The Effects of Skewness and Kurtosis on Heteroskedasticity-Robust Bootstrap Methods in Finite Samples

by

Daniel Indacochea

An essay submitted to the Department of Economics
in partial fulfillment of the requirements for the
degree of Master of Arts

Queen's University
Kingston, Ontario, Canada
August 2012

Acknowledgements

I am grateful to James G. MacKinnon for his guidance and expert insight as both my professor and advisor, to Tian Xie, Mark Babcock and Jason Rhinelander for their technical assistance, and to my parents, Rafael and Olga, for their steadfast support. Any mistakes within are *mea culpa*.

Contents

1 Introduction

White (1980) developed the first **heteroskedasticity-consistent covariance matrix estimator** (HCCME) to deal with linear regression models with independent but heteroskedastic error terms. One of the major drawbacks of White's original estimator—now referred to as HC0 in the literature—is that t tests based on it will tend to overreject when the null is true in finite samples.

There have been two major approaches to resolving this problem. The first is to use better HCCMEs that inflate the residuals by adjusting for degrees of freedom or by adjusting for the leverage of the observations. Many authors have found, however, that these estimators do not always work particularly well in finite samples; see MacKinnon (2011) for a comprehensive discussion. The second approach is to use a heteroskedasticity-robust bootstrap method. MacKinnon (2011) advocates a wild bootstrap approach that works better in terms of size and power than using HCCMEs directly in finite samples. Following the experimental design of MacKinnon (2011), Hausman and Palmer (2012) find that the variance bootstrap used along with the second-order Edgeworth expansion of the distribution of the robust t statistic—which the authors call the **second-order bootstrap** (SOB)—works comparably well to the wild bootstrap in terms of size and better in terms of power.

It is important to note that the experimental design used by both of these papers employs normal error terms to generate the data used for the simulations. However, practitioners will often encounter **data-generating processes** (DGPs) with error terms that deviate greatly from normality. Such deviations might have a substantial impact on the findings of both MacKinnon (2011) and Hausman and Palmer (2012).

The former paper advocates a particular version of the wild bootstrap that takes the symmetry of the error terms for granted. The latter paper implicitly makes the much stronger assumption of normality in the error terms. The goal of this paper is therefore to test the sensitivity of the wild bootstrap variations and the SOB to deviations from normality in the error terms by varying the skewness and kurtosis of those error terms. To the best of my knowledge, this topic has not been previously studied.

This study finds that both methods perform poorly in the presence of severe kurtosis, but that the SOB is considerably more sensitive. These results suggest that attempts to account for one type of misspecification—namely, heteroskedasticity—by using additional information about the DGP may leave the estimator vulnerable to other forms of misspecification, of which kurtosis is but one example.

2 Background

2.1 Heteroskedasticity-Consistent Covariance Matrix Estimators

White (1980) developed the first HCCME to deal with linear regression models with independent but heteroskedastic error terms,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \mathbb{E}(\boldsymbol{u}) = \boldsymbol{0}, \quad \mathbb{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Omega}, \tag{1}$$

where $\boldsymbol{y}$ is an n-vector of observations with typical element y_i, $\boldsymbol{u}$ is an n-vector of error terms with typical element u_i, $\boldsymbol{X}$ is an $n \times k$ matrix of regressors, $\boldsymbol{\beta}$ is a k-vector of parameter coefficients, and $\boldsymbol{\Omega}$ is the $n \times n$ error covariance matrix with unknown diagonal elements and all off-diagonal elements equal to 0. The **ordinary least squares** (OLS) estimator and its corresponding variance are given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}, \qquad \mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{2}$$

In the case of homoskedasticity, $\mathbb{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Omega}$ would simply equal $\sigma^2 \boldsymbol{I}_{n\times n}$, where σ^2 denotes the error variance, and the second expression in (2) would reduce to $\sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$. We could then compute an unbiased and consistent estimate of σ^2 with an appropriate method-of-moments estimator (see Davidson and MacKinnon 2004).

In the case of heteroskedasticity, however, we do not have the luxury of estimating $\boldsymbol{\Omega}$ so straightforwardly. Since $\boldsymbol{\Omega}$ is an $n \times n$ matrix, its dimensions grow as the number of observations, n, grows. Thus, $\boldsymbol{\Omega}$ cannot be estimated consistently without making further assumptions about the DGP in (1).

This seeming impasse is where the ingenuity of White (1980) comes into play. White realised that although the $n \times n$ matrix $\boldsymbol{\Omega}$ cannot be estimated consistently, the $k \times k$ matrix $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\Omega}\boldsymbol{X}$ *can* be using what he termed an HCCME. White's proposed HCCME is now referred to as HC0 in the literature, as it inspired later generations of HCCMEs, collectively referred to here as the HCj series.

2.2 HCCMEs in Finite Samples

The HC0 estimator is important in part because it allows one to compute heteroskedasticity-robust t statistics of the form:

$$\hat{\tau} = \frac{\boldsymbol{c}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{c}^\top \boldsymbol{\beta}_0}{\sqrt{\boldsymbol{c}^\top [(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \hat{\boldsymbol{\Omega}} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}] \boldsymbol{c}}}, \tag{3}$$

where $\boldsymbol{c}^\top$ is a vector of weights used to compute linear combinations of the parameter vector $\boldsymbol{\beta}$. One of the major drawbacks of White's original HC0 estimator is that t tests based on it will tend to overreject when the null is true in finite samples. Cribari-Neto and Zarkos (2001) identify cases when the rejection frequency is over four times the nominal level of 0.05.

There have been two major approaches to correcting this problem. The first is to adjust the way $\boldsymbol{X}^\top \boldsymbol{\Omega} \boldsymbol{X}$—the "meat" of the sandwich matrix—is estimated by using the later editions of the HCj estimators outlined below.

1. HC0: White's (1980) formulation; the first of the HCCMEs.
   $\hat{\boldsymbol{\Omega}} = \mathrm{diag}\{\hat{u}_i^2\}$,
   where $\hat{u}_i \equiv y_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}$ is the residual element from estimating (1) by OLS.

2. HC1: MacKinnon and White (1985) realised OLS residuals are on average too small, and therefore compensated for this by adjusting for the degrees of freedom
   $\hat{\boldsymbol{\Omega}} = \frac{n}{n-k} \mathrm{diag}\{\hat{u}_i^2\}$,
   where n is the number of observations and k is the number of regressors in the matrix $\boldsymbol{X}$.

3. HC2: MacKinnon and White (1985) also proposed an estimator that could

compensate for small residuals by adjusting for the leverage
$\hat{\boldsymbol{\Omega}} = \text{diag}\left\{\frac{\hat{u}_i^2}{1-h_i}\right\}$,
where $h_i = \boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top$.

4. HCJ: MacKinnon and White (1985) introduced an HCCME with similar properties to a statistical technique called the **jackknife**. It is especially useful when large variances correspond to large h_i
$\hat{\boldsymbol{\Omega}} = \frac{n-1}{n}\left(\text{diag}\left\{\ddot{u}_i^2\right\} - \frac{1}{n}\ddot{\boldsymbol{u}}\ddot{\boldsymbol{u}}^\top\right)$,
where $\ddot{u}_i \equiv \hat{u}_i/(1-h_i)$.

5. HC3: Introduced by Davidson and MacKinnon (1993) as a very close approximation to the HCJ estimator
$\hat{\boldsymbol{\Omega}} = \text{diag}\left\{(\frac{\hat{u}_i}{1-h_i})^2\right\}$.

This brief list of HCj estimators is by no means exhaustive. Cribari-Neto (2004) proposes an HC4 estimator, and Cribari-Neto et al. (2007) propose an HC5 estimator, both of which are based on the idea of inflating $\hat{u}_i$ more (less) when h_i is large (small) relative to the average h_i, which is always k/n by construction. However, under MacKinnon's (2011) experimental design, tests based on HC4 are found to underreject the most severely. Since this paper follows a slightly modified version of MacKinnon's (2011) experimental design, and since HC4 and HC5 are very similar, neither of these estimators are discussed in this paper.

It is important to note that although the HCj estimators are improvements on White's (1980) HC0 estimator, tests based on these estimators do not always work particularly well in terms of size and power in finite samples. MacKinnon (2011) finds that tests based on HC1 overreject quite severely in finite samples, as do tests based on HC2, albeit less so, and tests based on HC3 or HCJ may overreject in some cases and underreject in others.

2.3 The Wild Bootstrap

The second approach to dealing with the poor finite-sample properties of HCCME-based tests is to bootstrap the t tests. One common way of doing this is to use the **wild bootstrap** introduced by Liu (1988). MacKinnon (2011) advocates a wild bootstrap variation that works better in terms of size and power than using the HCj estimators in finite samples. Let $\boldsymbol{x}_1$ denote the rightmost regressor of $\boldsymbol{X}$ in (1) so that $\boldsymbol{X} \equiv [\boldsymbol{X}_1 \; \boldsymbol{x}_1]$. Without loss of generality, to test the hypothesis that the coefficient of $\boldsymbol{x}_1$ is significantly different from zero (or to test any single linear coefficient restriction) using the wild bootstrap, one would proceed as follows:

1. Compute $\hat{\tau}$ as in (3) from the original sample $\boldsymbol{y}$ using either HC1, HC2 or HC3 to estimate $\boldsymbol{X}^\top \boldsymbol{\Omega} \boldsymbol{X}$. One could use HCJ as well, but results would resemble those of HC3.

2. Compute either the restricted OLS residuals, denoted $\tilde{u}_i$, or unrestricted OLS residuals, denoted $\hat{u}_i$, from (1).

3. Rescale the residuals by using any of the transformations from the HCj series. For example, we could rescale the restricted residuals as in HC3 to form a new vector of transformed residuals $f(\tilde{u}_i) = (\frac{\tilde{u}_i}{1-h_{1i}})$ for $i \in \{1, ..., n\}$, where $h_{1i} = \boldsymbol{X}_{1i}(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_{1i}^\top$. This transformation is referred to as w3 in MacKinnon (2011), while the transformations that correspond to HC1 and HC2 are referred to as w1 and w2, respectively.

