TI 2013-211/III
Tinbergen Institute Discussion Paper

On Distributions of Ratios

Simon A. Broda[1]
Raymond Kan[2]

[1] *Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, The Netherlands.*
[2] *University of Toronto, Canada.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

On Distributions of Ratios

Simon A. Broda

Department of Economics and Econometrics, University of Amsterdam

Tinbergen Institute Amsterdam

Raymond Kan

Joseph L. Rotman School of Management, University of Toronto

Summary. A large number of exact inferential procedures in statistics and econometrics involve the sampling distribution of ratios of random variables. If the denominator variable is positive, then tail probabilities of the ratio can be expressed as those of a suitably defined difference of random variables. If in addition, the joint characteristic function of numerator and denominator is known, then standard Fourier inversion techniques can be used to reconstruct the distribution function from it. Most research in this field has been based on this correspondence, but which breaks down when both numerator and denominator are supported on the entire real line. The present manuscript derives inversion formulae and saddlepoint approximations that remain valid in this case, and reduce to known results when the denominator is almost surely positive. Applications include the IV estimator of a structural parameter in a just identified equation.

Keywords: Characteristic Function, Inversion Formula, Saddlepoint Approximation, Simultaneous Equations, Instrumental Variables, Weak Instruments, Bootstrap

1. Introduction

The distributions of many random variables of interest do not permit an analytic representation. By contrast, their characteristic and moment generating functions are often much more tractable. Results which express the distribution or density function in terms of these are commonly referred to as inversion formulae. They typically involve unsolved integrals, which have to be evaluated numerically in cases of interest.

In many cases, even the characteristic function is intractable, but the statistic may have a stochastic representation in terms of random variables for which it is readily available. The most common case, and the subject of this paper, is that of a ratio. Many test statistics and estimators in econometrics are of this form. The denominator of such ratios is typically related to some form of sample variance, and hence positive. This situation is quite fortunate, because there exist inversion formulae that express the density and distribution function of such a statistic in terms of the joint characteristic function of numerator and denominator. There are, however, important situations in which the quantity of interest is in the form of a ratio of random variables that both take values on the entire real line. Gurland (1948) has derived an inversion formula for this case, but its applicability is limited. The present manuscript derives expressions that are more expedient, in that they are amenable to numerical evaluation and lead to simple asymptotic expansions.

In fact, two results will be proven below. The first one shows that the standard results for ratios with positive denominator apply more generally than previously known. They are not, however, general enough to cover a number of situations of interest, including our applications. Our second result, on the other hand, is. The price that one has to pay is that it contains a double integral that must be evaluated numerically. Even with modern computers, this severely diminishes its usefulness for applications. We overcome this problem by providing a saddlepoint approximation to both the density and distribution function of a ratio of random variables. Saddlepoint approximations have been introduced to statistics by Daniels (1954) and have found numerous applications since. We do not attempt to provide a full bibliography here, but refer to the book-length treatment of Butler (2007) instead. Daniels had already considered the case of ratios of random variables, but his result is also limited to cases with positive denominator.

One of our numerical examples concerns the two stage least squares estimator of the structural parameter in a simultaneous equations model with one endogenous regressor and one, possibly weak, external instrument. The distribution of this estimator under normality has been studied intensively; see Richardson (1968), Sawa (1969), Anderson and Sawa (1973), and Holly and Phillips (1979). The realization that the asymptotic normality of the estimator is a poor approximation to the true sampling distribution if the instruments are weak has spurred renewed interest in the topic, as evidenced by the work of Nelson and Startz (1990a,b), Maddala and Jeong (1992), Woglom (2001), and the papers by Hillier (2006), Forchini (2006) and Phillips (2006), which comprise the entire 'Miscellanea' section of that issue of *Econometric Theory*. Only few authors have considered the distribution under non-Gaussianity. Knight (1986) discusses the case in which the error distribution is expandable in an Edgeworth-type series, and Broda (2013) assumes that the errors follow a multivariate generalized hyperbolic distribution. The results of the present manuscript facilitate the computation of the density and distribution functions under far more general assumptions. We give special attention to the bootstrap distribution.

The remainder of the paper is organized as follows. Section 2 derives two novel results concerning inversion formulae for ratios. Section 3 derives saddlepoint approximations for the density and distribution functions. Section 4 provides numerical examples, including the bootstrap distribution of the two stage least squares estimator. Section 5 concludes. Three appendices provide mathematical details.

2. Inversion Formulae for Ratios

We begin by recapitulating some known results from the literature. Consider a random variable X and denote by $F_X(x)$ and $\varphi_X(s)$, respectively, the associated distribution and characteristic functions. Gurland (1948) and Gil-Pelaez (1951) show that at every point of continuity of F_X,

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi}\int_0^\infty \operatorname{Im}\left[\frac{e^{-\mathrm{i}sx}\varphi_X(s)}{s} - \frac{e^{\mathrm{i}sx}\varphi_X(-s)}{s}\right] \mathrm{d}s$$

$$= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \mathrm{Im}\left[e^{-\mathrm{i}sx}\varphi_X(s)\right]\frac{\mathrm{d}s}{s}. \tag{1}$$

Care must be taken in interpreting the integral sign in (1). Wendel (1961) has shown that depending on φ_X, the integral may fail to converge absolutely. The weakest known condition for absolute convergence is

$$\mathbb{E}[\log(1 + |X|)] < \infty,$$

as given in Rosén (1961). Consequently, Gil-Pelaez relied on Riemann integrals in his derivation, and Gurland employed principal value integrals. Shephard (1991) provides a multivariate generalization of (1). In the bivariate case, his result is

$$F_{X,Y}(x, y) = \frac{1}{2}\left[F_X(x) + F_Y(y)\right] - \frac{1}{4} - \frac{1}{2\pi^2}\int_0^\infty\int_0^\infty \frac{\mathrm{Re}\left[e^{-\mathrm{i}sx-\mathrm{i}ty}\varphi_{X,Y}(s,t) - e^{-\mathrm{i}sx+\mathrm{i}ty}\varphi_{X,Y}(s,-t)\right]}{st}\mathrm{d}s\,\mathrm{d}t, \tag{2}$$

provided the mean of (X, Y) is finite and $\varphi_{X,Y}(s, t)$ is absolutely integrable, which implies that (X, Y) is absolutely continuous. These conditions remove the need for principal value integrals as in Gurland (1948).

We are interested in the probability density function (pdf) and and cumulative distribution function (cdf) of $R = X/Y$, the ratio of two absolutely continuous random variables. The standard approach is to express the pdf as

$$f_R(r) = \int_{-\infty}^{\infty} |y| f_{X,Y}(ry, y)\mathrm{d}y, \tag{3}$$

where $f_{X,Y}(x, y)$ is the joint pdf of X and Y. The cdf of R can then be obtained by integrating $f_R(r)$. Clearly this is only practical if $f_{X,Y}(x, y)$ is readily available. If this is not the case, then the joint characteristic function of X and Y may nevertheless be tractable. It is thus useful to express the pdf and cdf of R in terms of $\varphi_{X,Y}(s, t)$.

When Y is almost surely positive and r is not an atom of R, then it is straightforward to obtain the cdf of $R = X/Y$ from the relation

$$F_R(r) = \mathbb{P}[R < r] = \mathbb{P}[X < rY] = \mathbb{P}[X - rY < 0] = \mathbb{P}[W_r < 0],$$

where $W_r = X - rY$, and the subscript r in W_r will be suppressed below when there is no source of confusion. Provided that $\mathbb{E}[\log(1 + |X - rY|)] < \infty$, an application of (1) shows that

$$\begin{aligned} F_R(r) &= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \mathrm{Im}\left[\varphi_W(s)\right]\frac{\mathrm{d}s}{s} \\ &= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \mathrm{Im}\left[\varphi_{X,Y}(s, -rs)\right]\frac{\mathrm{d}s}{s}. \end{aligned} \tag{4}$$

Gurland (1948) derives a similar result using principal value integrals. If in addition, the mean of (X, Y) is finite, then $\varphi_{X,Y}$ is differentiable. Let $\varphi_2(s, t) = \frac{\partial}{\partial t}\varphi_{X,Y}(s, t)$. Provided that $\varphi_2(s, -rs)$ is absolutely integrable, then by dominated convergence, the pdf of R is

$$f_R(r) = \frac{1}{\pi}\int_0^\infty \mathrm{Im}[\varphi_2(s, -rs)]\mathrm{d}s. \tag{5}$$

Geary (1944) was the first to demonstrate such a result. The case with Y negative is treated analogously.

A more general expression is needed when both X and Y can take values on the entire real line. One such result is given in Gurland (1948, Theorem 2; see also Curtiss, 1941). Gurland shows that if 0 is not an atom of Y and F_R is continuous at r, then

$$F_R(r) = \frac{1}{2} - \frac{1}{2\pi}\int_{-\infty}^\infty \frac{\mathrm{Im}\left[\varphi^+_{X,Y}(s, -rs) + \varphi^-_{X,Y}(-s, rs)\right]}{s}\mathrm{d}s,$$

where

$$\begin{aligned} \varphi^+_{X,Y}(s, -rs) &= \int_{-\infty}^\infty \int_0^\infty e^{\mathrm{i}s(x-ry)} f(x, y)\mathrm{d}y\mathrm{d}x, \\ \varphi^-_{X,Y}(-s, rs) &= \int_{-\infty}^\infty \int_{-\infty}^0 e^{-\mathrm{i}s(x-ry)} f(x, y)\mathrm{d}y\mathrm{d}x, \end{aligned}$$

and principal values are to be taken if an integral fails to converge absolutely. The challenge in using this result is that explicit expressions for $\varphi^+_{X,Y}$ and $\varphi^-_{X,Y}$ are generally difficult to obtain. It is therefore preferable to express the pdf and cdf of R in terms of $\varphi_{X,Y}$ directly.

