf, M., 2000. Finite sample nonparametric inference and large sample efficiency. Annals of Statistics 28, 756–778.

Shi, X., 2011. Size distortion and modification of classical Vuoung tests, Unpublished Manuscript, Department of Economics, University of Wisconsin, Madison.

Silvapulle, M. J., 1996. A test in the presence of nuisance parameters. Journal of the American Statistical Association 91, 1690–1693.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Stock, J. H., 1991. Confidence intervals for the largest autoregressive root in the US macroeconomics time series. Journal of Monetary Economics 28, 435–459.

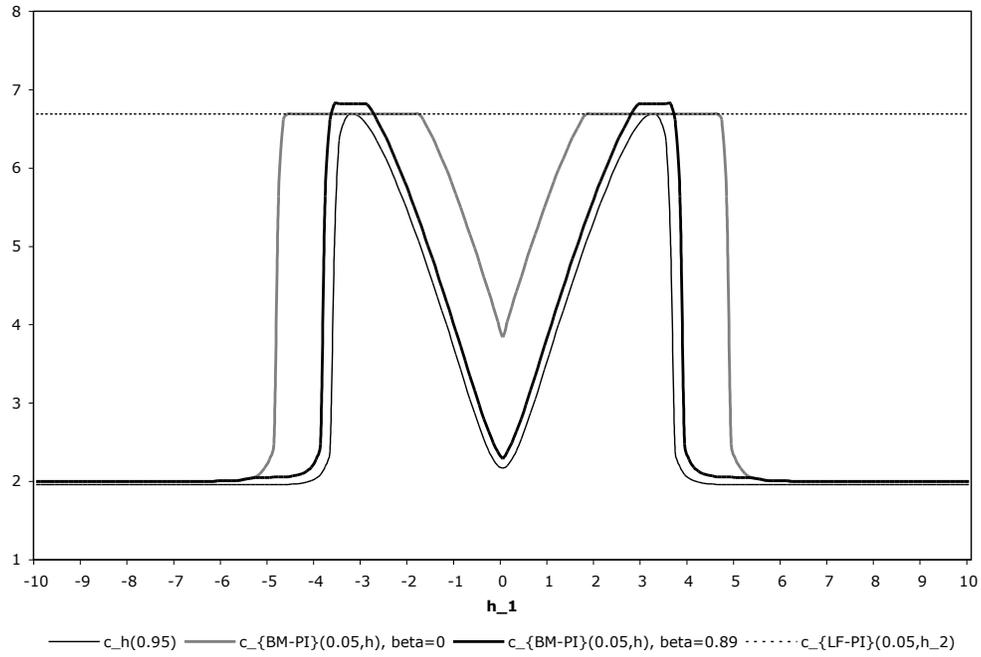Stoye, J., 2009. More on confidence intervals for partially identified parameters. Econometrica 77, 1299–1315.

Table 1: Maximum (Over β_2) Null Rejection Probabilities for Different Values of γ_2 in the Post-Conservative Model Selection Example

$|\gamma_2| =$	0	0.2	0.4	0.6	0.8	0.9	0.95	0.99	Max
$\beta = 0$, w/o SCF	0.051	0.052	0.051	0.048	0.046	0.048	0.047	0.048	0.052
$\beta = 0.89$, w/o SCF	0.052	0.050	0.051	0.050	0.048	0.048	0.048	0.086	0.086
$\beta = 0.89$, w/ SCF	0.052	0.050	0.050	0.050	0.048	0.048	0.048	0.086	0.086

Table 2: Maximum (Over β_2) Null Rejection Probabilities for Different Values of $\gamma_{2,2}$ in the Post Consistent Model Selection Example

$|\gamma_{2,2}| =$	0	0.2	0.4	0.6	0.8	0.9	0.95	0.99	Max
$\beta = 0$, w/o SCF $=$	0.052	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.052