4. For each of the B bootstrap samples, generate

$$y_i^* = \boldsymbol{X}_{1i}\tilde{\boldsymbol{\beta}} + f(\tilde{u}_i)v_i^*, \tag{4}$$

where $\tilde{\boldsymbol{\beta}}$ denotes the restricted parameter estimates, and v_i^* is a random variable with mean 0 and variance 1. If using unrestricted residuals $\hat{u}_i$, the B bootstrap samples are generated as

$$y_i^* = \boldsymbol{X}_i\hat{\boldsymbol{\beta}} + f(\hat{u}_i)v_i^*. \tag{4a}$$

MacKinnon (2011) looks at two choices for the distribution of v_i^*, both of which are two-point distributions. The first distribution, which MacKinnon (2011) refers to as F_1, was proposed by Mammen (1993):

$$F_1: \quad v_i^* = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } (\sqrt{5}+1)/(2\sqrt{5}), \\ (\sqrt{5}+1)/2 & \text{with probability } (\sqrt{5}-1)/(2\sqrt{5}). \end{cases}$$

One of the attractive features of this two-point distribution is that the skewness of the bootstrap error terms $f(\tilde{u}_i)v_i^*$ is the same as the skewness of the residuals. The second two-point distribution, which MacKinnon (2011) refers to as F_2, is

$$F_2: \quad v_i^* = \begin{cases} -1/2 & \text{with probability } 1/2, \\ 1/2 & \text{with probability } 1/2. \end{cases}$$

This distribution is called the **Rademacher distribution** and was proposed by Davidson and Flachaire (2008). An important feature of this distribution is that it imposes symmetry on the bootstrap error terms. For either distribution, since there are only n residuals, and each bootstrap error term can only take on two values, there exists 2^n possible bootstrap error vectors.

5. If using the restricted DGP in (4), then for each of the bootstrap samples $\boldsymbol{y}_j^*$, a statistic τ_j^* is computed as

$$\tau_j^* = \frac{\boldsymbol{c}^\top \hat{\boldsymbol{\beta}}_j^* - \boldsymbol{c}^\top \boldsymbol{\beta}_0}{\sqrt{\boldsymbol{c}^\top [(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}_j^* \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}] \boldsymbol{c}}}, \tag{5}$$

where $\hat{\boldsymbol{\beta}}_j^* \equiv (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}_j^*$, and $\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}_j^* \boldsymbol{X}$ is estimated with the same HCCME used in step 1, but using unrestricted bootstrap residuals $\hat{u}_{ij}^* \equiv y_{ij}^* - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_j^*$. If using the unrestricted DGP in (4a), then τ_j^* is instead computed as

$$\tau_j^* = \frac{\boldsymbol{c}^\top \hat{\boldsymbol{\beta}}_j^* - \boldsymbol{c}^\top \hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{c}^\top [(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}_j^* \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}] \boldsymbol{c}}}. \tag{5a}$$

The key difference between (5) and (5a) is that $\hat{\boldsymbol{\beta}}$ replaces $\boldsymbol{\beta}_0$ in the numerator of (5) in order to impose the null.

6. The **equal-tail bootstrap $\boldsymbol{P}$ value** is then computed as

$$\hat{p}^*(\hat{\tau}) = 2 \cdot \min \left(\frac{1}{B} \sum_{j=1}^{B} I(\tau_j^* \leq \hat{\tau}), \frac{1}{B} \sum_{j=1}^{B} I(\tau_j^* > \hat{\tau}) \right), \tag{6}$$

which is then compared to the nominal significance level of 0.05 for hypothesis testing. Here $I(\cdot)$ denotes the indicator function.

In sum, there are two ways of computing the residuals and fitted values (restricted or unrestricted, denoted r and u), three ways of transforming the residuals (w1, w2 and w3) and two distributions from which to draw the v_i^* (F_1 and F_2, or simply "1" and “2”), giving a total of twelve variations of the wild bootstrap which MacKinnon (2011) considers. Each of these variations can be used along with any of the HCj estimators. MacKinnon (2011) finds that the w3r2 variation (*i.e.* w3 transformation, restricted residuals, F_2 Rademacher distribution) used along with the HC1 estimator works best in the following Monte Carlo experiment:

$$y_i = \beta_1 + \sum_{k=2}^{5} \beta_k X_{ik} + u_i, \qquad u_i = \sigma_i \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0,1), \tag{7}$$

where $\beta_k = 1$ for $k \leq 4$, $\beta_k = 0$ for $k = 5$, and

$$\sigma_i = z(\gamma)(\beta_1 + \sum_{k=2}^{5} \beta_k X_{ik})^{\gamma}. \tag{8}$$

All the X_{ik} regressors are drawn from a lognormal distribution. The $z(\gamma)$ term is a scaling factor that mechanically satisfies the requirement that the average variance of u_i equals 1. One can vary the severity of the heteroskedasticity by varying $0 \leq \gamma \leq 2$, where $\gamma = 0$ corresponds to homoskedasticity and $\gamma = 2$ corresponds to severe heteroskedasticity.

2.4 The Second-Order Bootstrap

Following the experimental design of MacKinnon (2011), Hausman and Palmer (2012) find that the **variance bootstrap** (see Equation (11) below) used along with the second-order Edgeworth expansion of the distribution of the robust t statistic—which the authors call the **second-order bootstrap** (SOB)—works comparably well to the wild bootstrap in terms of size and better in terms of power. The SOB is based on Rothenberg’s (1988) derivation of the Edgeworth expansion for distributions of test statistics of the form (3). The SOB method scales the traditional normal critical values $z_{\alpha/2}$ as follows:

$$t = \pm z_{\alpha/2}\left(1 - \frac{1}{12}\left(1 + z_{\alpha/2}^2\right)V + \frac{a\left(z_{\alpha/2}^2 - 1\right) + b}{2n}\right), \tag{9}$$

where n is the sample size and the remaining terms are defined as

$$\begin{aligned}
V &= \frac{\sum_{i=1}^n f_i^4 \hat{u}_i^4}{(\sum_{i=1}^n f_i^2 \hat{u}_i^2)^2},\\
a &= \frac{\sum_{i=1}^n f_i^2 g_i^2}{\sum_{i=1}^n f_i^2 \hat{u}_i^2},\\
b &= \frac{\sum_{i=1}^n f_i^2 Q_{ii}}{\sum_{i=1}^n f_i^2 \hat{u}_i^2},\\
\boldsymbol{f} &= n\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{c},\\
\boldsymbol{g} &= \frac{(\boldsymbol{I}_{n\times n} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\Omega}\boldsymbol{f}}{\sqrt{\boldsymbol{f}^\top \boldsymbol{\Omega}\boldsymbol{f}/n}},\\
\boldsymbol{Q} &= n\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\Omega}(\boldsymbol{P}_{\boldsymbol{X}} - 2\boldsymbol{I}_{n\times n}),
\end{aligned}$$

and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is a projection matrix that projects orthogonally onto the column span of the matrix $\boldsymbol{X}$, $\boldsymbol{\Omega}$ is estimated with HC0 and, as in (3), $\boldsymbol{c}^\top$ is a vector of weights of the parameter vector $\boldsymbol{\beta}$. An important feature of the SOB approach is that instead of computing test statistics in (3) using an HCCME, Hausman and Palmer (2012) compute the test statistic using the **bootstrap covariance matrix**, which is defined as:

$$\widehat{\text{Var}}^*(\hat{\boldsymbol{\beta}}) = \frac{1}{B-1}\sum_{j=1}^{B}(\hat{\boldsymbol{\beta}}_j^* - \bar{\boldsymbol{\beta}}^*)(\hat{\boldsymbol{\beta}}_j^* - \bar{\boldsymbol{\beta}}^*)^\top, \tag{10}$$

where $\hat{\boldsymbol{\beta}}_j^*$ is computed for each of the B bootstrap samples, and $\bar{\boldsymbol{\beta}}^*$ is their mean. Each of the B bootstrap sample pairs, $[\boldsymbol{y}_j^* \ \boldsymbol{X}_j^*]$, is generated via the **pairs bootstrap** by making n draws with replacement from the rows of the matrix $[\boldsymbol{y} \ \boldsymbol{X}]$. In contrast to the $\hat{\boldsymbol{\beta}}_j^*$ in (5) and (5a), which condition on $\boldsymbol{X}$, here we have $\hat{\boldsymbol{\beta}}_j^* \equiv (\boldsymbol{X}_j^{*\top}\boldsymbol{X}_j^*)^{-1}\boldsymbol{X}_j^{*\top}\boldsymbol{y}_j^*$. The estimated bootstrap covariance matrix in (10) is then used to compute test statistics identical to (3), except that the HCCME is replaced with $\widehat{\mathrm{Var}}^*(\hat{\boldsymbol{\beta}})$

$$\hat{\tau} = \frac{\boldsymbol{c}^\top\hat{\boldsymbol{\beta}} - \boldsymbol{c}^\top\boldsymbol{\beta}_0}{\sqrt{\boldsymbol{c}^\top\big[\widehat{\mathrm{Var}}^*(\hat{\boldsymbol{\beta}})\big]\boldsymbol{c}}}. \tag{11}$$

For hypothesis testing, then, the SOB proposed by Hausman and Palmer (2012) simply compares the test statistic in (11) with the scaled critical values in (9).

2.5 Assumptions in Bootstrap Methods

It is important to note that the experimental design used by both of these papers employs normal error terms to generate the data used for the simulations. However, practitioners will often encounter DGPs with error terms that deviate greatly from normality. Such deviations might have a substantial impact on the findings of both MacKinnon (2011) and Hausman and Palmer (2012). The former paper advocates a particular version of the wild bootstrap that takes the symmetry of the error terms for granted. The latter, which makes the much stronger assumption of normality in the error terms, relies on an Edgeworth expansion of the distribution of test statistics of the form (3), first worked out in Rothenberg (1988). As Rothenberg explains, "In the present paper we examine the properties of a family of tests in the *normal* [emphasis added] regression model when the error covariance matrix is unknown". The goal of this paper is therefore to test the sensitivity of the wild bootstrap variations and the SOB to deviations from normality in the error terms by varying the skewness and kurtosis of those error terms. To the best of my knowledge, such analysis has not been previously included in studies on this topic.

3 Experimental Design

This study extends the experimental design developed in MacKinnon (2011) and used by Hausman and Palmer (2012) by examining the size and power performance of the wild bootstrap variations as well as the SOB when the error terms deviate from normality. The key difference is that here the ϵ_i terms in (7) are generated from a **skew-normal distribution**:

$$y_i = \beta_1 + \sum_{k=2}^{5} \beta_k X_{ik} + u_i, \qquad u_i = \sigma_i \epsilon_i, \qquad \epsilon_i \sim \mathcal{SN}(0, 1, \alpha), \tag{12}$$

where $\beta_k = 1$ for $k \leq 4$, $\beta_k = 0$ for $k = 5$, and

$$\sigma_i = z(\gamma)(\beta_1 + \sum_{k=2}^{5} \beta_k X_{ik})^{\gamma}. \tag{13}$$

The skew-normal distribution was first described by Azzalini (1985), and the mathematical details have been relegated to Appendix I. Notice the presence of the extra α argument in the distribution from (12). This α is a **shape parameter** whereby one can vary the skewness of the distribution. The role of the shape parameter is immediately apparent from Figure 1. Evidently, the skewness of the distribution is fairly extreme for $|\alpha| \geq 5$. Thus, in this study the values of $\alpha \in \{-5, -3, -1, 0, 1, 3, 5\}$ are used, where $\alpha = 0$ corresponds to the normal distribution used in MacKinnon (2011) and Hausman and Palmer (2012).