We will derive two such results. The first one shows that (5) remains valid if some linear combination of X and Y is almost surely positive or negative, that is, if X and Y form a definite pair, defined as follows.

Definition 1 (Definite Pair). *We call two real-valued random variables a* definite pair *if* $\exists \beta \in \{\mathbb{R} \cup \infty\} : \mathbb{P}[X - \beta Y < 0] = \delta$, *for* $\delta \in \{0, 1\}$.

Trivially, if X is positive with probability one, then $X - \beta Y$ is a positive random variable for $\beta = 0$, but it is less apparent that two random variables can form a definite pair even if both X and Y can take positive and negative values. The following is a simple example. Let $X = 2Z_1^2 - Z_2^2$ and $Y = Z_1^2 - 2Z_2^2$, where Z_1 and Z_2 are independent standard Gaussian, so that $W = X - rY = (2-r)Z_1^2 + (2r-1)Z_2^2$. Then $\mathbb{P}[W < 0] = 0$ for $1/2 \le r \le 2$.

Our result is based on the following identity, which appears not to be well known in the literature.

Lemma 1. *If X and Y form a definite pair such that $\mathbb{P}[X - \beta Y < 0] = \delta$ for $\delta \in \{0, 1\}$ and 0 is not an atom of Y, then*

$$\mathbb{P}[R < r] = 2\delta H(r - \beta) + (1 - 2\delta)\Big\{\mathbb{P}[Y < 0] + \operatorname{sgn}(r - \beta)\mathbb{P}[W < 0]\Big\},$$

where $R = X/Y$ and $H(\cdot)$ is the Heaviside function.

Proof. Appendix C. □

The following result follows at once.

Theorem 1. *If X and Y form a definite pair such that $\mathbb{P}[X - \beta Y < 0] = \delta$ for $\delta \in \{0, 1\}$, 0 is an atom of neither Y nor $W \equiv X - rY$, $\mathbb{E}[\log(1 + |Y|)] < \infty$, and $\mathbb{E}[\log(1 + |W|)] < \infty$, then*

$$F_R(r) = H(r - \beta) - \frac{(1 - 2\delta)}{\pi} \int_0^\infty \operatorname{Im}\Big[\varphi_{X,Y}(0, s) + \operatorname{sgn}(r - \beta)\varphi_{X,Y}(s, -rs)\Big]\frac{\mathrm{d}s}{s}. \quad (6)$$

If in addition, Y has a finite mean and $\varphi_2(s, -rs)$ is absolutely integrable, then R has a density and

$$f_R(r) = \frac{\operatorname{sgn}(r - \beta)}{\pi(2\delta - 1)} \int_0^\infty \operatorname{Im}\left[\varphi_2(s, -rs)\right] \mathrm{d}s = \left|\frac{1}{\pi}\int_0^\infty \operatorname{Im}\left[\varphi_2(s, -rs)\right] \mathrm{d}s\right|. \quad (7)$$

Proof. Observe that r is an atom of R if and only if 0 is an atom of W. For the cdf, use (1) in Lemma 1, together with $\varphi_{W,Y}(s,t) = \varphi_{X,Y}(s,t-rs)$. For the pdf, finiteness of the mean guarantees the existence of φ_2. The result follows by dominated convergence, provided that $\varphi_2(s,-rs)$ is absolutely integrable. □

Hence, provided that one takes the absolute value of the result, Geary's formula remains valid even if Y can take positive and negative values, as long as X and Y form a definite pair. We remark that at a set of isolated points, (7) may differ from $F_R'(r)$. Consider the following example. Let $X = Z_1^2$ and $Y = Z_1 Z_2$, where Z_1 and Z_2 are independent standard Gaussian. It can readily be shown that (7) yields $f_R(0) = 0$ for $R = X/Y = Z_1/Z_2$, whereas $F_R'(0) = 1/\pi$.

Our second result provides general inversion formulae for ratios that remain valid when (6) and (7) fail. We start from the following observation.

Lemma 2. *If 0 is an atom of neither Y nor $W \equiv X - rY$, then*

$$F_R(r) = \mathbb{P}[W<0] + \mathbb{P}[Y<0] - 2\mathbb{P}[W<0, Y<0]. \tag{8}$$

Proof. Observe that r is an atom of R if and only if 0 is an atom of W. Hence

$$\begin{aligned} F_R(r) &= \mathbb{P}\left[\frac{X}{Y} < r\right] = \mathbb{P}\left[\frac{X}{Y} < r, Y<0\right] + \mathbb{P}\left[\frac{X}{Y} < r, Y>0\right] \\ &= \mathbb{P}\left[X > rY, Y<0\right] + \mathbb{P}\left[X < rY, Y>0\right] \\ &= \mathbb{P}\left[W > 0, Y<0\right] + \mathbb{P}\left[W<0, Y>0\right] \\ &= \mathbb{P}[W<0] + \mathbb{P}[Y<0] - 2\mathbb{P}[W<0, Y<0] \end{aligned}$$

□

We then have the following result.

Theorem 2. *If (X,Y) has a finite mean, $\varphi_{X,Y}$ is absolutely integrable and 0 is not an atom of $W \equiv X - rY$, then for $|r| < \infty$,*

$$F_R(r) = \frac{1}{2} + \frac{1}{\pi^2}\int_0^\infty \int_0^\infty \frac{\mathrm{Re}\left[\varphi_{X,Y}(s,t-rs) - \varphi_{X,Y}(s,-t-rs)\right]}{st} \mathrm{d}s\, \mathrm{d}t \tag{9}$$

and

$$f_R(r) = \frac{1}{\pi^2}\int_0^\infty \int_{-\infty}^\infty \mathrm{Re}[\varphi_2(s,-t-rs)]\mathrm{d}s\, \frac{\mathrm{d}t}{t} \tag{10}$$

whenever this integral converges absolutely.

Proof. The assumptions imply that (W, Y) has a finite mean, and that its characteristic function $\varphi_{W,Y}(s,t) = \varphi_{X,Y}(s, t - rs)$ is absolutely integrable. Hence (2) applies. Combining it with (8) completes the proof for the cdf. Integrability of $\varphi_{W,Y}(s,t)$ ensures that R has a density. By dominated convergence,

$$f_R(r) = \frac{1}{\pi^2} \int_0^\infty \int_0^\infty \frac{\mathrm{Re}\,[\varphi_2(s, -t - rs) - \varphi_2(s, t - rs)]}{t} \mathrm{d}s\, \mathrm{d}t$$

whenever the integral is absolutely convergent. The result follows upon noting that $\varphi_{X,Y}(s,t) = \overline{\varphi}_{X,Y}(-s,-t)$, so that $-\mathrm{Re}[\varphi_2(s, t - rs)] = \mathrm{Re}[\varphi_2(-s, -t + rs)]$. □

To see how (10) reduces to (7) when X and Y form a definite pair, assume that $r \neq \beta$ and rewrite (10) as

$$f_R(r) = -\frac{1}{\pi^2}\,\mathrm{Re} \int_0^\infty \oint_{-\infty}^\infty \varphi_2(s, t - rs) \frac{\mathrm{d}t}{t}\, \mathrm{d}s,$$

where the circled integral represents the Cauchy principal value. We consider the case with $\delta = 1$ and $\beta < \infty$; the other cases can be treated analogously. Make the change of variables $s \mapsto s + t$, $t \mapsto (r - \beta)t$, so that the inner integrand becomes $f_s(t) \equiv \mathrm{sgn}(r - \beta)\varphi_2(s + t, -sr - \beta t)$. Observe that $\varphi_{X,Y}(s + t, -sr - \beta t) = \mathbb{E}[\exp(\mathrm{i}sW_r + \mathrm{i}tW_\beta)]$, where $W_\beta < 0$ almost surely. This implies that for real s and as a function of t, $\varphi_{X,Y}(s + t, -sr - \beta t)$ (and hence $f_s(t)$) is analytic for $\mathrm{Im}\, t < 0$. Consider a contour that consists of a line segment from $-T$ to $-1/T$, a small counterclockwise loop half way around the origin, another line segment from $1/T$ to T, and a large semicircle in the lower half of the complex plane back to $-T$. The contour encloses no singularities, hence the integral along it is zero. As $T \to \infty$, the integral along the large semicircle converges to zero. The integral along the half loop around the origin is equal to minus one half the residue at the origin, and hence

$$\oint_{-\infty}^\infty f_s(t) \frac{\mathrm{d}t}{t} = \mathrm{i}\pi f_s(0).$$

3. Saddlepoint Approximation

3.1. Pdf Approximation

In this section, we derive a saddlepoint approximation to the density of a ratio of two random variables that do not necessarily form a definite pair. The first step is to rewrite (10) in a form amenable to saddlepoint methods.