Figure 1: Post-Conservative MS Critical Values, $h_2 = 0.9$
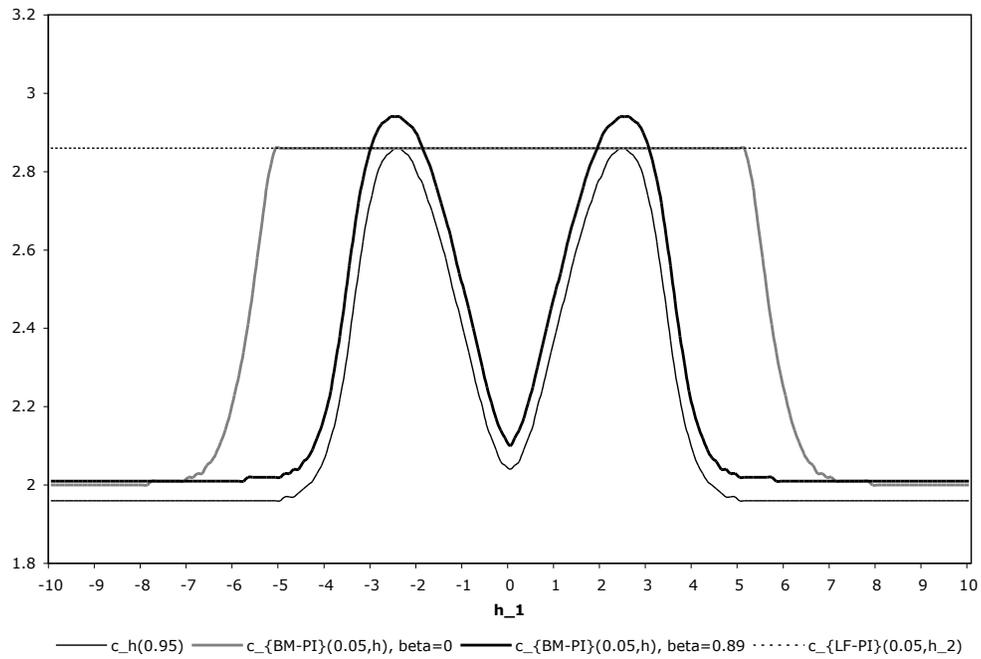
Figure 2: Post-Conservative MS Critical Values, $h_2 = 0.6$
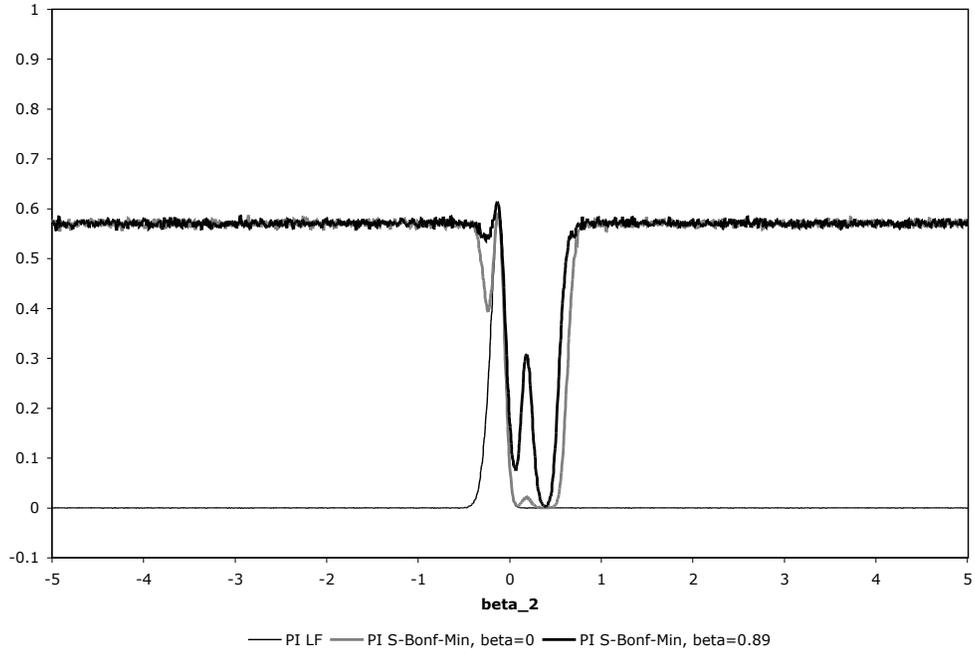
Figure 3: Post-Conservative MS Power, $\gamma_2 = 0.9$, $\theta = 0.2$