For each replication, the statistic in (3) is calculated if using the wild bootstrap, or the statistic in (11) if using the SOB. Since MacKinnon (2011) finds the wild bootstrap to work best with HC1, our attention is limited to this pairing when testing the wild bootstrap. Both MacKinnon (2011) and Hausman and Palmer (2012) use $k = 5$ regressors, including a constant. We thus have a parameter vector $\boldsymbol{\beta}^\top = [\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5]$. To test H_0: $\beta_5 = 0$, then, one must use the vector $\boldsymbol{c}^\top = [0\ 0\ 0\ 0\ 1]$.

In addition to studying the effects of skewness on the size and power of the bootstrap methods, this paper studies the effect of kurtosis. Arellano-Valle and Azzalini (2011) describe a **skew-*t* distribution** that allows one to vary the kurtosis of the error terms. The Monte Carlo experimental design is analogous to the skew-normal:

$$y_i = \beta_1 + \sum_{k=2}^{5} \beta_k X_{ik} + u_i, \qquad u_i = \sigma_i \epsilon_i, \qquad \epsilon_i \sim \mathcal{ST}(0, 1, \alpha, \nu), \tag{14}$$

where $\beta_k = 1$ for $k \leq 4$, $\beta_k = 0$ for $k = 5$, and

$$\sigma_i = z(\gamma)(\beta_1 + \sum_{k=2}^{5} \beta_k X_{ik})^\gamma. \tag{15}$$

Notice the presence of the extra ν argument in the skew-t distribution from (14). This ν is the degrees of freedom of the skew-t distribution whereby one can vary the kurtosis of the error terms. The mathematical details of the skew-t distribution are relegated to Appendix II. That said, the role of the parameter ν is clear from Figure

2. As one might expect, the skew-t converges to the skew-normal as $\nu \to \infty$. Again, for each replication the test statistic in (3) or (11) is computed depending on the bootstrap method used.

4 Results

4.1 Size with Normal Errors

To illustrate the relative size performance of the different heteroskedasticity-robust bootstrap methods when the error terms are normal, Figure 3 presents rejection frequencies for eight of the wild bootstrap variations based on HC1 and for the SOB when testing H_0: $\beta_5 = 0$. The four w1 transformations are omitted because MacKinnon (2011) finds them to usually perform worse than w2 and w3. For this and the remaining size figures, the rejection frequencies are plotted against $\gamma \in \{0.0, 0.1, ..., 2.0\}$, with M = 400,000 Monte Carlo replications, B = 399 bootstrap samples, and a sample size of n = 40. In terms of size, we see that with normal errors the w3r2 variation and the SOB work comparably well to each other and outperform the remaining bootstrap variations. All of the unrestricted variations overreject at $\gamma = 0$ and decline monotonically as γ increases. The w2r1 and w3r1 variations—which make use of the Mammen distribution—consistently overreject for all values of γ.

4.2 Size with Skew-Normal Errors

Having identified w3r2 and SOB as the best bootstrap methods in terms of size, the effect of skewness on rejection frequencies for these methods is examined in Figures 4 and 5, respectively. Perhaps surprisingly, even with the extreme skewness exhibited when $|\alpha| = 5$, the rejection frequencies are fairly steady. It does appear in Figure 5, however, that $|\alpha| = 5$ has the most deleterious effect on rejection frequencies.

4.3 Size with Skew-t Errors

To study the effects of kurtosis, $\alpha = -5$ is chosen along with values of $\nu \in \{\infty,$ 8, 5, 3, 2$\}$, where $\nu = \infty$ corresponds to the skew-normal. As with the standard t distribution, the skew-t has no moments for $\nu = 1$, a mean but no higher moments for $\nu = 2$, and so forth. We now see in Figures 6 and 7 that increasing kurtosis increases the downward bias of t tests based on either w3r2 or SOB. At $\nu = 2$ the SOB underrejects more severely than does w3r2 for low to mid values of γ, and less severely for high values of γ. Figure 8 presents the results for w3r1, which was found to perform the least poorly out of the Mammen variations of the wild bootstrap in terms of size. Here it is apparent that kurtosis increases the upward bias of tests based on w3r1. Interestingly, for $\gamma >> 0$ these tests appear markedly less sensitive to kurtosis than the two leading bootstrap candidates, w3r2 and SOB.

4.4 Power with Normal Errors

All of the remaining figures deal with statistical power and were generated using $M = 50,000$ replications, $B = 399$ bootstraps, and $n = 40$ observations. Using normal error terms, Figure 9 compares the power functions of w3r2 and SOB for H_0: $\beta_5 = 0$ against the 71 true values of $\beta_5 \in \{-0.70, -0.68, ..., 0.70\}$. Here we see the key result in Hausman and Palmer (2012) that the SOB outperforms w3r2 in terms of power against true values of $\beta_5 \in \{-0.70, ..., -0.14\} \cup \{0.04, ..., 0.70\}$—or about 89% of the domain of β_5. Moreover, SOB only performs modestly less well than w3r2 against the true values of $\beta_5 \in \{-0.12, ..., 0.02\}$. One of the notable features of Figure 9 is that the power function for w3r2—and to a lesser extent SOB—are asymmetrical, exhibiting more power against $\beta_5 < 0$ than against $\beta_5 > 0$. This is due to the way heteroskedasticity is generated in the experimental setup. Recall from (8) that σ_i is dependent on a weighted sum of the columns of $\boldsymbol{X}$. For $\gamma > 0$, this sum is typically greater for $\beta_5 > 0$ than for $\beta_5 < 0$. As such, there is more heteroskedasticity when β_5 is positive than when β_5 is negative, which explains the asymmetry observed in the power functions.

4.5 Power with Skew-Normal Errors

To examine the effects of skewness on power, the errors in Figure 10 are generated from the skew-normal distribution and a zoomed-in view is provided in Figure 11. Here we see that power actually *increases* for $\alpha > 0$. These results make sense when we again consider how the data are generated. With a skew corresponding to $\alpha > 0$, the error terms are drawn from a distribution where the mode is less than the mean, as is clear in Figure 1. The error terms therefore tend to take on negative values, which decreases the heteroskedasticity produced in equation (13) and in turn yields greater power. As α becomes negative, the errors become increasingly positive, which increases the heteroskedasticity in (13) and in turn reduces power. Figure 12 shows the effect of skewness on the power function of the SOB, which is noticeably less sensitive than w3r2. Thus far we see that skewness alone does not have a very sizable effect on power in our experimental design, which one might have expected based on the near-negligible effect skewness was shown to have on size in Figures 4 and 5.

4.6 Power with Skew-t Errors

Figures 13 – 15 present the effects of kurtosis on the power functions of the different bootstrap methods. The kurtosis is defined by $\nu \in \{\infty, 8, 5, 3, 2\}$ and skewness is held constant at $\alpha = 0$. Now we see a reduction in power for all the bootstrap tests as kurtosis increases. For comparison, Figure 16 plots w3r1, w3r2 and SOB on the same graph for $\nu = 2$. One of the striking features of Figure 16 is that the power function for the SOB can no longer be said to outperform that of w3r2. SOB performs better in the domain $\beta_5 \in \{-0.70, ..., -0.46\} \cup \{0.26, ..., 0.70\}$—or about 51% of the domain of β_5—which suggests that it is more sensitive to kurtosis than w3r2.

It is difficult to make meaningful comparisons between w3r1 and the other two bootstrap methods in Figure 16 since all have disparate and distorted sizes. The former always overrejects against $\beta_5 = 0$ and the latter two slightly underreject. Davidson and MacKinnon (2006) examine ways to compare power functions with different sizes, but one of their findings is that it is difficult to do so in an unambiguous way.

Up to this point we have looked at the effects of kurtosis on power in isolation, but we are also interested in the effects of kurtosis coupled with skewness. Figures 17 – 19 allow us to examine these combined effects by fixing the kurtosis to an extreme level of $\nu = 2$ and varying the shape parameter α. Recall from Figures 10 – 12 that when we introduced skewness into the error terms but not kurtosis, the power actually *increased* for $\alpha > 0$. In Figure 17, by contrast, we now see a substantial reduction in power for the SOB as the skewness increases in *either* direction. It seems that with sufficiently severe kurtosis, greater skewness leads to a greater reduction in power.

However, as one might expect based on the asymmetry of the experimental design in (15), more power is lost when the skew is negative. To avoid clutter, Figures 18 and 19 only show results for $\alpha < 0$. One unexpected feature of Figure 19 is that the reduction in power for w3r1 appears to be just as severe as with w3r2 in Figure 18. This is surprising since the Mammen two-point distribution in w3r1 is designed to deal precisely with the kind of asymmetry introduced here. This result echoes that of Davidson and Flachaire (2008), who find that when the error terms are asymmetrically distributed, Mammen-based wild bootstrap variations offer no advantage in terms of size over bootstrap methods that do not take explicit account of skewness.

Figure 20 compares the power functions of the main bootstrap methods discussed under the most stressed conditions considered in this paper: $\nu = 2$, $\alpha = -5$ and $\gamma = 1$. Remarkably, the SOB actually performs *worse* than w3r2 in most of the domain of β_5. In particular, SOB only outperforms w3r2 for $\beta_5 \in \{0.36, ..., 0.70\}$—or 25% of the domain of β_5—and underperforms everywhere else. This is a dramatic drop from the case of normal errors in Figure 9, where SOB was found to outperform w3r2 in 89% of the domain of β_5. From these results, it appears that the more the error terms deviate from normality, the worse SOB performs relative to w3r2. These findings make sense considering that the Edgeworth expansion on which the SOB is based is derived in Rothenberg (1988) under the assumption that the error terms are normal.

Figure 20 also features two of the "runner-up" bootstrap variations, w2r1 and w2r2. Just as with normal errors, w3r2 outperforms w2r2 when the errors are skew-t distributed. There also appears to be little to choose from between the two Mammen distributions, w2r1 and w3r1, and surprisingly neither seems to handle skewness any better than the bootstrap variations based on the symmetric Rademacher distribution.

5 Conclusion

White's (1980) HCCME filled a substantial gap in the econometric literature as for the first time practitioners could make asymptotically valid inferences for linear regression models with independent but heteroskedastic error terms. The main drawback of White's proposed HCCME is that t tests based on it tend to over-reject in finite samples since the corresponding residuals are typically too small. There have been two major approaches to correcting this problem. The first is to use better HCCMEs that inflate the residuals by adjusting for degrees of freedom or leverage. There is a large body of evidence, however, that suggests a better approach is to use one of several heteroskedasticity-robust bootstrap methods, of which the Rademacher-based wild bootstrap and the SOB are the two leading candidates. These methods come with a set of assumptions. The former takes the symmetry of the distribution of the error terms for granted; the latter implicitly makes the much stronger assumption of normality in the error terms.