Suppose that X and Y have a joint density, and that their joint cumulant generating function (cgf) $\mathbb{K}(s,t) \equiv \log \mathbb{E}[\exp(sX + tY)]$ converges on the open set $\mathcal{T} \ni (0,0)$. Let $\bar{X}$ and $\bar{Y}$ denote the mean of n independent copies of X and Y, respectively, and consider the ratio $R \equiv \bar{X}/\bar{Y}$. From (8), the distribution function of R is

$$F_R^n(r) = \mathbb{P}[\bar{W} < 0] + \mathbb{P}[\bar{Y} < 0] - 2\mathbb{P}[\bar{W} < 0, \bar{Y} < 0], \tag{11}$$

where $\bar{W} = \bar{X} - r\bar{Y}$. Our goal is to arrive at an expression for the density $f_R^n(r)$ by differentiation. We will therefore require an expression for the tail probabilities appearing in (8), $F_{\bar{W}}^n(\bar{w})$ and $F_{\bar{W},\bar{Y}}^n(\bar{w},\bar{y})$, say. By a standard Laplace inversion argument, the joint density of $\bar{W}$ and $\bar{Y}$ is

$$f_{\bar{W},\bar{Y}}^n(\bar{w},\bar{y}) = \left(\frac{n}{2\pi \mathrm{i}}\right)^2 \int_{c_2 - \mathrm{i}\infty}^{c_2 + \mathrm{i}\infty} \int_{c_1 - \mathrm{i}\infty}^{c_1 + \mathrm{i}\infty} e^{n(\mathbb{K}(s,t-rs) - s\bar{w} - t\bar{y})} \mathrm{d}s\, \mathrm{d}t,$$

where c_1 and c_2 are such that $(c_1, c_2 - rc_1) \in \mathcal{T}$. Note that $\mathbb{K}(s, t - rs)$ is the joint cgf of $\bar{W}$ and $\bar{Y}$.

Choosing $c_1 < 0$, $c_2 < 0$ and integrating between $-\infty$ and zero yields the orthant probability

$$F_{\bar{W},\bar{Y}}^n(0,0) = \left(\frac{1}{2\pi \mathrm{i}}\right)^2 \int_{c_2 - \mathrm{i}\infty}^{c_2 + \mathrm{i}\infty} \int_{c_1 - \mathrm{i}\infty}^{c_1 + \mathrm{i}\infty} e^{n\mathbb{K}(s,t-rs)} \frac{\mathrm{d}s}{s} \frac{\mathrm{d}t}{t}. \tag{12}$$

Similarly, the marginal density of $\bar{W}$ is

$$f_{\bar{W}}^n(\bar{w}) = \frac{n}{2\pi \mathrm{i}} \int_{c_3 - \mathrm{i}\infty}^{c_3 + \mathrm{i}\infty} e^{n(\mathbb{K}(s,-rs) - s\bar{w})} \mathrm{d}s,$$

where c_3 is such that $(c_3, -rc_3) \in \mathcal{T}$. Choosing $c_3 < 0$ and integrating between $-\infty$ and zero yields

$$F_{\bar{W}}^n(0) = -\frac{1}{2\pi \mathrm{i}} \int_{c_3 - \infty}^{c_3 + \infty} e^{n\mathbb{K}(s,-rs)} \frac{\mathrm{d}s}{s}. \tag{13}$$

Differentiating (11) therefore produces

$$f_R^n(r) = I_1 + 2I_2,$$

where

$$I_1 = \frac{n}{2\pi \mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} e^{n\mathbb{K}(s,-rs)} \mathbb{K}_2(s,-rs) \mathrm{d}s, \tag{14}$$

$$I_2 = \frac{1}{2\pi \mathrm{i}} \int_{c_2-\mathrm{i}\infty}^{c_2+\mathrm{i}\infty} \frac{n}{2\pi \mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} e^{n\mathbb{K}(s,t-rs)} \mathbb{K}_2(s,t-rs) \mathrm{d}s \frac{\mathrm{d}t}{t}, \tag{15}$$

$\mathbb{K}_i(\cdot,\cdot)$ denotes the derivative of the joint cgf with respect to its ith argument, and we have set $c_1 = 0$, which is permissible because differentiation has removed the pole at $s = 0$. If $c_2 > 0$, then the residue at the origin must be subtracted, which is precisely I_1. We can therefore write

$$f_R^n(r) = 2I_2 - \mathrm{sgn}(c_2) I_1, \quad (0, c_2) \in \mathcal{T} \setminus (0,0). \tag{16}$$

We begin by deriving a saddlepoint approximation to I_2. The plan is to apply a standard Laplace approximation to the inner integral, and then approximate the integral in t by a saddlepoint approximation, modified as in Chester et al. (1957) and Bleistein (1966) to accommodate the pole at the origin. This parallels Skovgaard's (1987) derivation of a saddlepoint approximation for conditional distribution functions. The relevant integral in Skovgaard's problem has a similar form; the essential difference is the presence of the term involving $\mathbb{K}_2(\cdot,\cdot)$ in I_2.

Choose a compact subset $\mathcal{R}$ of the range of $\mathbb{K}_1(s,t)/\mathbb{K}_2(s,t)$. Applying a standard Laplace approximation to the integral in s yields

$$I_2 = \left(\frac{n}{2\pi}\right)^{1/2} \frac{1}{2\pi \mathrm{i}} \int_{c_2-\mathrm{i}\infty}^{c_2+\mathrm{i}\infty} e^{nh(t)} g_0(t) \left\{1 + \mathcal{O}\left(n^{-1}\right)\right\} \frac{\mathrm{d}t}{t}, \tag{17}$$

where $h(t) \equiv \mathbb{K}(\tilde{s}, t - r\tilde{s})$,

$$g_0(t) = \frac{\mathbb{K}_2(\tilde{s}, t - r\tilde{s})}{\sqrt{\mathbf{c}_r' \mathbb{K}''(\tilde{s}, t - r\tilde{s}) \mathbf{c}_r}},$$

$\mathbf{c}_r \equiv (1, -r)'$, $\mathbb{K}''(\cdot,\cdot) = \{\mathbb{K}_{ij}(\cdot,\cdot)\}$ denotes the Hessian of the cgf, and for each value of t, the saddlepoint $\tilde{s} = \tilde{s}(t)$ solves

$$\mathbb{K}_1(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_2(\tilde{s}, t - r\tilde{s}) = 0. \tag{18}$$

We refer to $\tilde{s}$ as the *inner saddlepoint*.

In order to approximate the integral in (17), we will need the following result, which is a special case of a theorem due to Bleistein (1966); for a simple derivation see Broda (2012). We state the result as a lemma, with the notation adapted to the current manuscript.

LEMMA 3. *If $g_0(t)$ and $h(t)$ are real functions of t, analytic in a strip containing $c \neq 0$ and the imaginary axis, and $h(t)$ has a unique saddle point $\hat{t}_r \neq 0$ on the real axis in the interior of this strip, then*

$$\frac{1}{2\pi\mathrm{i}}\int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} g_0(t)e^{nh(t)}\frac{\mathrm{d}t}{t} = e^{nh(0)}g_0(0)\Big(1_{c>0} - \Phi(\hat{w}\sqrt{n})\Big) + \frac{e^{nh(\hat{t}_r)}}{\sqrt{2\pi n}}\left(\frac{g_0(\hat{t}_r)}{\hat{u}} - \frac{g_0(0)}{\hat{w}} + \mathcal{O}\left(n^{-1}\right)\right),$$

where $\Phi(\cdot)$ is the standard normal cdf, $\hat{w} \equiv \operatorname{sgn}(\hat{t}_r)\sqrt{-2\big(h(\hat{t}_r) - h(0)\big)}$, $\hat{u} \equiv \hat{t}_r\sqrt{h''(\hat{t}_r)}$, and for each r, the saddlepoint $\hat{t}_r$ solves $h'(\hat{t}_r) = 0$.

In order to apply this result to the problem at hand, we require $h'(t)$ and $h''(t)$, the first and second derivatives of $h(t)$. By virtue of (18),

$$h'(t) = \tilde{s}'(t)[\mathbb{K}_1(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_2(\tilde{s}, t - r\tilde{s})] + \mathbb{K}_2(\tilde{s}, t - r\tilde{s}) = \mathbb{K}_2(\tilde{s}, t - r\tilde{s}), \quad (19)$$

where $\tilde{s}'(t)$ denotes the derivative of $\tilde{s}$ with respect to t. This is easily found by differentiating (18), which yields

$$\tilde{s}'(t) = -\frac{\mathbb{K}_{12}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{22}(\tilde{s}, t - r\tilde{s})}{\mathbf{c}_r'\mathbb{K}''(\tilde{s}, t - r\tilde{s})\mathbf{c}_r}.$$

The second derivative evaluates to

$$\begin{aligned} h''(t) &= \tilde{s}'(t)[\mathbb{K}_{12}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{22}(\tilde{s}, t - r\tilde{s})] + \mathbb{K}_{22}(\tilde{s}, t - r\tilde{s}) \\ &= \mathbb{K}_{22}(\tilde{s}, t - r\tilde{s}) - \frac{[\mathbb{K}_{12}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{22}(\tilde{s}, t - r\tilde{s})]^2}{\mathbf{c}_r'\mathbb{K}''(\tilde{s}, t - r\tilde{s})\mathbf{c}_r} \\ &= \frac{|\mathbb{K}''(\tilde{s}, t - r\tilde{s})|}{\mathbf{c}_r'\mathbb{K}''(\tilde{s}, t - r\tilde{s})\mathbf{c}_r}. \end{aligned}$$

The saddlepoint $\hat{t}_r = \hat{t}_r(\tilde{s})$ is found by equating (19) to zero. Equivalently, $\hat{t}_r = \hat{t} + r\hat{s}$, where $(\hat{s}, \hat{t})$ — the *outer saddlepoint* — solves the system

$$\mathbb{K}'\left(\hat{s}, \hat{t}\right) \equiv \left[\mathbb{K}_1(\hat{s}, \hat{t})\ \mathbb{K}_2(\hat{s}, \hat{t})\right]' = \mathbf{0}. \tag{20}$$