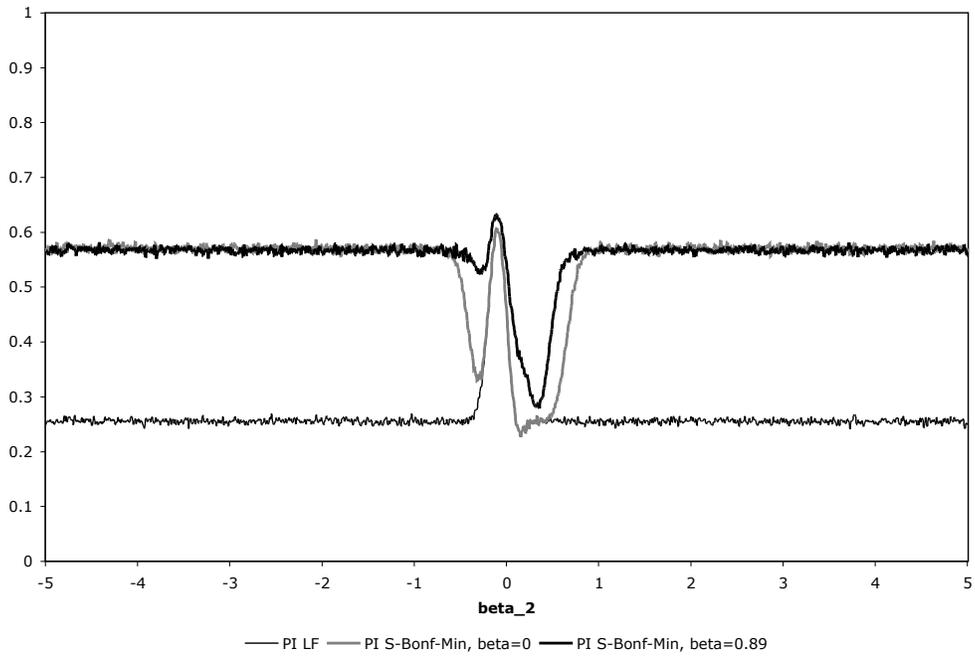
Figure 4: Post-Conservative MS Power, $\gamma_2 = 0.6$, $\theta = 0.2$
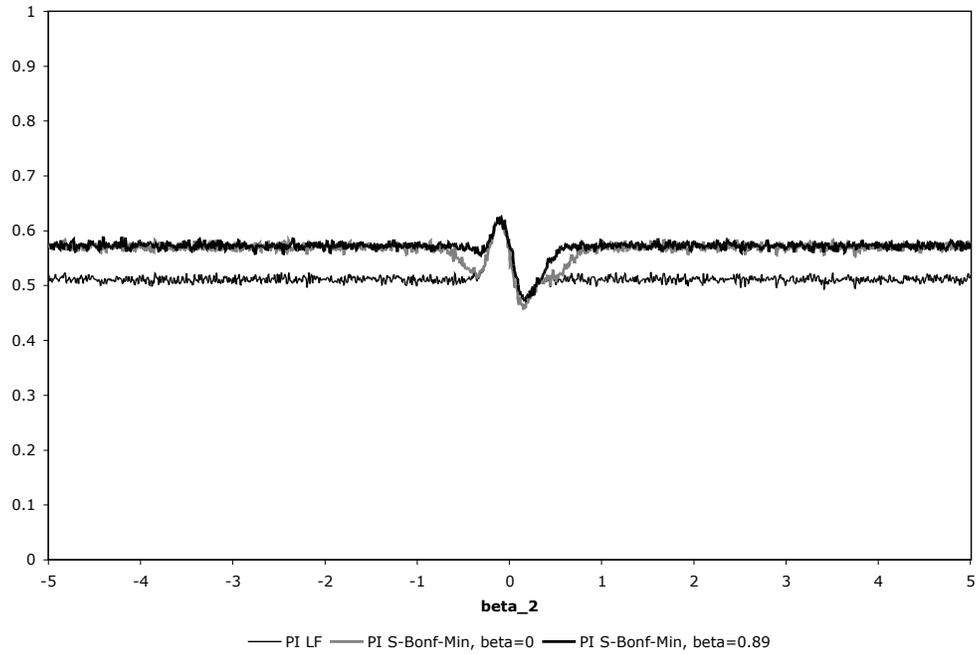
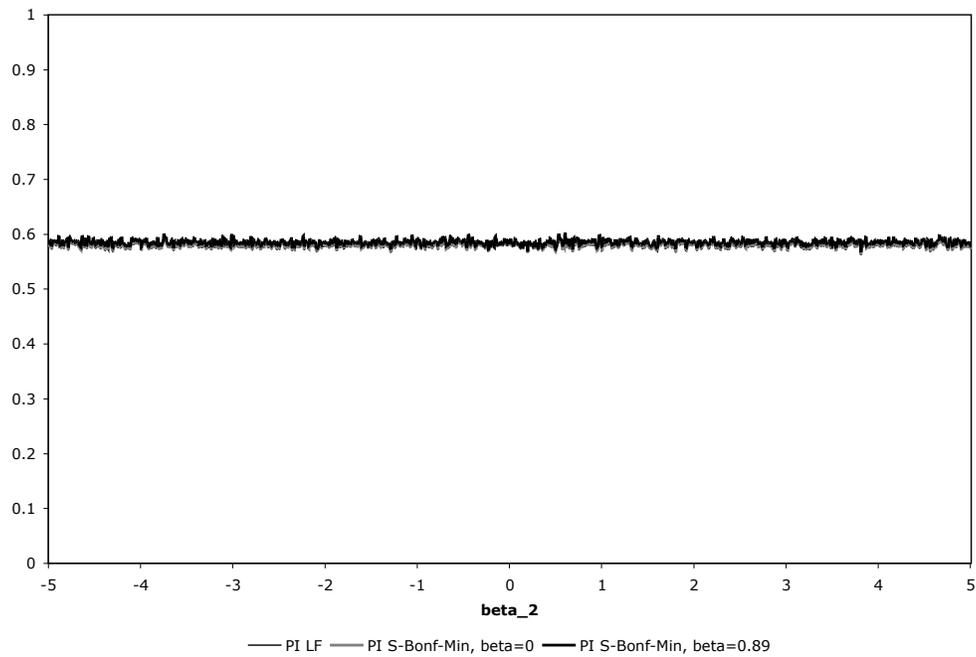Figure 5: Post-Conservative MS Power, $\gamma_2 = 0.3$, $\theta = 0.2$