In the presence of normal error terms, this study shows that the SOB performs comparably well to the wild bootstrap in terms of size and considerably better in terms of power. However, it seems that this superior performance may be due to making use of information about the DGP that practitioners seldom have in practice. Indeed, this study finds that the more the distribution of the error terms deviates from normality, the less well the SOB performs relative to the wild bootstrap. Moreover, this paper identifies cases where the SOB even performs worse. It therefore appears that attempts to correct for one type of misspecification—namely, heteroskedasticity—by using *a priori* information about the DGP can leave the estimator vulnerable to other forms of misspecification.

This study also shows that the choice of two-point distribution used in the wild bootstrap DGP has an enormous effect on both size and power. Quite unexpectedly, tests based on the Mammen distribution—which take explicit account of skewness—actually perform substantially worse than those based on the symmetric Rademacher distribution, even in the presence of severe skewness and kurtosis. These results corroborate earlier findings that the Mammen distribution has little to recommend it, and that the Rademacher-based variation of the wild bootstrap is to be preferred in practice.

6 References

Arellano-Valle, R. and Azzalini, A. (2011). "The centred parameterization and related quantities of the skew-t distribution". *Journal of Multivariate Analysis*, doi: 10.1016/j.jmva.2011.05.016.

Azzalini, A. (1985). "A class of distributions which includes the normal ones". *Scandanavian Journal of Statistics*, 12:171–178.

Cribari-Neto, F. (2004). "Asymptotic inference under heteroskedasticity of unknown form". *Computational Statistics and Data Analysis*, 45:215–233.

Cribari-Neto, F., Souza, T. C., and Vasconcellos, K. L. P. (2007). "Asymptotic inference under heteroskedasticity of unknown form". *Communications in Statistics: Theory and Methods*, 37:3329–3330.

Cribari-Neto, F. and Zarkos, S. (2001). "Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap". *Journal of Statistical Computation and Simulation*, 68:391–411.

Davidson, R. and Flachaire, E. (2008). "The wild bootstrap, tamed at last". *Journal of Econometrics*, 146:162–169.

Davidson, R. and MacKinnon, J. (1993). *Estimation and Inference in Econometrics.* Oxford University Press, New York.

Davidson, R. and MacKinnon, J. (2004). *Econometric Theory and Methods.* Oxford University Press, New York.

Davidson, R. and MacKinnon, J. (2006). “The power of bootstrap and asymptotic tests”. *Journal of Econometrics*, 133:441–444.

Hausman, J. and Palmer, C. (2012). “Heteroskedasticity-robust inference in finite samples”. *Economics Letters*, 116:232–235.

Liu, R. (1988). “Bootstrap procedures under some non-I.I.D. models”. *Annals of Statistics*, 16:1696–1708.

MacKinnon, J. (2011). “Thirty years of heteroskedasticity-robust inference”. *Queen's Economics Department Working Paper*, No. 1268.

MacKinnon, J. and White, H. (1985). “Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties”. *Journal of Econometrics*, 29:305–325.

Mammen, E. (1993). “Bootstrap and wild bootstrap for high dimensional linear models”. *Annals of Statistics*, 21:255–285.

Rothenberg, T. (1988). “Approximate power functions for some robust tests of regression coefficients”. *Econometrica*, 56:997–1019.

White, H. (1980). “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity”. *Econometrica*, 48:817–838.

7 Appendix I: The Skew-Normal Distribution

7.1 Mathematical Properties of the Skew-Normal

The setup for the skew-normal distribution is as follows:

$$x_i \sim f(x), \qquad f(x) = 2\phi(x)\Phi(\alpha x), \tag{A1}$$

where α is the **shape parameter** that allows one to vary the skewness of the distribution, and

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad \Phi(\alpha x) = \int_{-\infty}^{\alpha x} \phi(t)\, dt. \tag{A2}$$

The density $f(x)$ exhibits various interesting properties, described by Azzalini (1985):

1. when $\alpha = 0$, the skewness vanishes and we are left with a normal distribution
2. skewness increases as α increases in magnitude
3. as $\alpha \to \infty$, $f(x)$ converges to a half-normal density
4. as the sign of α changes, the skewness is reflected on the opposite side of the vertical axis

We can construct ϵ_i from x_i as follows:

$$\epsilon_i = \xi + \omega x_i, \tag{A3}$$

which is said to be skew-normally distributed with **location parameter** ξ and **scale parameter** ω^2. It is important to note that ξ and ω^2 are *not* the moments of this distribution. In fact, for the first four non-central moments we have

$$\mathbb{E}(\epsilon_i) = \xi + \omega\sqrt{\frac{2}{\pi}}\delta, \tag{A4}$$

$$\mathrm{Var}(\epsilon_i) = \omega^2(1 - 2\delta^2/\pi), \tag{A5}$$

$$\gamma_1 = \frac{4-\pi}{2}\frac{(\sqrt{2/\pi}\delta)^3}{(1-2\delta^2/\pi)^{3/2}}, \tag{A6}$$

$$\gamma_2 = 2(\pi - 3)\frac{(\sqrt{2/\pi}\delta)^4}{(1-2\delta^2/\pi)^2}, \tag{A7}$$

where δ is defined as

$$\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

In order to ensure $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = 1$, the random variables are standardised as follows:

$$\omega = \frac{1}{\sqrt{1-(2\delta^2/\pi)}}; \qquad \xi = -\omega\sqrt{\frac{2}{\pi}}\delta.$$

Now we indeed have $\epsilon_i \sim \mathcal{SN}(0, 1, \alpha)$.

7.2 Generating Skew-Normal Random Variables

To actually generate skew-normal random variables with a pseudo-random number generator, the algorithm used in this study is similar to an acceptance-rejection algorithm. However, one need not draw new random variables when the algorithm rejects. First, draw two independent random variables from the standard normal distribution.

$$z_{i1} \sim \mathcal{N}(0,1), \qquad z_{i2} \sim \mathcal{N}(0,1).$$

Second, produce x_i as follows:

$$x_i = \begin{cases} z_{i1} & if\ z_{i2} \leq \alpha z_{i1}, \\ -z_{i1} & if\ z_{i2} > \alpha z_{i1}, \end{cases}$$

where α is the shape parameter. The resulting x_i is distributed as in (A1).

8 Appendix II: The Skew-t Distribution

8.1 Mathematical Properties of the Skew-t

Azzalini (2011) also describes a skew-t distribution that allows for kurtosis. The setup is given below:

$$x_i \sim f(x), \qquad f(x) = \frac{2}{\omega} t(x;\nu) T\left(\alpha x \sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1\right), \tag{A8}$$

where α is the shape parameter, ν is the degrees of freedom that allows one to vary kurtosis, and

$$t(x;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad T(\alpha x;\nu) = \int_{-\infty}^{\alpha x} t(s;\nu)\, ds. \tag{A9}$$

Analogous to the skew-normal distribution, we can construct ϵ_i with the following transformation:

$$\epsilon_i = \xi + \omega x_i, \tag{A10}$$

which is said to follow a skew-t distribution with location parameter ξ and scale parameter ω^2. The moments of the skew-t distribution are generalisations of those of the skew-normal. In order to describe these moments, it is easiest to first state the **cumulants** for x_i with location parameter $\xi = 0$ and scale parameter $\omega^2 = 1$, and then derive the non-central moments of ϵ_i from those cumulants (see Arellano-Valle and Azzalini 2011):

$$\begin{aligned}
\mu_1(\delta,\nu) &= b_\nu\delta,\\
\mu_2(\delta,\nu) &= \frac{\nu}{\nu-2} - b_\nu^2\delta^2,\\
\mu_3(\delta,\nu) &= \mathbb{E}\left([u_i-\mu_1(\delta,\nu)]^3\right),\\
&= b_\nu\delta\left[\frac{3\nu}{(\nu-3)(\nu-2)} - \delta^2\Big(\frac{\nu}{\nu-3} - 2b_\nu^2\Big)\right],\\
\mu_4(\delta,\nu) &= \mathbb{E}\left([u_i-\mu_1(\delta,\nu)]^4\right),\\
&= \frac{3\nu^2}{(\nu-2)(\nu-4)} - 6\delta^2 b_\nu^2\frac{\nu(\nu-1)}{(\nu-2)(\nu-3)} + \delta^4 b_\nu^2\Big(\frac{4\nu}{\nu-3} - 3b_\nu^2\Big),
\end{aligned}$$

where ν must be larger than the cumulant $\mu_i(\delta,\nu)$ considered, and

$$b_\nu = \left(\frac{\nu}{\pi}\right)^{1/2}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, \qquad (\nu > 1).$$

Now that we have the cumulants for x_i, we can state the moments for $\epsilon_i = \xi + \omega x_i$:

$$\mathbb{E}\left(\epsilon_i\right) = \xi + \omega\mu_1(\delta,\nu), \tag{A11}$$

$$\mathrm{Var}\left(\epsilon_i\right) = \omega^2\mu_2(\delta,\nu), \tag{A12}$$

$$\gamma_1(\delta,\nu) = \frac{\mu_3(\delta,\nu)}{\mu_2(\delta,\nu)^{3/2}}, \tag{A13}$$

$$\gamma_2(\delta,\nu) = \frac{\mu_4(\delta,\nu)}{\mu_2(\delta,\nu)^2} - 3. \tag{A14}$$

The first two moments can be used to centre and standardise the skew-t distribution.

8.2 Generating Skew-t Random Variables

To generate a skew-t random variable, one could proceed by first generating a skew-normal random variable x_i in a similar manner to that described in Appendix I, but with location parameter $\xi = 0$ and the scale parameter $\omega^2 = 1$. Next, one would draw $q_i \sim \chi^2(\nu)$, where ν is the degrees of freedom. Most modern statistical software packages have built-in pseudo-random number generators capable of simulating the chi-square distribution. One then simply computes ϵ_i as

$$\epsilon_i = \xi' + \omega' \frac{x_i}{\sqrt{q_i/\nu}},$$

where ξ' and ω' are chosen in such a way that (A11) equals 0 and (A12) equals 1. We now indeed have $\epsilon_i \sim \mathcal{ST}(0, 1, \alpha, \nu)$.