In order to apply the lemma, we assume that $\hat{t}_r \neq 0$ (so that $\hat{t} \neq -r\hat{s}$) for the remainder of the proof; the other case will be dealt with separately. It is further observed that $(\hat{s}, \hat{t})$ is independent of r (so that this system needs only be solved once for any given cgf), and that (20) implies that

$$g_0(\hat{t}_r) = \frac{\mathbb{K}_2(\hat{s}, \hat{t})}{\sqrt{\mathbf{c}_r' \mathbb{K}''(\hat{s}, \hat{t}) \mathbf{c}_r}} = 0.$$

Let $\tilde{s}_0 \equiv \tilde{s}(0)$, i.e., the inner saddlepoint corresponding to $t = 0$, and define $\tilde{w}_0 \equiv \operatorname{sgn}\left(\tilde{s}_0\right) \sqrt{-2\mathbb{K}\left(\tilde{s}_0, -r\tilde{s}_0\right)}$ and

$$\tilde{g}_0 \equiv g_0(0) = \frac{\mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)}{\sqrt{\mathbf{c}_r' \mathbb{K}''(\tilde{s}_0, -r\tilde{s}_0) \mathbf{c}_r}}.$$

Then

$$I_2 = \sqrt{n}\phi(\sqrt{n}\tilde{w}_0)\tilde{g}_0 \left[1_{c_2>0} - \Phi(\sqrt{n}\hat{w}) - \frac{\phi(\sqrt{n}\hat{w})}{\sqrt{n}\hat{w}} + \mathcal{O}\left(n^{-1}\right)\right], \tag{21}$$

where ϕ is the standard normal pdf.

It remains to approximate I_1, but this is a straightforward exercise, as I_1 is the special case of the inner integral in I_2 with $t = 0$. The arguments that led to (17) therefore immediately yield

$$I_1 = \sqrt{n}\phi(\sqrt{n}\tilde{w}_0)\tilde{g}_0 \left(1 + \mathcal{O}\left(n^{-1}\right)\right).$$

Combining the two approximations according to (16) produces the desired result.

THEOREM 3. *Suppose that X and Y have a joint density with respect to Lebesgue measure on $\mathbb{R}^2$, and that their joint cgf $\mathbb{K}(s,t) \equiv \log \mathbb{E}[\exp(sX + tY)]$ converges on the open set $\mathcal{T} \ni (0,0)$, with gradient $\mathbb{K}'(s,t)$ and Hessian $\mathbb{K}''(s,t)$. Let $\bar{X}$ and $\bar{Y}$ denote the mean of n independent copies of X and Y, respectively. For*

$r \in \mathcal{R}$, a compact subset of the range of $\mathbb{K}_1(s,t)/\mathbb{K}_2(s,t)$, define the outer *and* inner saddlepoints $(\hat{s}, \hat{t})$ *and* $\tilde{s}_0$ *as the solutions to* $\mathbb{K}'\left(\hat{s},\hat{t}\right) = \mathbf{0}$ *and*

$$\mathbf{c}_r' \mathbb{K}'\left(\tilde{s}_0, -r\tilde{s}_0\right) = 0, \tag{22}$$

respectively, where $\mathbf{c}_r \equiv (1, -r)'$. *Then, provided that* $\hat{t} \neq -r\hat{s}$, *the density of the ratio* $R \equiv \bar{X}/\bar{Y}$ *is* $f_R^n(r) = \hat{f}_n^{(1)}(r)\left(1 + \mathcal{O}\left(n^{-1}\right)\right)$, *where*

$$\begin{aligned} \hat{f}_n^{(1)}(r) &= \sqrt{n}\phi(\sqrt{n}\tilde{w}_0)\tilde{g}_0 \left\{1 - 2\left[\Phi(\sqrt{n}\hat{w}) + \frac{\phi(\sqrt{n}\hat{w})}{\sqrt{n}\hat{w}}\right]\right\}, \qquad (23)\\ \tilde{g}_0 &\equiv \frac{\mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)}{\sqrt{\mathbf{c}_r'\mathbb{K}''(\tilde{s}_0, -r\tilde{s}_0)\mathbf{c}_r}},\\ \tilde{w}_0 &\equiv \operatorname{sgn}\left(\tilde{s}_0\right)\sqrt{-2\mathbb{K}\left(\tilde{s}_0, -r\tilde{s}_0\right)}, \quad \text{and}\\ \hat{w} &\equiv \operatorname{sgn}\left(\hat{t} + r\hat{s}\right)\sqrt{-2\left[\mathbb{K}\left(\hat{s},\hat{t}\right) - \mathbb{K}\left(\tilde{s}_0, -r\tilde{s}_0\right)\right]}. \end{aligned}$$

Higher order approximations are provided in Appendix A. In particular, the second order approximation is given in (37) and (42), and the third order approximation is given in (47) and (48).

A few remarks are in order. First, the approximate density is always non-negative. This is seen as follows. Consider the function $f : \mathbb{R} \to \mathbb{R}, x \mapsto \Phi(x) + \phi(x)/x$. Then $f(x) < 0, x < 0$ and $f(x) > 1, x > 0$. This follows directly from Gordon (1941, Eq. 7), who shows that for $x > 0$, $\Phi(-x)/\phi(-x) < 1/x$. Thus the term in curly braces in (23) is greater than one if $\hat{w} < 0$ and smaller than minus one if $\hat{w} > 0$. The result follows because $\tilde{g}_0$ and $\hat{w}$ have opposite signs, as shown in Appendix C. Second, the term in front of the curly braces (and thus the approximation for I_1) is the standard saddlepoint approximation derived in Daniels (1954) for the case with $\mathbb{P}[Y < 0] = 0$. One may therefore interpret the term in brackets as a correction for cases in which this requirement fails. Indeed, if X and Y form a definite pair, then $\hat{w}$ diverges to $\pm\infty$, and the two approximations coincide in absolute value. Since the term in curly braces is always greater than unity, the correction is, in general, upwards; hence using the absolute value of Daniels's approximation when it is not applicable will tend to underestimate the density. Third, it is seen that (23) is undefined when $\hat{w}=0$, which happens whenever $\hat{t}=-r\hat{s}$. This singularity is, however, removable. Two cases can be distinguished: (i) $\mu_X \equiv \mathbb{E}[X] \neq 0$ or $\mu_Y \equiv \mathbb{E}[Y] \neq 0$ (or both), so that $\hat{s}, \hat{t} \neq 0$. Then $f_n(r)$ has a removable singularity at $r^* \equiv -\hat{t}/\hat{s}$, and the limiting value is derived in Appendix A as

$$\hat{f}_n^{(1)}(r^*) = \sqrt{\frac{2}{\pi}}\phi(\sqrt{n}\tilde{w}_0)\frac{|\mathbb{K}''(\hat{s},\hat{t})|^{1/2}}{\mathbf{c}_{r^*}'\mathbb{K}''(\hat{s},\hat{t})\mathbf{c}_{r^*}}. \tag{24}$$

The limiting values of the second and third order approximations are given in (43), (46) and (49), (50), respectively. If, on the other hand, (ii) $\mu_X = \mu_Y = 0$, then $\hat{s}=\hat{t}=0$. Consequently (23) is undefined for all r and should be replaced by the limit

$$\hat{f}_n^{(1)}(r) = \frac{1}{\pi}\frac{|\mathbf{\Sigma}|^{1/2}}{\mathbf{c}_r'\mathbf{\Sigma}\mathbf{c}_r},$$

where $\mathbf{\Sigma} \equiv \mathbb{K}''(0,0)$ is the covariance matrix of (X,Y). We note that in this case, the accuracy of the approximation relative to the main term is only $\mathcal{O}\left(n^{-1/2}\right)$, because in (21), the $\mathcal{O}(1)$ terms vanish between the curly braces. Furthermore, comparison with the example below reveals that in this zero-means case, the density is approximated by that of a ratio of two correlated normals with matching first and second moments, which is correct to the order stated. The asymptotic distribution in the non-zero mean case is quite different: suppose that $\mu_Y \neq 0$ and let $\lambda \equiv \mu_X/\mu_Y$. It is a standard result (see, e.g. Fuller, 1990, Theorem 1.3.7) that $\sqrt{n}\mu_Y(R-\lambda) \to \mathrm{N}(0, \mathbf{c}_\lambda'\mathbf{\Sigma}\mathbf{c}_\lambda)$ in distribution. In approximation (23), the term in curly braces tends to unity as $n \to \infty$ for fixed r, so that the approximation will converge to that derived in Daniels (1954), and hence ultimately to a Gaussian. The case with $\mu_Y = 0, \mu_X \neq 0$ can be treated by considering R^{-1}. It can be verified longhand that the saddlepoint approximation to the density of R^{-1}, $\hat{g}_n^{(1)}(r)$, say, obeys the symmetry relation $\hat{g}_n^{(1)}(r) = \hat{f}_n^{(1)}(r^{-1})/r^2$.