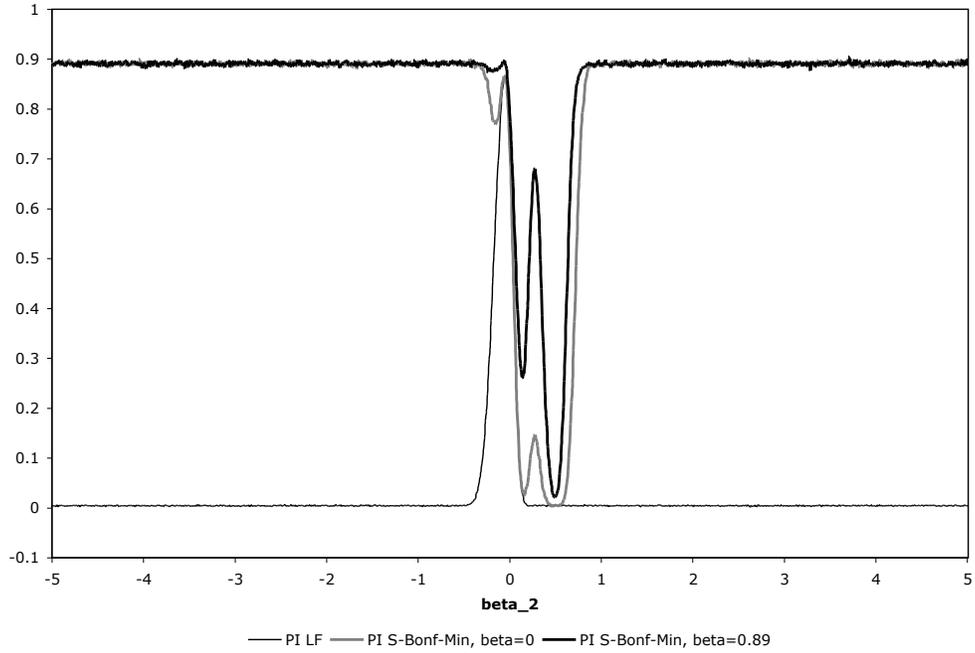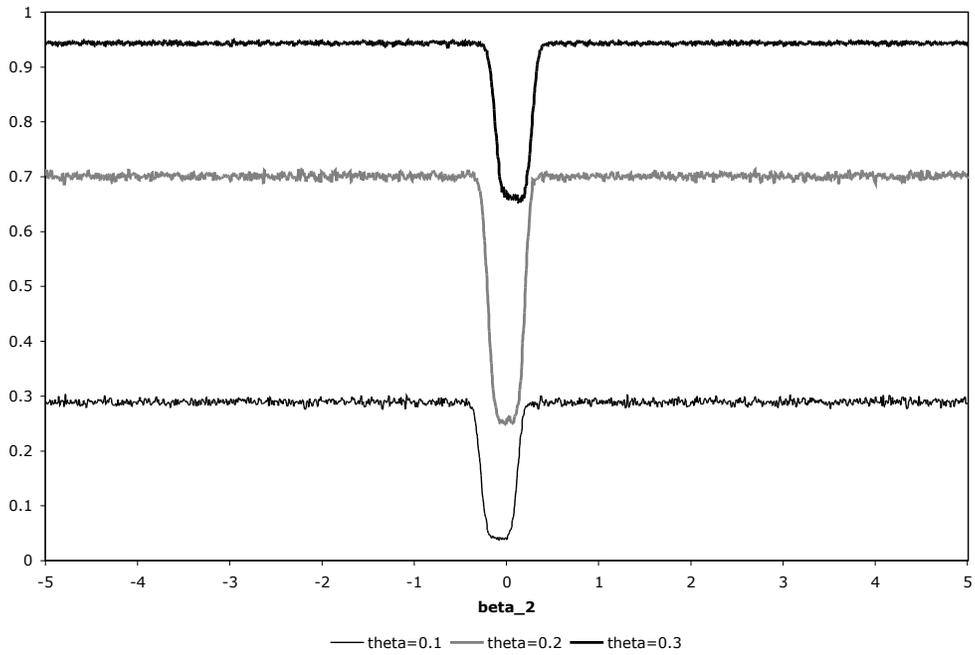Figure 6: Post-Conservative MS Power, $\gamma_2 = 0$, $\theta = 0.2$