9 Figures

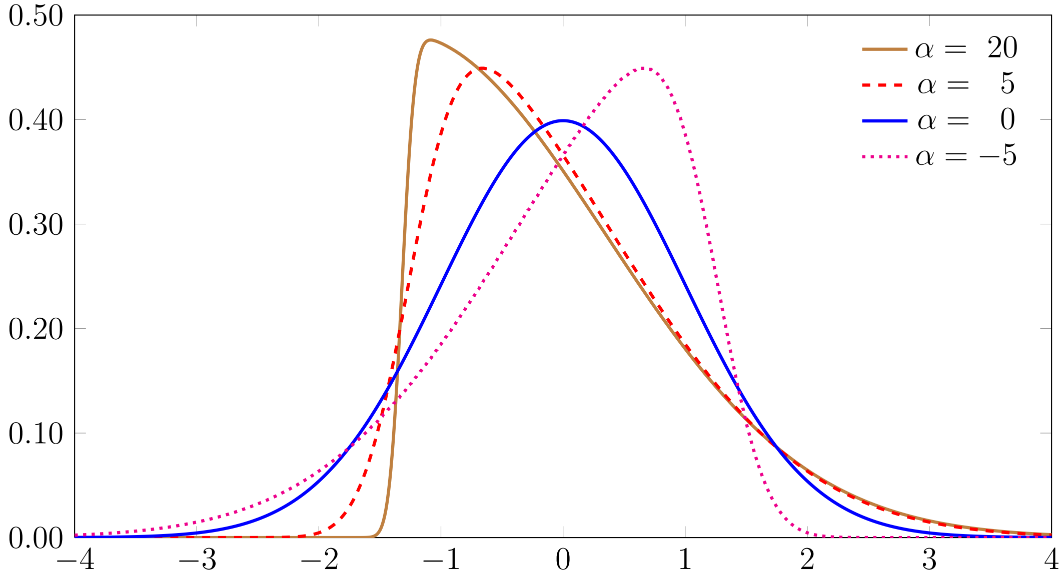

Figure 1: Skew-normal distributions for various values of shape parameter α

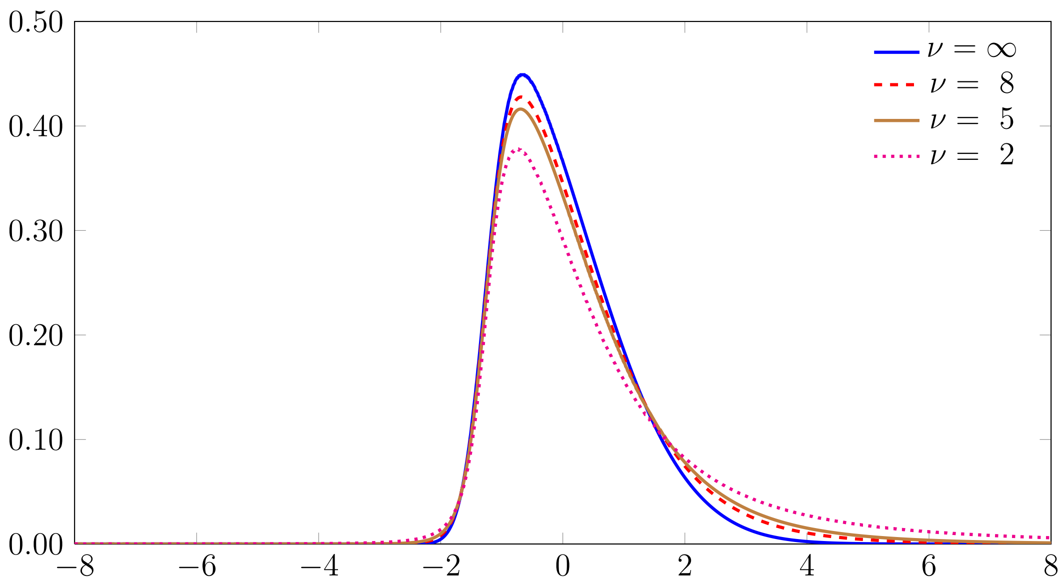

Figure 2: Skew-t distributions for shape parameter $\alpha = 5$ and various degrees of freedom ν

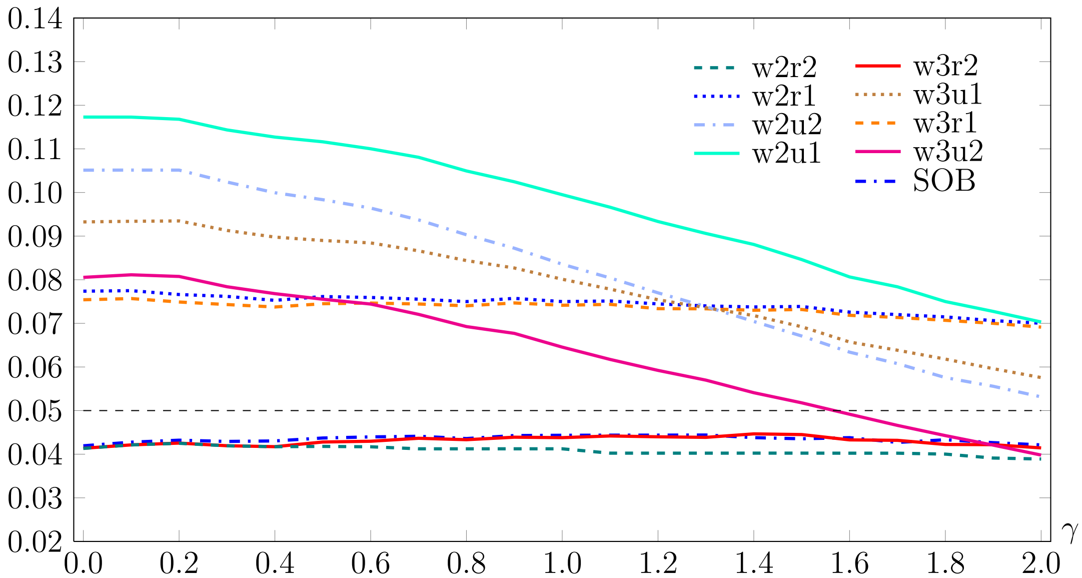

Figure 3: Rejection frequencies for SOB–VB and various WB-HC1 t tests with normal errors ($M = 400,000; B = 399; n = 40$)

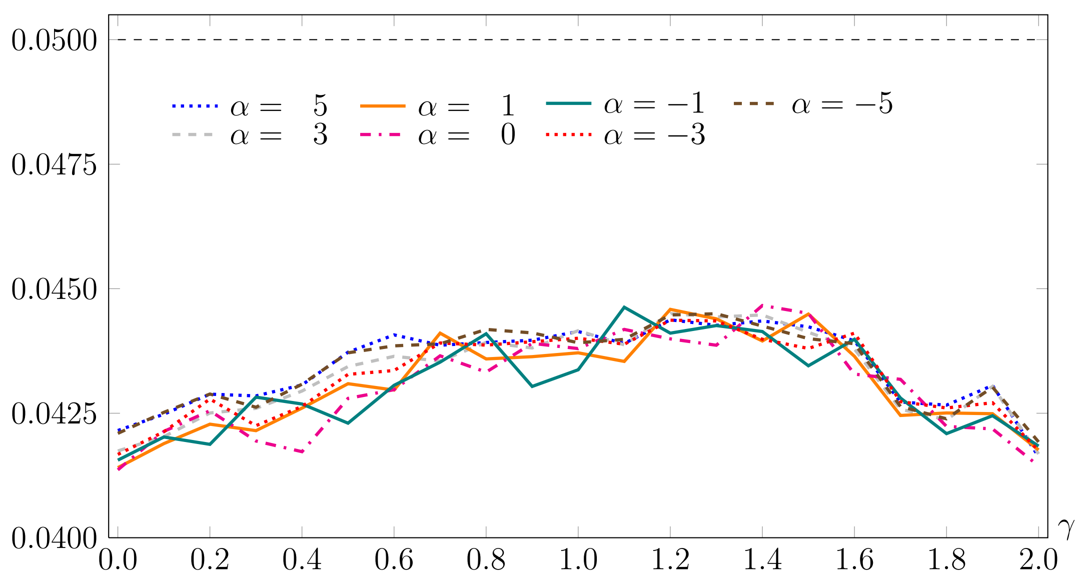

Figure 4: Rejection frequencies for w3r2–HC1 bootstrap t tests for various values of α ($M = 400,000; B = 399; n = 40$)

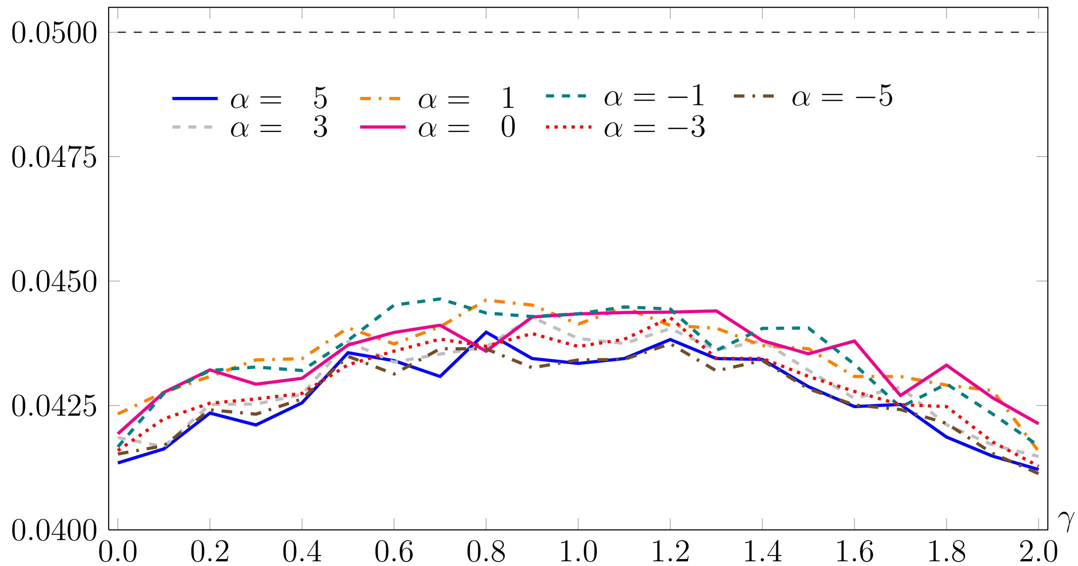

Figure 5: Rejection frequencies for SOB–VB bootstrap t tests for various values of α ($M = 400,000; B = 399; n = 40$)