Example 1 (Ratio of correlated normals). *Suppose X and Y are jointly Gaussian with respective means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation ρ. The density of $R = X/Y$ has been found in Fieller (1932); see also Hinkley (1969). The cgf of X and Y is $\mathbb{K}(s,t) = s\mu_X + t\mu_Y + (s^2\sigma_X^2 + 2st\rho\sigma_X\sigma_Y + t^2\sigma_Y^2)/2$.*

Applying Theorem 3 with $n = 1$, it is found that both saddlepoints are explicit in terms of the parameters and given by

$$\tilde{s}_0 = \frac{r\mu_Y - \mu_X}{a^2\sigma_X^2\sigma_Y^2}, \quad \hat{s} = \frac{\mu_Y\rho\sigma_X - \mu_X\sigma_Y}{\sigma_X^2\sigma_Y(1-\rho^2)}, \quad \hat{t} = \frac{\mu_X\rho\sigma_Y - \mu_Y\sigma_X}{\sigma_Y^2\sigma_X(1-\rho^2)}.$$

Defining

$$a \equiv \sqrt{\frac{r^2}{\sigma_X^2} - \frac{2r\rho}{\sigma_X\sigma_Y} + \frac{1}{\sigma_Y^2}} \quad \textit{and} \quad b \equiv \frac{r\mu_X}{\sigma_X^2} - \frac{\rho(\mu_X + r\mu_Y)}{\sigma_X\sigma_Y} + \frac{\mu_Y}{\sigma_Y^2},$$

the other relevant quantities are

$$\tilde{g}_0 = \frac{b}{a^3\sigma_X\sigma_Y}, \quad \tilde{w}_0 = \frac{r\mu_Y - \mu_X}{a\sigma_X\sigma_Y}, \quad \textit{and} \quad \hat{w} = -\frac{b}{\sqrt{(1-\rho^2)a}}.$$

Plugging in and rearranging, this is exactly the expression given in Fieller (1932) and Hinkley (1969); in other words, the saddlepoint approximation is exact in this case.

A final remark concerns the uniqueness of the approximation. By way of example, consider a Cauchy random variable; that is, take $n = 1$, $X = Z_1$, and $Y = Z_2$, where Z_1 and Z_2 are independent standard Gaussian. As shown above, the saddlepoint approximation is exact in this case, so that $\hat{f}_n^{(1)}(r) = (\pi(1+r^2))^{-1}$. Alternatively, one might take $X = Z_1Z_2$ and $Y = Z_2^2$ with joint cgf $\mathbb{K}(s,t) = -1/2\log(1 - 2s - t^2)$, so that $R = X/Y = Z_1/Z_2$ as before. Now Y is almost surely positive; thus X and Y form a definite pair and the approximation reduces to $\hat{f}_n^{(1)}(r) = \phi(\tilde{w}_0)\tilde{g}_0 = (\sqrt{2\pi}(1+r^2))^{-1}$, which is clearly different (albeit in agreement after normalization). Apparently, different representations for the ratio can result in different approximations. It would appear that in general, a choice had to be made as to which representation to use. It is however quite rare that the cumulant generating functions of both Z_2 and Z_2^2 are available, let alone tractable. In fact, the latter only exists if the tails of Z_2 are as least as thin as those of a Gaussian, a requirement that fails even for the Exponential distribution.

3.2. Cdf Approximation

While an approximation to the distribution function can always be obtained by integrating (23), the ability to approximate the cdf of R directly is undoubtedly of value. Equation (11) is a convenient starting point. Expressed in terms of distribution functions, it states that

$$F_R^n(r) = F_{\bar{W}}^n(0) + F_{\bar{Y}}^n(0) - 2F_{\bar{W},\bar{Y}}^n(0,0).$$

The inversion formulae for $F_{\bar{W},\bar{Y}}^n(\bar{w},\bar{y})$ and $F_{\bar{W}}^n(\bar{w})$ are repeated here for convenience:

$$F_{\bar{W},\bar{Y}}^n(0,0) = \left(\frac{1}{2\pi \mathrm{i}}\right)^2 \int_{c_2-\mathrm{i}\infty}^{c_2+\mathrm{i}\infty} \int_{c_1-\mathrm{i}\infty}^{c_1+\mathrm{i}\infty} e^{n\mathbb{K}(s,t-rs)} \frac{\mathrm{d}s}{s}\frac{\mathrm{d}t}{t}, \tag{12}$$

$$F_{\bar{W}}^n(0) = \frac{\mathrm{i}}{2\pi} \int_{c_3-\infty}^{c_3+\infty} e^{n\mathbb{K}(s,-rs)} \frac{\mathrm{d}s}{s}, \tag{13}$$

and, analogously,

$$F_{\bar{Y}}^n(0) = \frac{\mathrm{i}}{2\pi} \int_{c_4-\infty}^{c_4+\infty} e^{n\mathbb{K}(0,t)} \frac{\mathrm{d}t}{t}, \quad (0,c_4) \in \mathcal{T}. \tag{25}$$

Each integral in (12), (13), and (25) will be treated separately. Ideally, one would expand each integral in a uniform asymptotic expansion. For the latter two (one-dimensional) integrals, this is a simple task. All that is required is an application of Lemma 3. Doing so leads to the cdf approximation of Lugannani and Rice (1980). For the two-dimensional integral in (12), matters are less straightforward. In order to appreciate the difficulties involved, compare (12) and the bivariate integral (15) whose uniform asymptotic expansion formed the basis for the pdf approximation (23). The essential difference is the presence of the pole in the inner integrand. Applying a standard Laplace approximation as in (17) would therefore result in an expansion which is nonuniform in r as the saddlepoint crosses the pole. Instead, the inner integral could be approximated by another application of Lemma 3; this is the approach taken by Wang (1990a) in deriving a saddlepoint approximation for bivariate distributions. Unfortunately, when applied to the present problem, the approximation contains a term $\mathbb{K}(0,\hat{t}+r\hat{s})$ (w_{u_0} in Wang's notation). Depending on the structure of the problem, $(0,\hat{t}+r\hat{s})$ may fall outside the convergence region $\mathcal{T}$

for some values of r, rendering the approximation invalid. Although not discussed by Wang, this problem occurs not only in the present context, but more generally in the approximation of a bivariate distribution function, the subject of his paper. Kolassa and Li (2010) develop an alternative to Wang's approximation which is also applicable in higher dimensional problems, but it suffers from the same deficiency (see also Li, 2009, in particular Eq. 3.2.3).

In order to avoid these problems, we approximate the integrals in (12), (13) and (25) using an expression due to Kolassa (2003). After correcting for a typo, Kolassa's result, which is essentially a multivariate version of the cdf approximation of Hauschildt (1969) and Robinson (1982), is as follows.

Theorem 4 (Kolassa, 2003). *Suppose the d-dimensional random vector $\mathbf{X}$ has a density and a joint cgf $\mathbb{K}(\mathbf{t}) \equiv \log \mathbb{E}[\exp\{\mathbf{t}'\mathbf{X}\}]$ with gradient $\mathbb{K}'(\mathbf{t})$, Hessian $\mathbb{K}''(\mathbf{t})$, and third order derivatives $\mathbb{K}_{ijk}(\mathbf{t}) \equiv \partial^3/(\partial t_i \partial t_j \partial t_k)\mathbb{K}(\mathbf{t})$, $i, j, k \in \{1, \ldots, d\}$. Choose a compact subset $\mathcal{C}$ of the range of $\mathbb{K}'(\mathbf{t})$. Let $\overline{\mathbf{X}}$ denote the mean of n independent copies of $\mathbf{X}$, and for fixed $\bar{\mathbf{x}} \in \mathcal{C}$, define the saddlepoint $\hat{\mathbf{t}}$ as the solution to $\mathbb{K}'(\hat{\mathbf{t}}) = \bar{\mathbf{x}}$. Then, provided that $\hat{\mathbf{t}} > \mathbf{0}$,*

$$\mathbb{P}[\overline{\mathbf{X}} > \bar{\mathbf{x}}] = e^{n(\hat{\mathbb{K}} - \hat{\mathbf{t}}'\bar{\mathbf{x}})} \Bigg\{ e^{\frac{n\hat{\mathbf{t}}'\hat{\mathbb{K}}''\hat{\mathbf{t}}}{2}} \Bigg[I(\mathbf{0}, n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) + \frac{n}{6} \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \mathbb{K}_{ijk}(\hat{\mathbf{t}}) I(\mathbf{e}_i + \mathbf{e}_j + \mathbf{e}_k, n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) \Bigg] + \mathcal{O}(n^{-1}) \Bigg\}. \quad (26)$$

Here, $\widehat{\mathbb{K}}$ and $\widehat{\mathbb{K}}''$ denote the cgf and its Hessian evaluated at $\hat{\mathbf{t}}$, $\mathbf{e}_j$ is a d-vector with all components zero except for a 1 at position j, and, for $\mathbf{\Sigma}$ a positive definite matrix and $\mathbf{m} = [m_1, \ldots, m_d]$,

$$I(\mathbf{m}, \mathbf{\Sigma}, \hat{\mathbf{t}}) \equiv \frac{1}{(2\pi \mathrm{i})^d} \int_{\hat{\mathbf{t}} - \mathrm{i}\infty}^{\hat{\mathbf{t}} + \mathrm{i}\infty} e^{\frac{\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}{2} - \mathbf{t}'\mathbf{\Sigma}\hat{\mathbf{t}}} \prod_{j=1}^{d} \frac{(t_j - \hat{t}_j)^{m_j}}{t_j} \mathrm{d}\mathbf{t}. \quad (27)$$

Applying Theorem 4 requires a means of evaluating the function I. Kolassa provides a recursive algorithm for this purpose, which expresses I in terms of the multivariate normal distribution and its derivatives. Recently, Broda and Kan (2013) have obtained a simpler recursion, which is also sufficiently general to allow evaluation of the terms required when (26) is expanded to higher order. Appendix B presents the explicit expressions for the relevant cases when $d = 1$ and $d = 2$. Kolassa defines the function I only for $\hat{\mathbf{t}} > 0$; when some elements of $\hat{\mathbf{t}}$ are zero, it can be defined as the appropriate limit. For our purposes, it will prove convenient to also allow $\hat{t}_j < 0$ for some or all j. Let $\mathbf{D} = \mathrm{diag}(\{d_j\})$, where $d_j = 1$ if $\hat{t}_j \geq 0$ and $d_j = -1$ if $\hat{t}_j < 0$. Then the following relationship holds.

$$I(\mathbf{m}, \mathbf{\Sigma}, \hat{\mathbf{t}}) = I(\mathbf{m}, \mathbf{D\Sigma D}, \mathbf{D}\hat{\mathbf{t}}) \prod_{j=1}^{d} d_j^{m_j+1}. \tag{28}$$

All elements of $\mathbf{D}\hat{\mathbf{t}}$ are nonnegative, so that the expressions in Appendix B apply.