Figure 7: Post-Conservative MS Power, $\gamma_2 = 0.9$, $\theta = 0.3$

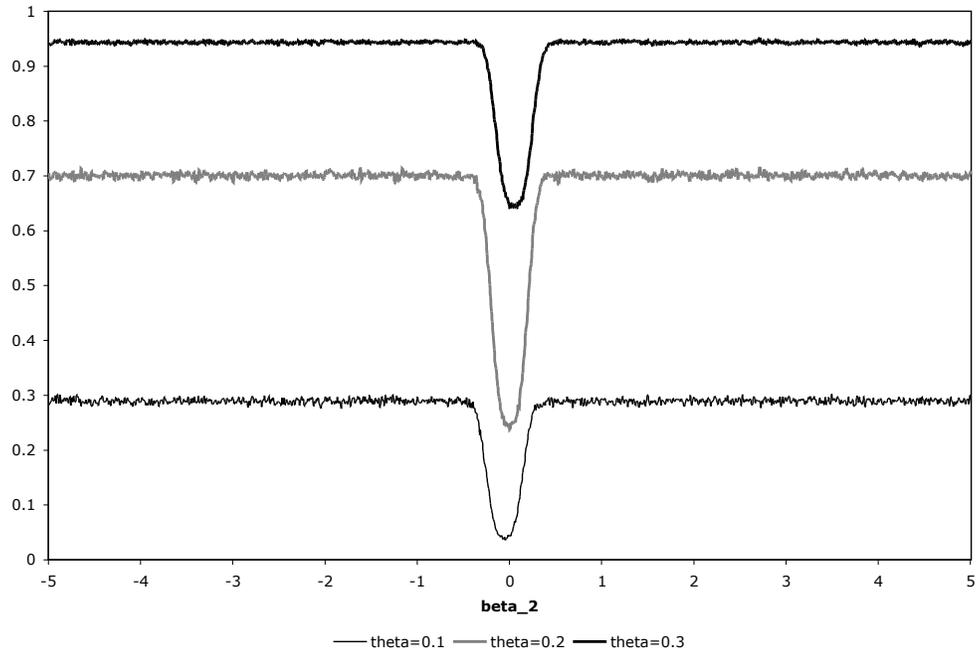Figure 8: Post-Consistent MS Power Using PI M-Bonf-Min CVs, $\gamma_{2,2} = 0.9$