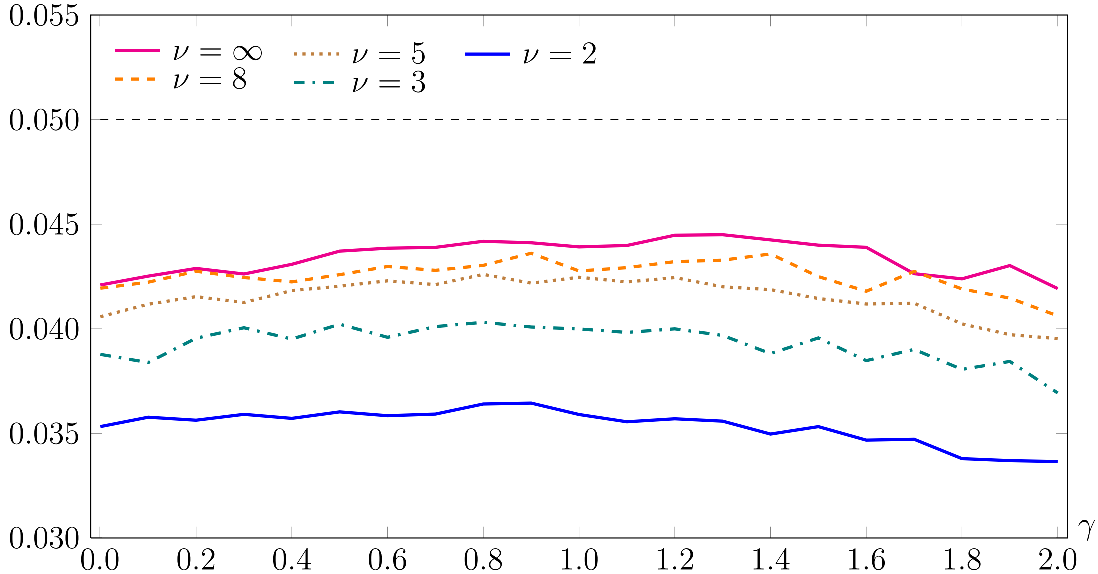

Figure 6: Rejection frequencies for w3r2–HC1 bootstrap t tests for various values of ν and $\alpha = -5$ ($M = 400,000; B = 399; n = 40$)

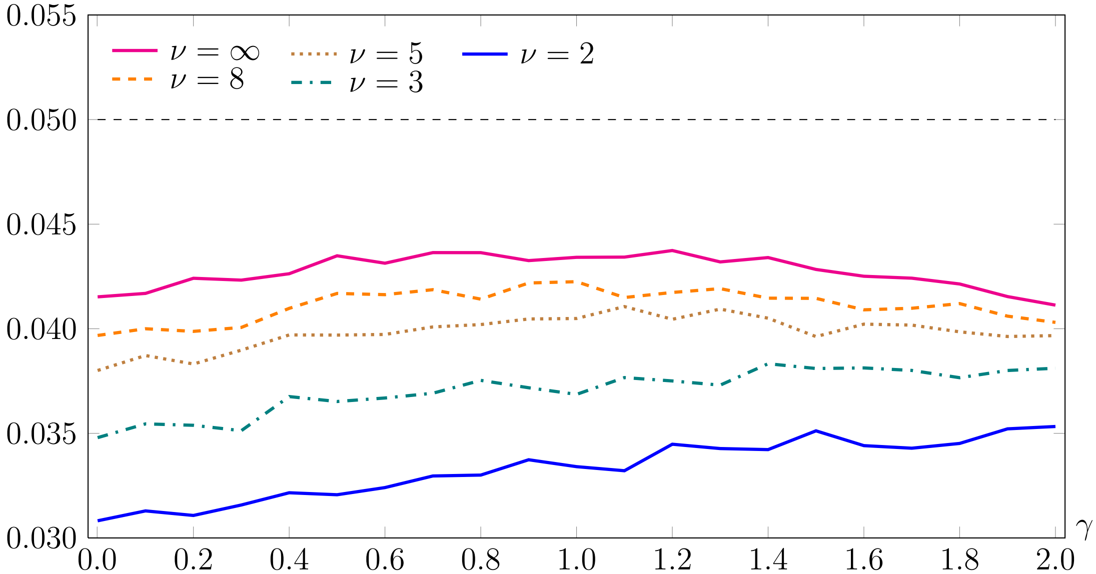

Figure 7: Rejection frequencies for SOB–VB t tests for various values of ν and $\alpha = -5$ $(M = 400{,}000; B = 399; n = 40)$

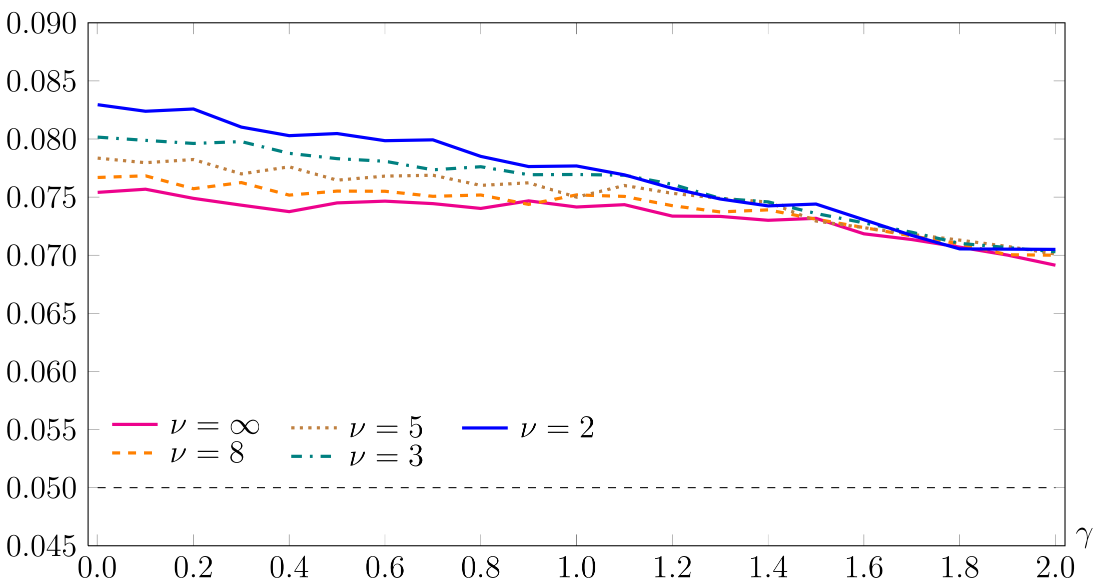

Figure 8: Rejection frequencies for w3r1–HC1 t tests for various values of ν and $\alpha = -5$ $(M = 400{,}000; B = 399; n = 40)$

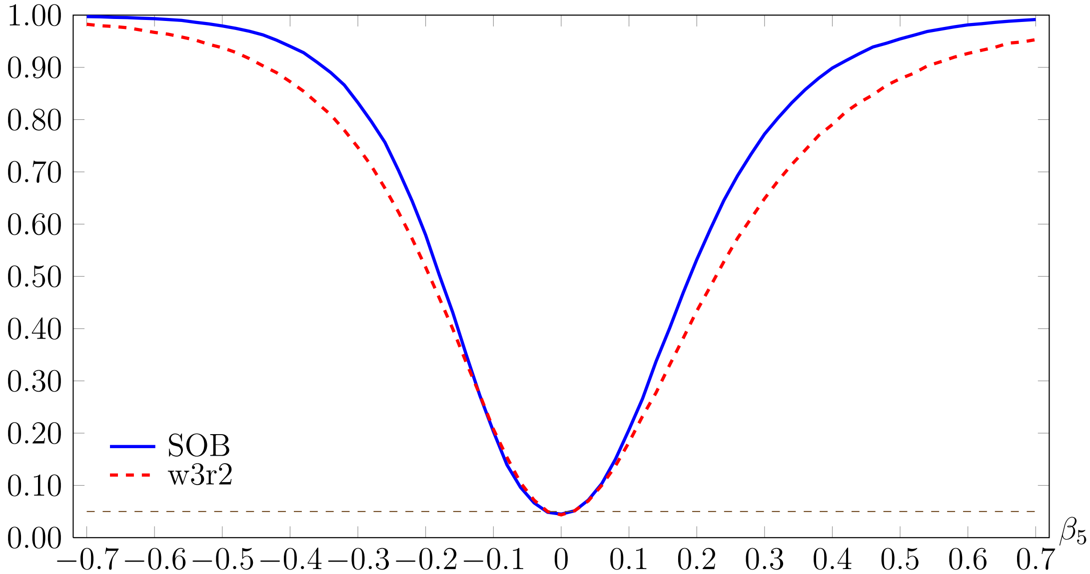

Figure 9: Power of SOB–VB and w3r2–HC1 bootstrap t tests for $\alpha = 0$, $\nu = \infty$ and $\gamma = 1$ ($M = 50,000; B = 399; n = 40$)

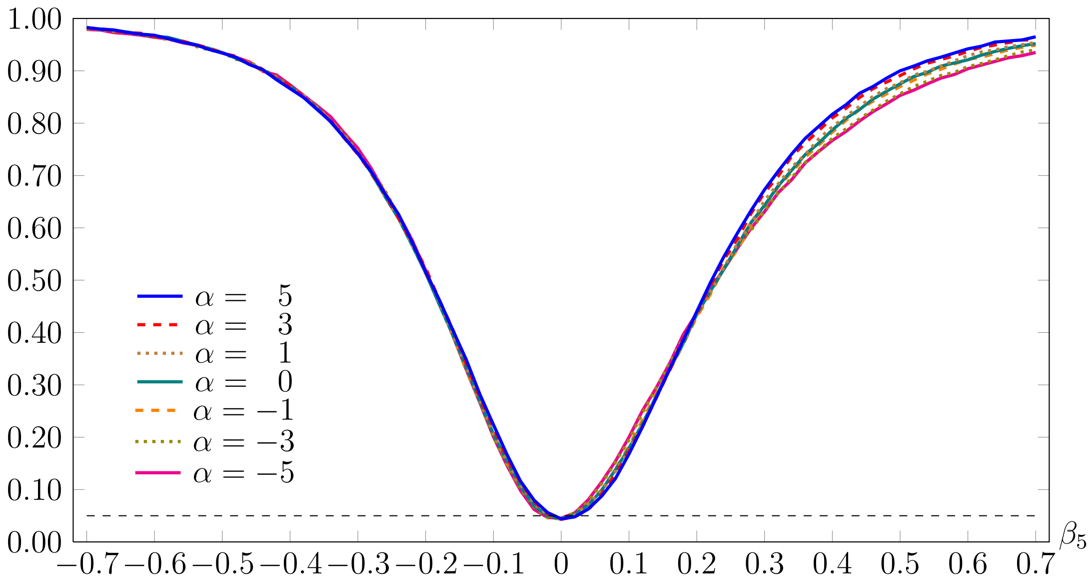

Figure 10: Power for w3r2–HC1 bootstrap t tests for various values of α, $\nu = \infty$ and $\gamma = 1$ ($M = 50,000; B = 399; n = 40$)