If any component of $\hat{\mathbf{t}}$ is negative, say component j, then the approximation in Theorem 4 is not applicable. In that case, Kolassa recommends defining $\bar{\mathbf{Y}} = \mathbf{D}\bar{\mathbf{X}}$ and $\bar{\mathbf{y}} = \mathbf{D}\bar{\mathbf{x}}$ and exploiting the relationship $\mathbb{P}[\overline{\mathbf{X}} > \bar{\mathbf{x}}] = \mathbb{P}[\overline{\mathbf{X}}_{-j} > \bar{\mathbf{x}}_{-j}] - \mathbb{P}[\overline{\mathbf{Y}} > \bar{\mathbf{y}}]$. Here a negative subscript on a vector denotes omission of the indicated component. When applied to the present problem, the following result is obtained.

THEOREM 5. *Under the conditions of Theorem 3,* $F_R^n(r) = \hat{F}_n^{(1)}(r) + \mathcal{O}\left(n^{-1}\right)$, *where*

$$\begin{aligned}
\hat{F}_n^{(1)}(r) \equiv{}& \Big(H^*(\tilde{s}_0) - P_1\Big)\Big(1 - 2H^*(\hat{t}_r)\Big) + \Big(H^*(\check{t}_0) - P_2\Big) \\
&\times \Big(1 - 2H^*(\hat{s})\Big) + 2\Big(H^*(\hat{t}_r)H^*(\hat{s}) - P_3\Big), \\
P_1 \equiv{}& e^{n[\tilde{\kappa}_0^{(0)} + \tilde{s}_0^2\tilde{\kappa}_0^{(2)}/2]} \left[I(0, n\tilde{\kappa}_0^{(2)}, \tilde{s}_0) + n\tilde{\kappa}_0^{(3)} I(3, n\tilde{\kappa}_0^{(2)}, \tilde{s}_0)/6\right] \\
P_2 \equiv{}& e^{n[\check{\kappa}_0^{(0)} + \check{t}_0^2\check{\kappa}_0^{(2)}/2]} \left[I(0, n\check{\kappa}_0^{(2)}, \check{t}_0) + n\check{\kappa}_0^{(3)} I(3, n\check{\kappa}_0^{(2)}, \check{t}_0)/6\right], \\
P_3 \equiv{}& e^{n[\hat{\kappa}^{(0,0)} + \hat{\mathbf{t}}'\widehat{\mathbb{K}}''\hat{\mathbf{t}}/2]} \times \\
&\left[I(\mathbf{0}, n\hat{\mathcal{K}}, \hat{\mathbf{t}}_r) + \frac{n}{6}\sum_{j=0}^{3}\binom{3}{j}\hat{\kappa}^{(3-j,j)} I([3-j,\ j], n\hat{\mathcal{K}}, \hat{\mathbf{t}}_r)\right],
\end{aligned}$$

$H^*(s) \equiv \mathbf{1}_{s\geq 0}$, $\hat{\mathcal{K}} \equiv \{\hat{\kappa}^{(i,j)}\}$, $\check{t}_0$ *solves* $\mathbb{K}_2(0, \check{t}_0) = 0$, $\tilde{s}_0$ *is as in* (22), $(\hat{s}, \hat{t})$ *is as in* (20), $\hat{t}_r \equiv \hat{t} + r\hat{s}$, $\hat{\mathbf{t}}_r \equiv (\hat{s}, \hat{t}_r)$, $\hat{\mathbf{t}} = (\hat{s}, \hat{t})$, $\widehat{\mathbb{K}}'' \equiv \mathbb{K}''(\hat{s}, \hat{t})$, $\check{\kappa}_0^{(j)} \equiv \mathbb{K}_{2^j}(0, \check{t}_0)$,

$$\tilde{\kappa}_0^{(j)} \equiv \sum_{k=0}^{j}\binom{j}{k}(-r)^k \mathbb{K}_{1^{j-k}2^k}(\tilde{s}_0, -r\tilde{s}_0),$$

$$\hat{\kappa}^{(i,j)} \equiv \sum_{k=0}^{k} \binom{i+j}{k} (-r)^k \mathbb{K}_{1^{i+j-k}2^k}(\hat{s}, \hat{t}),$$

$\mathbb{K}_{1^i 2^j}(s,t) \equiv \partial^{i+j}\mathbb{K}(s,t)/\partial s^i \partial t^j$, *and explicit expressions for evaluating the function* I *are given in Appendix B.*

Proof. Appendix C. □

Similar to the pdf approximation, it can be verified that the cdf approximation in Theorem 5 is exact if X and Y are jointly Gaussian. Higher order approximations can be obtained by expanding the probabilities P_i in the theorem to higher order. This is easy for the univariate quantities P_1 and P_2, for which Hauschildt (1969, Eq. 1.4.25) provides the terms as far as n^{-2}. Higher order expansions for P_3 will be considered in a companion paper by the same authors (Broda and Kan, 2013).

4. Examples

4.1. Ratio of Mixtures of Normals

Before we turn to our main application (which pertains to a discrete distribution), we demonstrate the accuracy of the saddlepoint approximations to the pdf and the cdf using a ratio of mixtures of Gaussian random variables. Besides being a natural generalization of Example 1, this is a convenient choice, because the density and distribution functions of this ratio can be evaluated analytically. This avoids having to numerically evaluate the double integral in (7) for comparison with the saddlepoint approximation.

Mixtures of normals form a very flexible family of distributions. We restrict ourselves to two-component mixtures with unit variances. Specifically, let X_i and Y_i be independently distributed with respective densities

$$f_X(x) = w_X \phi(x - \mu_{X,1}) + (1 - w_X)\phi(x - \mu_{X,2}) \quad \text{and}$$
$$f_Y(y) = w_Y \phi(y - \mu_{Y,1}) + (1 - w_Y)\phi(y - \mu_{Y,2}).$$

Let $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$. An n-fold convolution shows that the density of $\sqrt{n}\bar{X}$ is

$$f_{\sqrt{n}\bar{X}}(x) = \sum_{i=0}^{n} p_{X,i}\phi\left(x - \frac{i\mu_{X,1} + (n-i)\mu_{X,2}}{\sqrt{n}}\right),$$

where $p_{X,i} = \binom{n}{i} w_X^i (1-w_X)^{n-i}$. A similar expression holds for $\sqrt{n}\bar{Y}$. An application of (3) yields the pdf of $R = \bar{X}/\bar{Y}$ as

$$f_R^n(r) = \sum_{i=0}^{n}\sum_{j=0}^{n} p_{X,i}p_{Y,j}\, g\left(r; \frac{i\mu_{X,1} + (n-i)\mu_{X,2}}{\sqrt{n}}, \frac{i\mu_{Y,1} + (n-i)\mu_{Y,2}}{\sqrt{n}}\right),$$

where $p_{Y,j} = \binom{n}{j} w_Y^j (1-w_Y)^{n-j}$,

$$g(r;\mu_1,\mu_2) = \phi\left(\frac{r\mu_2 - \mu_1}{\sqrt{1+r^2}}\right)\frac{r\mu_1 + \mu_2}{(1+r^2)^{\frac{3}{2}}}\left\{1 - 2\left[\Phi(\hat{w}) + \frac{\phi(\hat{w})}{\hat{w}}\right]\right\},$$

and $\hat{w} = -(r\mu_1 + \mu_2)/\sqrt{1+r^2}$. Note that $g(r;\mu_1,\mu_2)$ is the pdf of Z_1/Z_2, a ratio of two independent Gaussians with respective means μ_1 and μ_2 and unit variances. In other words, R is distributed as a mixture of ratios of independent Gaussians. Similarly, the cdf of R is

$$F_R^n(r) = \sum_{i=0}^{n}\sum_{j=0}^{n} p_{X,i}p_{Y,j}\, G\left(r; \frac{i\mu_{X,1} + (n-i)\mu_{X,2}}{\sqrt{n}}, \frac{i\mu_{Y,1} + (n-i)\mu_{Y,2}}{\sqrt{n}}\right),$$

where

$$G(r;\mu_1,\mu_2) = \Phi\left(\frac{r\mu_2 - \mu_1}{\sqrt{1+r^2}}\right) + \Phi(-\mu_2) - 2\Phi_2\left(\frac{r\mu_2 - \mu_1}{\sqrt{1+r^2}}, -\mu_2; -\frac{r}{\sqrt{1+r^2}}\right)$$

is the cdf of Z_1/Z_2. Here $\Phi_2(\cdot,\cdot;\rho)$ denotes the cdf of a standard bivariate Gaussian with correlation ρ.

For illustration, let $\mu_{X,1} = -1$, $\mu_{X,2} = 4$, $\mu_{Y,1} = -4$, $\mu_{Y,2} = 14$, $w_X = 0.2$, and $w_Y = 0.8$. Figure 1 shows the exact pdf and cdf of R for $n = 1$, 5, and 20, together with their first and second order saddlepoint approximations. The second order approximation for the cdf requires expanding P_3 in Theorem 5 to second order. This is done in Broda and Kan (2013). For the choice of parameters under consideration, the pdf of R has several modes, making it a rather challenging target for approximation. This is particularly apparent for $n = 1$. As n increases, the approximation improves as expected. It is seen that the second order approximation offers little improvement in this example, for neither the pdf nor the cdf. There are, however, situations in which it does. One example is the distribution of a ratio of indefinite quadratic forms in normal random vectors. This is the subject of a separate paper by the same authors.