Figure 9: Post-Consistent MS Power Using PI M-Bonf-Min CVs, $\gamma_{2,2} = 0.6$
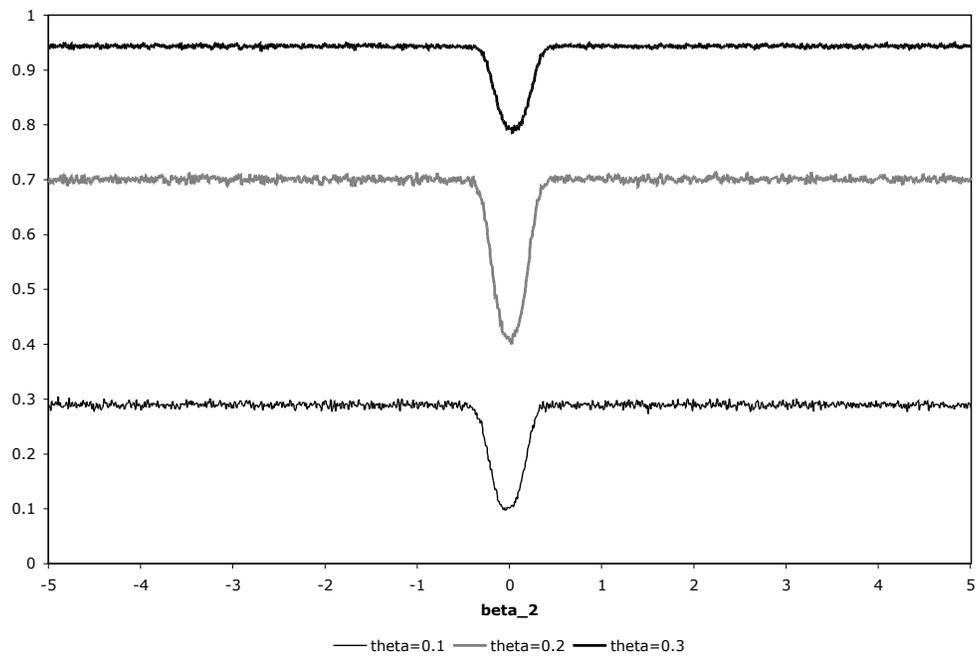

Figure 10: Post-Consistent MS Power Using PI M-Bonf-Min CVs, $\gamma_{2,2} = 0.3$
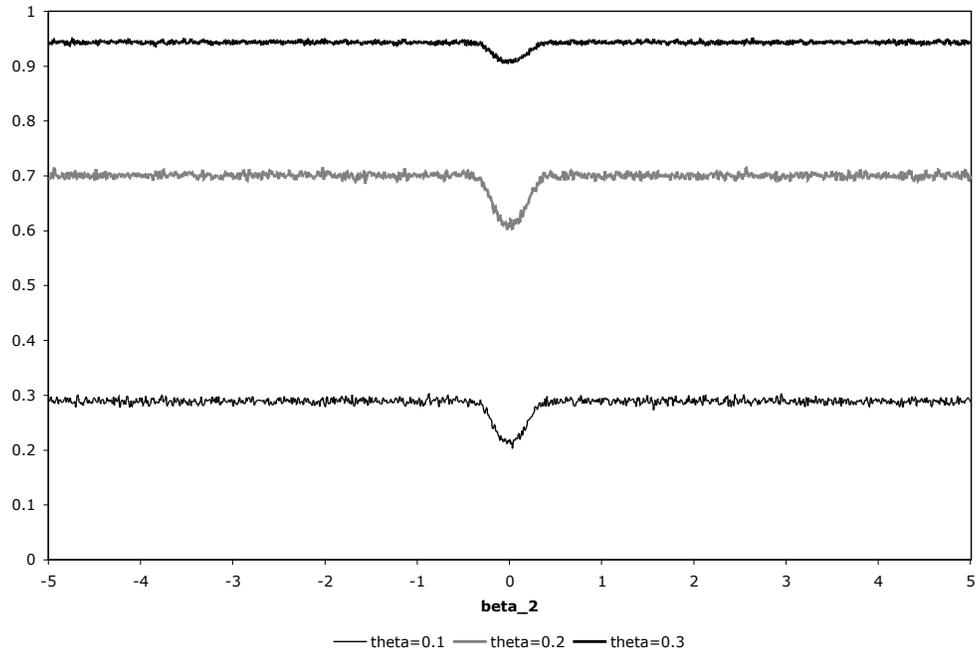
Figure 11: Post-Consistent MS Power Using PI M-Bonf-Min CVs, $\gamma_{2,2} = 0$