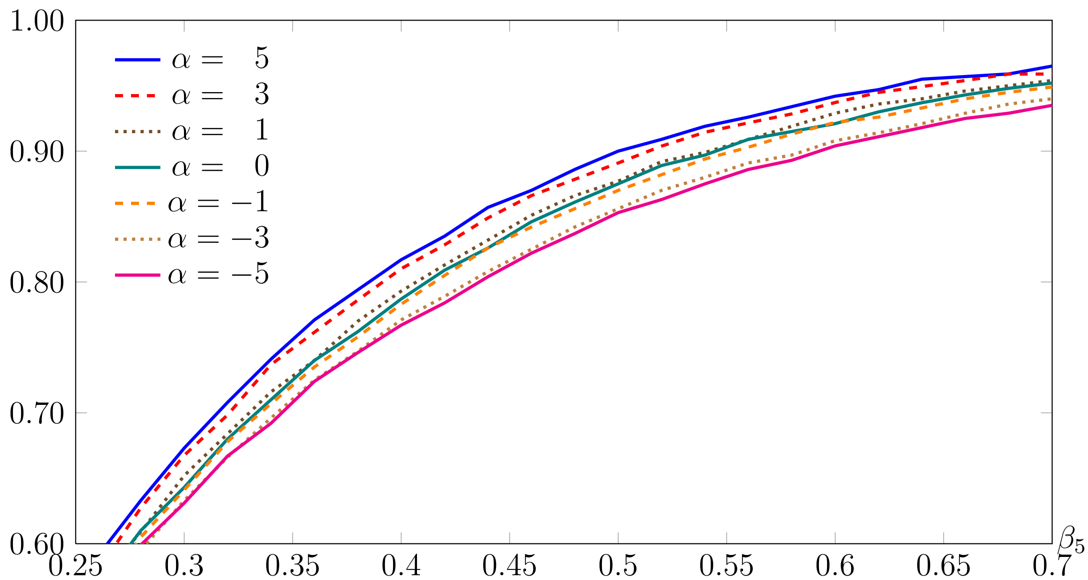

Figure 11: Zoomed-in view of **Figure 10**

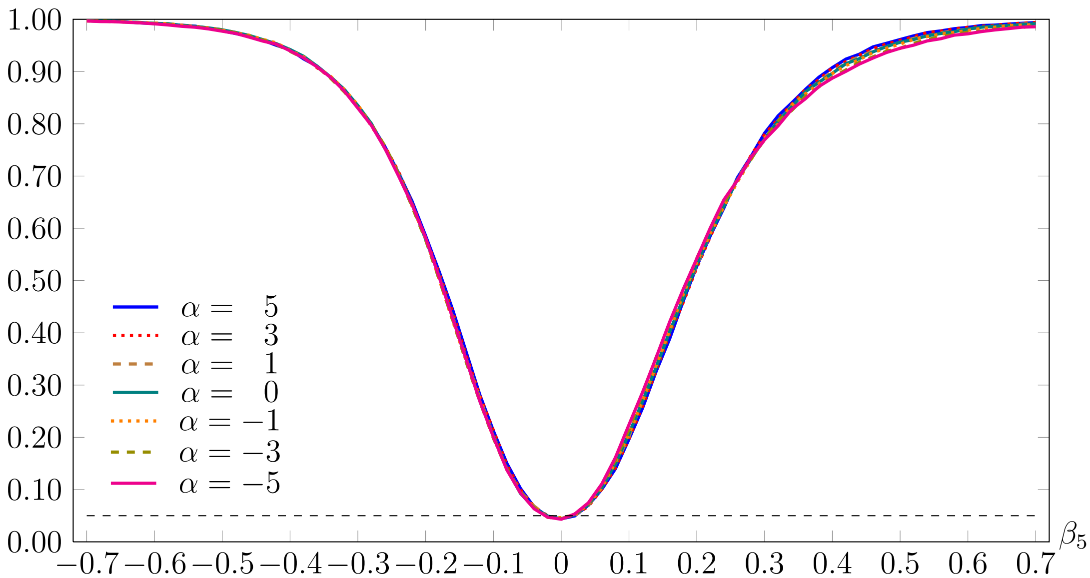

Figure 12: Power of SOB–VB bootstrap t tests for various values of α, $\nu = \infty$ and $\gamma = 1$ ($M = 50,000; B = 399; n = 40$)

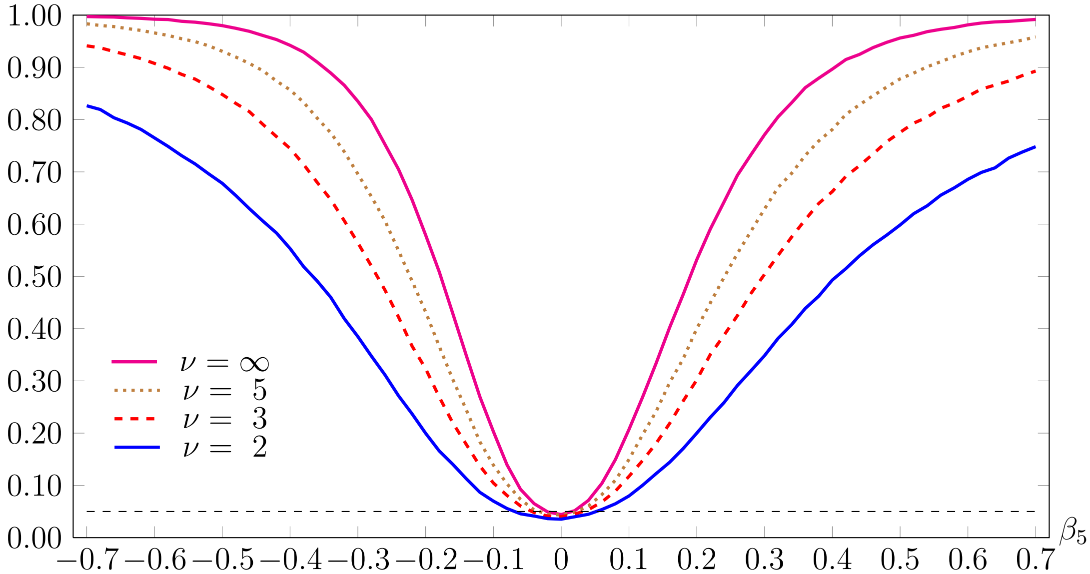

Figure 13: Power of SOB–VB bootstrap t tests for various values of ν, $\alpha = 0$, and $\gamma = 1$ ($M = 50,000; B = 399; n = 40$)

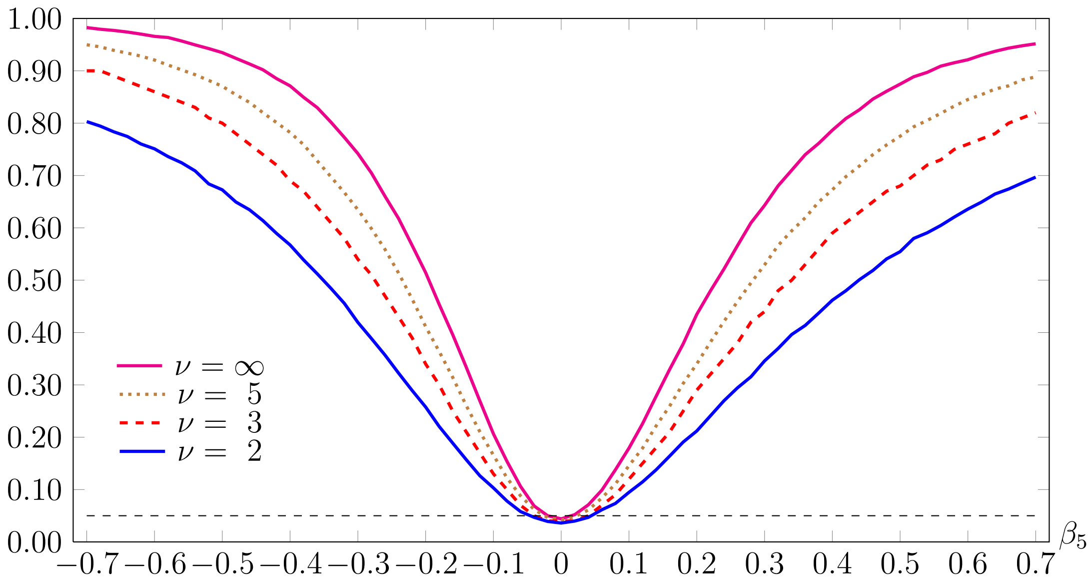

Figure 14: Power of w3r2–HC1 bootstrap t tests for various values of ν, $\alpha = 0$, and $\gamma = 1$ ($M = 50,000; B = 399; n = 40$)

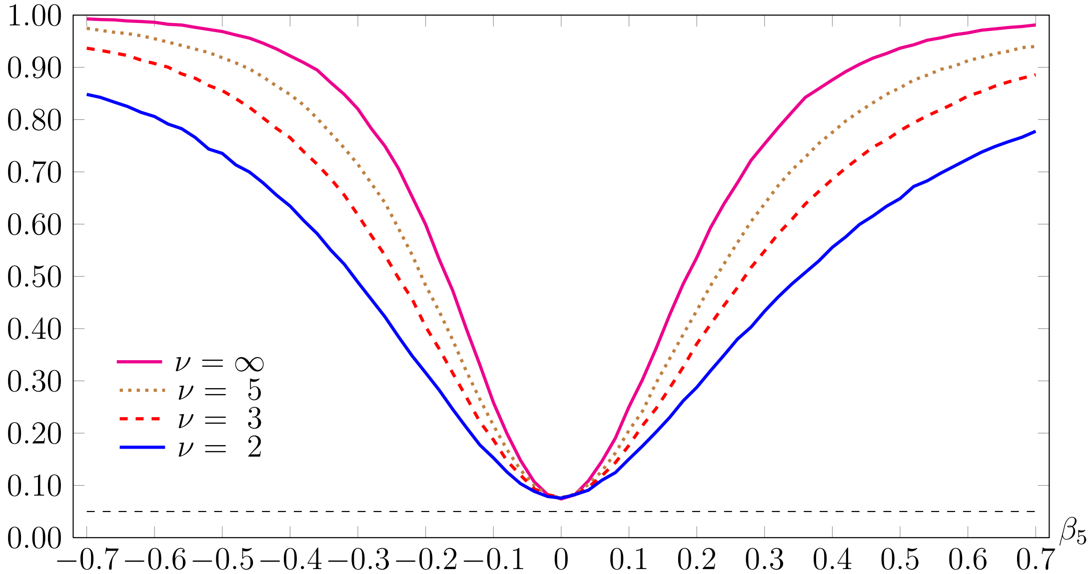

Figure 15: Power of w3r1–HC1 bootstrap t tests for various values of ν, $\alpha = 0$, and $\gamma = 1$ $(M = 50,000; B = 399; n = 40)$