4.2. *The IV Estimator in a Just Identified Model*

Consider the just identified simultaneous equations model with one endogenous regressor,

$$\mathbf{y}_1 = \mathbf{y}_2\beta + \mathbf{X}\boldsymbol{\gamma} + \mathbf{u}, \tag{29}$$

$$\mathbf{y}_2 = \mathbf{z}_1\pi + \mathbf{X}\boldsymbol{\delta} + \mathbf{v}, \tag{30}$$

where $\mathbf{y}_1 \equiv (y_{1,1}, \ldots, y_{1,T})'$, $\mathbf{y}_2 \equiv (y_{2,1}, \ldots, y_{2,T})'$, $\mathbf{u} \equiv (u_1, \ldots, u_T)'$, $\mathbf{v} \equiv (v_1, \ldots, v_T)'$, $\mathbf{X}$ is a $T \times k$ matrix of exogenous regressors, the $T \times 1$ vector $\mathbf{z}_1$ represents a non-stochastic instrument, $[\mathbf{z}_1\ \mathbf{X}]$ has full column rank, and

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \left(\mathbf{0}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right).$$

Let $\mathbf{M_X} \equiv \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and define $\mathbf{z} \equiv \mathbf{M_X}\mathbf{z}_1$. Then the IV estimator for β can be written as

$$\hat{\beta} = \frac{\mathbf{z}'\mathbf{y}_1}{\mathbf{z}'\mathbf{y}_2}, \tag{31}$$

and the associated estimation error $\hat{B} \equiv \hat{\beta} - \beta$ is

$$\hat{B} = \frac{\mathbf{z}'\mathbf{u}}{\pi\mathbf{z}'\mathbf{z} + \mathbf{z}'\mathbf{v}}. \tag{32}$$

The distribution of the estimator is invariant with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, but depends on the value of π, or more precisely, on the concentration parameter, defined as

$$\mu^2 \equiv \frac{\pi^2}{\sigma_v^2}\mathbf{z}'\mathbf{z}.$$

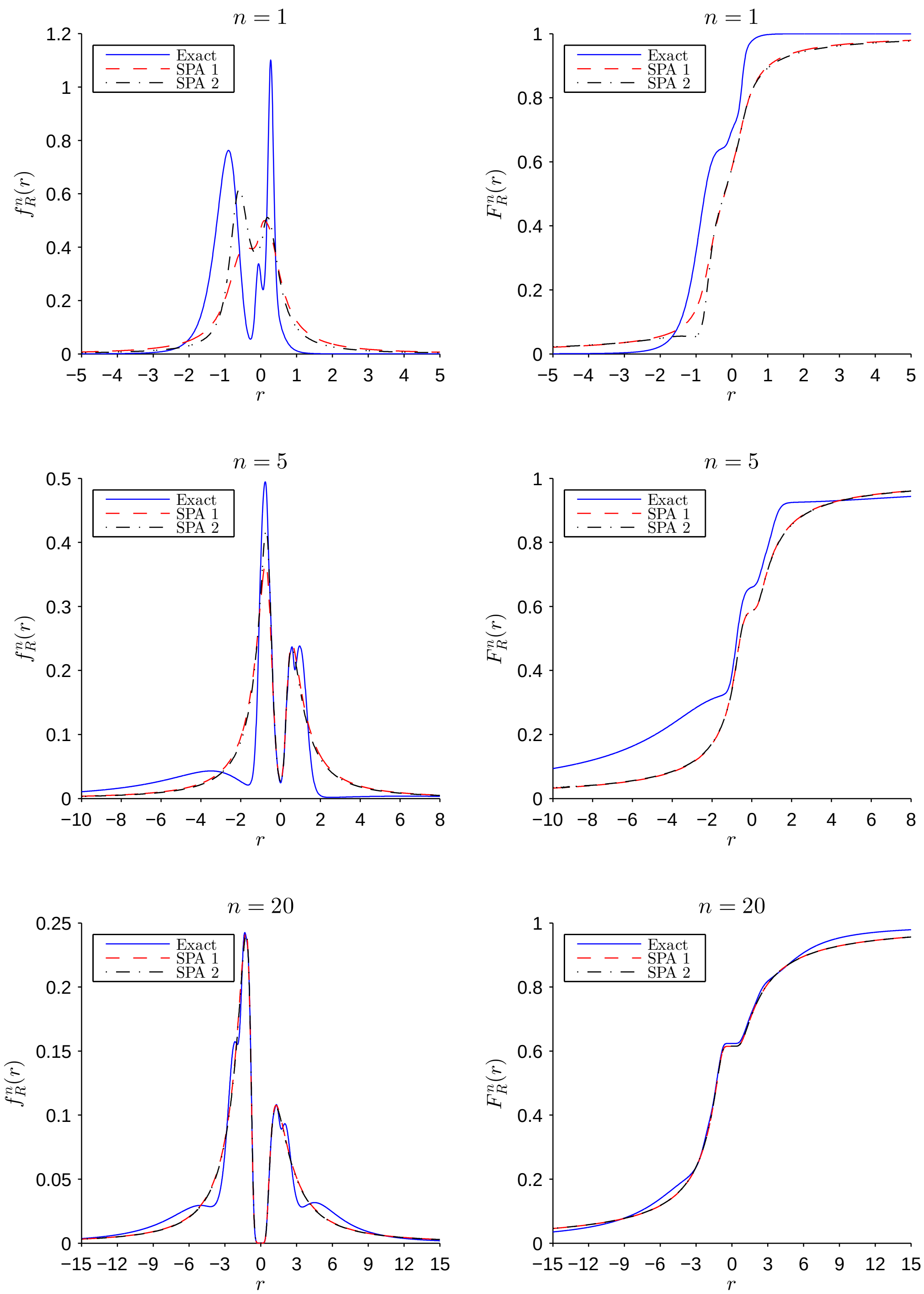

Fig. 1. Density and Distribution of a Ratio of Mixtures of Normals, with Parameters $\mu_{X,1} = -1$, $\mu_{X,2} = 4$, $\mu_{Y,1} = -4$, $\mu_{Y,2} = 14$, $w_X = 0.2$, and $w_Y = 0.8$.

The concentration parameter determines the strength of the instruments. In the strong instruments setting where π is $\mathcal{O}(1)$, the asymptotic distribution is Gaussian. If π is $\mathcal{O}(n^{-1/2})$, then the instruments are termed weak, and the asymptotic distribution is that of a ratio of normals, see Staiger and Stock (1997). Several authors have considered bootstrap inference in this setting; examples are Flores-Lagunes (2007), Zhan (2010), and Davidson and MacKinnon (2008, 2010). In general, saddlepoint approximations facilitate the (approximate) computation of bootstrap distributions without the need for computationally expensive simulations. This has prompted a number of authors to consider their application to such problems; early examples include Davison and Hinkley (1988), Daniels and Young (1991), DiCiccio et al. (1994), and Jing et al. (1994). The results of the present paper extend the applicability of the saddlepoint method to bootstrapping in the simultaneous equation model. The fact that the cdf approximation is a continuous function of the parameter of interest make it attractive for constructing confidence intervals by inverting a sequence of tests.

Davidson and MacKinnon (2010) discuss several bootstrap schemes. We consider what they refer to as the wild restricted efficient residual (WRE) bootstrap, which they found to perform most favorably. In it, a bootstrap replication is obtained from the resampled residuals $(\hat{\mathbf{u}} \odot \mathbf{r}, \hat{\mathbf{v}} \odot \mathbf{r})$, where $\mathbf{r}$ is a $T \times 1$ vector of i.i.d. Rademacher random variables and $\odot$ denotes elementwise multiplication. Here, $\hat{u}_t$ are the residuals from the structural equation (29) under the null (that is, imposing $\beta = \beta_0$), and $\hat{v}_t$ are the OLS residuals from the first stage regression (30). This sampling scheme is designed to replicate any possible heteroscedasticity. Davidson and MacKinnon scale $\hat{u}_t$ by a factor $\sqrt{T/(T-k)}$ and $\hat{v}_t$ by $\sqrt{T/(T-k-1)}$ to match the OLS variance estimates. In jointly sampling (u_t, v_t), any dependence between the errors in the structural and reduced form equations, and hence the endogeneity, is replicated. In the weak instrument setting, π, which enters the (bootstrap) distribution of $\hat{\beta}$, cannot be estimated consistently. Davidson and MacKinnon attempt to mitigate this by using an estimator that is asymptotically equivalent to three stage least squares applied to the system (29) and (30). This estimator is more efficient than the OLS estimator from the first stage regression (30).