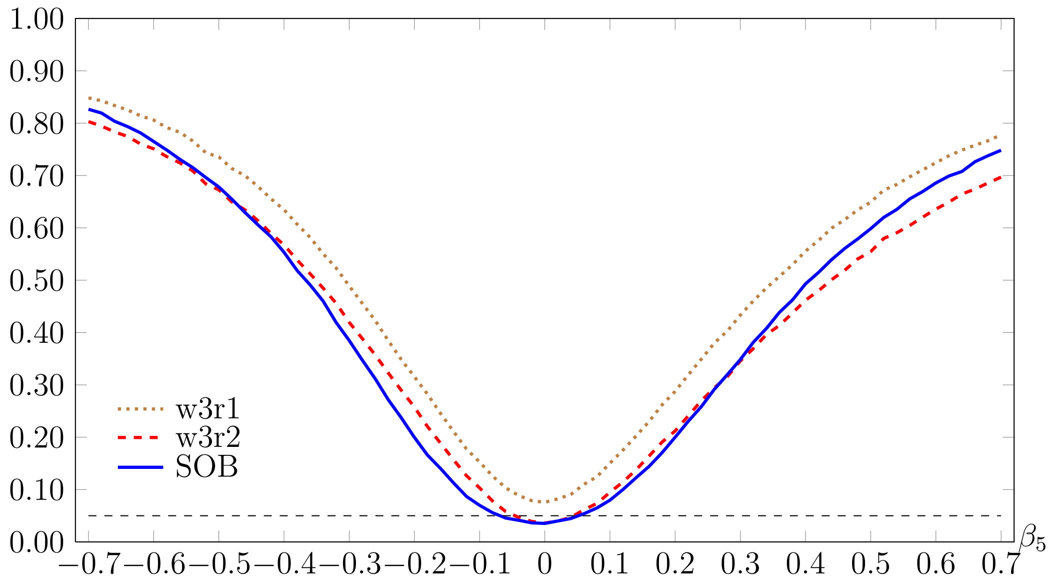

Figure 16: Power of SOB–VB, w3r1 and w3r2 HC1 bootstrap t tests for $\nu = 2$, $\alpha = 0$, $\gamma = 1$ $(M = 400,000; B = 399; n = 40)$

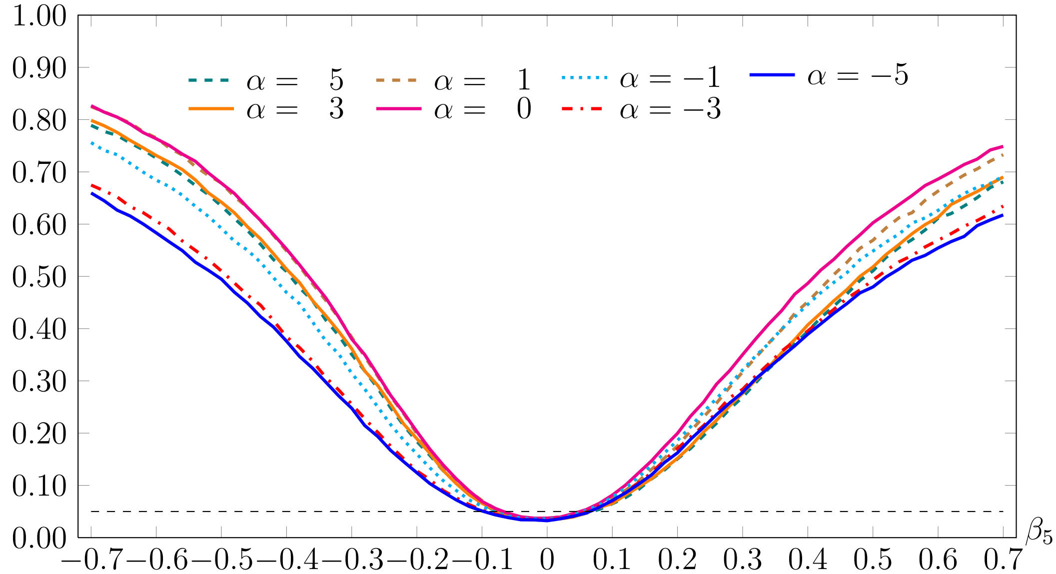

Figure 17: Power of SOB–VB bootstrap t tests for various values of α, $\nu = 2$ and $\gamma = 1$ $(M = 50,000; B = 399; n = 40)$

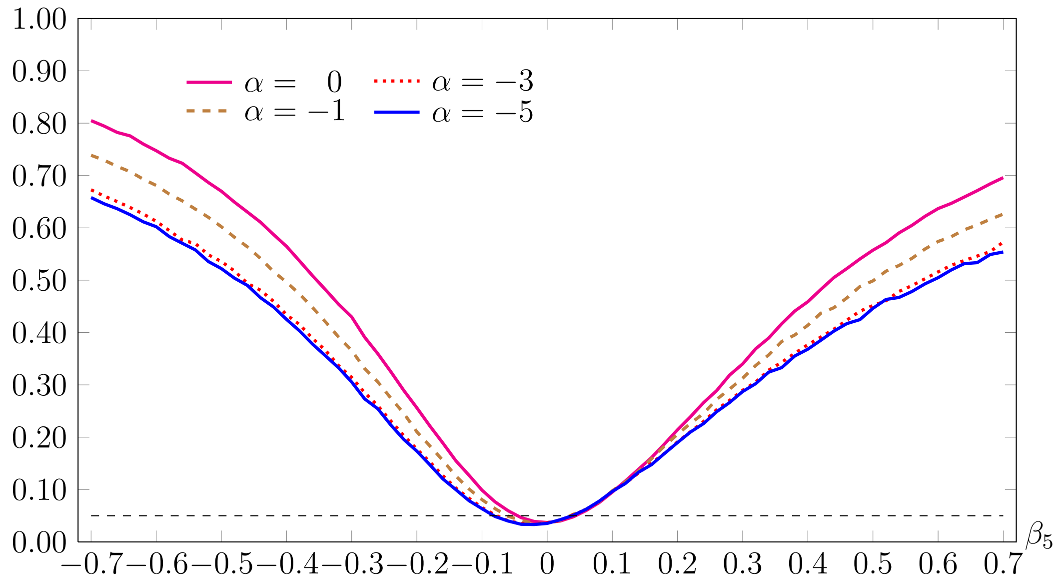

Figure 18: Power of w3r2–HC1 bootstrap t tests for various values of α, $\nu = 2$ and $\gamma = 1$ $(M = 50,000; B = 399; n = 40)$

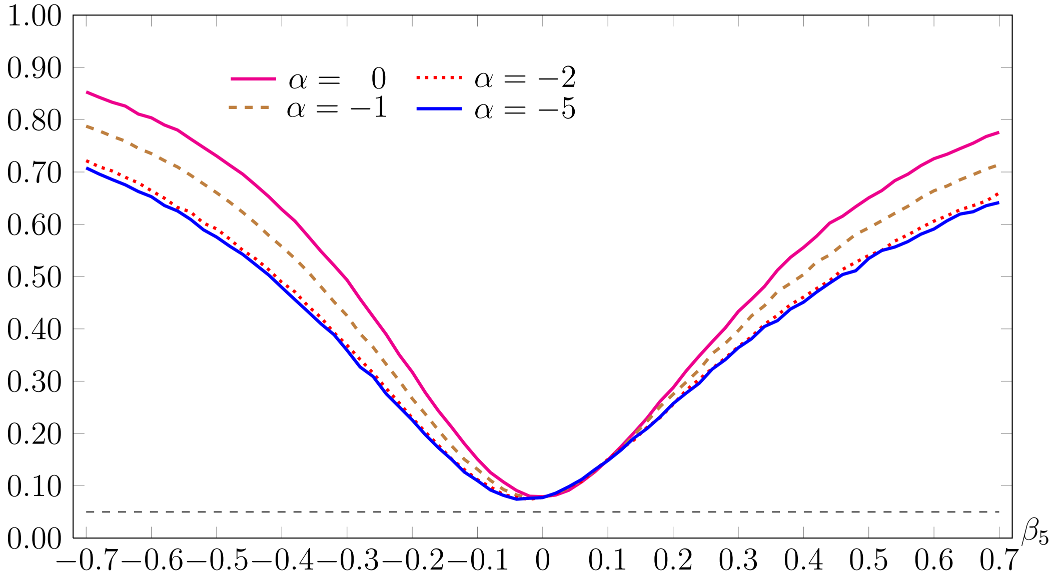

Figure 19: Power of w3r1–HC1 bootstrap t tests for various values of α, $\nu = 2$ and $\gamma = 1$ $(M = 50,000; B = 399; n = 40)$

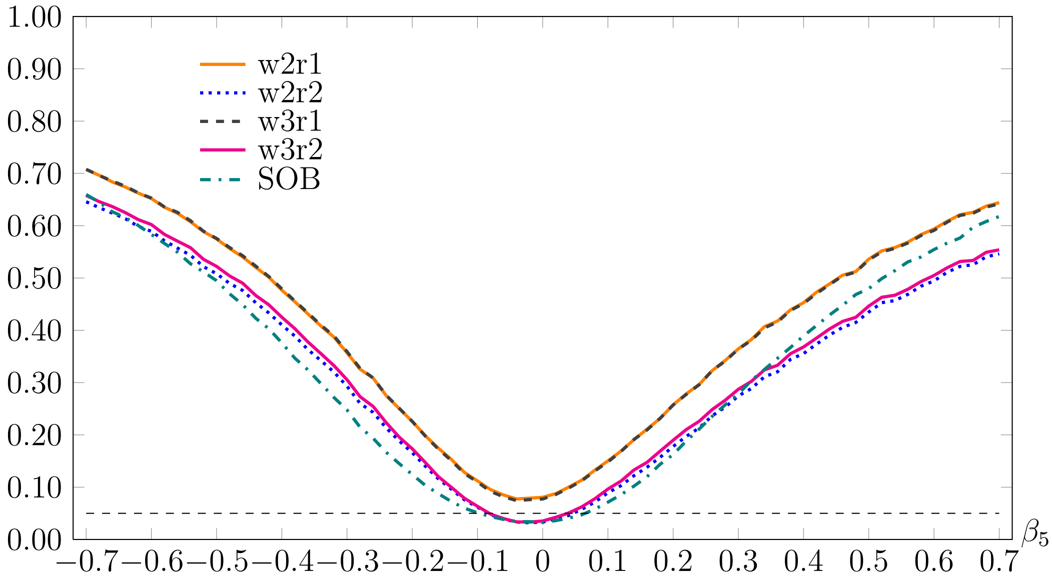

Figure 20: Power of SOB–VB and various WB-HC1 bootstrap t tests for α = -5, $\nu = 2$ and $\gamma = 1$ $(M = 50,000; B = 399; n = 40)$