Technically, the bootstrap distribution is discrete, and a continuity-corrected version of the saddlepoint approximation would appear to be warranted, as discussed in Daniels (1987) and Butler (2007, Sec. 1.2.3). We do not pursue this here because the improvements afforded by such a modification are typically small in bootstrap applications; see Davison and Hinkley (1988, Sec. 8), Wang (1990b, Sec. 2.2), and DiCiccio et al. (1994, p. 285). Unlike Davidson and MacKinnon, we refrain from studentization and resample the estimator directly. We require the joint cgf of $\mathbf{z}'\mathbf{u}$ and $\pi\mathbf{z}'\mathbf{z} + \mathbf{z}'\mathbf{v}$, given by

$$\mathbb{K}(s,t) = t\pi\mathbf{z}'\mathbf{z} - T\log(2) + \sum_{j=1}^{T} z_j(s\hat{u}_j + t\hat{v}_j) + \log\left(1 + e^{-2z_j(s\hat{u}_j + t\hat{v}_j)}\right).$$

A numerical example will exemplify the virtues of the saddlepoint approximation. We fix the sample size at $T = 15$ and set $\pi = 1/2$ and $\beta = \beta_0 = 0$. The small sample size allows us to evaluate the exact bootstrap cdf by enumerating all 2^T possible outcomes (recall that we have derived (9) under the assumption of absolute continuity, whereas the bootstrap distribution is discrete). We use the notation T rather than n because the summands in the numerator and denominator of our statistic are not i.i.d. Instead, we apply the saddlepoint approximation formally with $n = 1$. We nevertheless expect the approximation to improve as T grows and the joint distribution of numerator and denominator converges to a Gaussian. We draw the instrument from a uniform distribution and generate the structural innovations u_t from the Gaussian GARCH process $\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2$ with $\omega = 0.01$, $\alpha = 0.059$, and $\beta = 0.94$. The reduced form error is generated as $v_t = \rho u_t + \sqrt{1-\rho^2}\varepsilon_t$, where ε_t is generated by an independent GARCH process with the same parameters. We include a constant, a time trend, and the indicator $\mathbf{1}_{t>T/2}$ as exogenous regressors.

The bootstrap distribution is largely determined by $\hat{\pi}$. In the data set we chose for illustration, it is estimated at $\hat{\pi} = 0.4$. The concentration parameter is estimated as 0.49, which corresponds to rather weak instruments. Nevertheless, Figure 2 shows that the saddlepoint approximation tracks the exact bootstrap distribution accurately, particularly in the tails. Also depicted is the weak instrument asymptotic distribution, with nuisance parameters replaced by estimates. In this particular example, it differs considerably from the bootstrap distribution. We remark that for some data sets, the numerator and denominator of (32) can form a definite pair, but they do not for the one under scrutiny here.

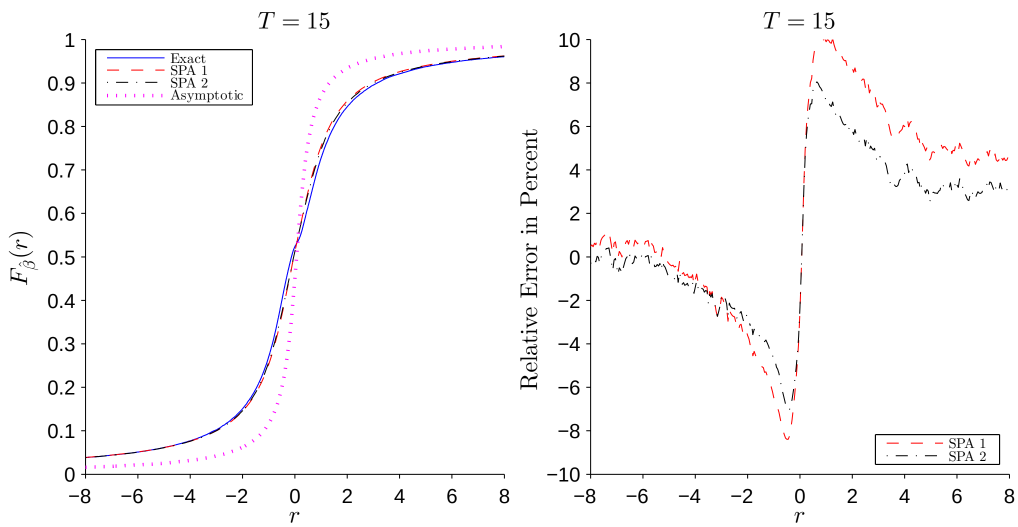

Fig. 2. Wild bootstrap distribution of the IV estimator for the sample described in the text. The right panel shows the relative error $(\hat{F} - F)/\min(F, 1 - F)$ in percent.

It is interesting to note that the IV estimator in the system (29) and (30) can equivalently be written as

$$\hat\beta = \frac{\mathbf{y}_2'\mathbf{P_z}\mathbf{y}_1}{\mathbf{y}_2'\mathbf{P_z}\mathbf{y}_2},$$

where $\mathbf{P_z} = \mathbf{z}'(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}$. This renders the associated estimation error as

$$\hat{B} = \frac{\pi\mathbf{z}'\mathbf{u} + \mathbf{v}'\mathbf{P_z}\mathbf{u}}{\pi^2\mathbf{z}'\mathbf{z} + 2\pi\mathbf{z}'\mathbf{v} + \mathbf{v}'\mathbf{P_z}\mathbf{v}}. \tag{33}$$

The matrix $\mathbf{P}_z$ is positive semidefinite. Therefore the denominator in (33) is almost surely positive, and the bootstrap distribution could be approximated by the standard result of Daniels (1954) if the joint bootstrap cgf of $\pi\mathbf{z}'\mathbf{u} + \mathbf{v}'\mathbf{P_z u}$ and $\pi\mathbf{z}'\mathbf{z}^2 + 2\pi\mathbf{z}'\mathbf{v} + \mathbf{v}'\mathbf{P_z v}$ were tractable. Unfortunately this is not the case. To see this, consider $\mathbb{E}[e^{\mathbf{v}'\mathbf{P_z v}}]$. Unlike in the Gaussian case, one cannot use the spectral theorem to reduce $\mathbf{v}'\mathbf{P_z v}$ to a sum of independent random variables. Consequently, computing $\mathbb{E}[e^{\mathbf{v}'\mathbf{P_z v}}]$ requires enumerating all 2^T possible realizations for $\mathbf{v}$, which becomes infeasible quickly (and renders the use of the approximation moot).

5. Conclusion

Ratios of random variables play a vital role in statistics and econometrics. The results of this paper facilitate the evaluation of their density and distribution functions, even if both numerator and denominator are supported on the entire real line. An important instance of such a random variable is the IV estimator in a just identified system with one endogenous regressor. We have considered its wild bootstrap distribution and demonstrated that the saddlepoint approximation is able to reproduce it accurately. It will be interesting to investigate the size and power properties of the procedure, as compared to resampling the studentized statistic by simulation. Important extensions of the results of the present paper concern the case of lattice variables, and the possibility of expanding P_3 in Theorem 5 to higher order. The latter is the subject of another paper by the same authors. Matlab code for evaluating the expressions in the paper is available from the authors.

Acknowledgements

The authors thank Ron Butler and John Kolassa for their helpful comments. The second author gratefully acknowledges financial support from National Bank Financial of Canada and the Social Sciences and Humanities Research Council of Canada, as well as travel support from Tinbergen Institute for his visit to Amsterdam during which this work was completed.

A. Higher Order Terms for the Pdf Approximation

This appendix presents higher order approximations for the pdf. Let $\tilde{s} = \tilde{s}(t)$ be the inner saddlepoint, that is, the solution to the equation

$$\mathbb{K}_1(\tilde{s}, t - r\tilde{s}) = r\mathbb{K}_2(\tilde{s}, t - r\tilde{s}).$$

Applying a standard Laplace approximation to the inner integral in (15) yields

$$\frac{n}{2\pi \mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} e^{n\mathbb{K}(s,t-rs)} \mathbb{K}_2(s, t - rs) \mathrm{d}s = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{nh(t)} \left[\sum_{j=0}^{m-1} \frac{g_j(t)}{n^j} + \mathcal{O}(n^{-m}) \right], \tag{34}$$

where $h(t) \equiv \mathbb{K}(\tilde{s}, t - r\tilde{s})$. An explicit expression for $g_j(t)$ can be obtained by using Eq. (103) of Rice (1968). It is given by

$$g_j(t) = \sum_{k=0}^{2j} \frac{\tilde{J}_k(t)}{k! \tilde{h}_2(t)^{\frac{k+1}{2}}} \tilde{a}_{j,2j-k}(t), \tag{35}$$

where

$$\tilde{h}_k(t) = \left. \frac{\partial^k \mathbb{K}(s, t - rs)}{\partial s^k} \right|_{s=\tilde{s}} = \sum_{j=0}^{k} \binom{k}{j} (-r)^j \mathbb{K}_{1^{k-j}2^j}(\tilde{s}, t - r\tilde{s}),$$

$$\tilde{J}_k(t) = \left. \frac{\partial^k \mathbb{K}_2(s, t - rs)}{\partial s^k} \right|_{s=\tilde{s}} = \sum_{j=0}^{k} \binom{k}{j} (-r)^j \mathbb{K}_{1^{k-j}2^{j+1}}(\tilde{s}, t - r\tilde{s}),$$

$\mathbb{K}_{1^i 2^j}(s, t) \equiv \partial^{i+j}\mathbb{K}(s, t)/\partial s^i \partial t^j$, and the coefficients $\tilde{a}_{i,j}(t)$ satisfy

$$\tilde{a}_{i,j}(t) = \sum_{k=0}^{j} \tilde{d}_{k,j}(t)(-2)^{i+k} \left(\frac{1}{2}\right)_{i+k},$$

with $\tilde{d}_{i,j}(t)$ obtained from the recurrence relation $\tilde{d}_{0,0}(t) = 1$, $\tilde{d}_{0,j}(t) = 0$, $j \geq 1$, and

$$\tilde{d}_{i,j}(t) = \frac{1}{i} \sum_{k=1}^{j-i+1} \frac{\tilde{h}_{k+2}(t)}{(k+2)! \tilde{h}_2(t)^{\frac{k+2}{2}}} \tilde{d}_{i-1,j-k}(t), \ j \geq i \geq 1.$$

Using (34), one has

$$I_2 = \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{1}{2\pi \mathrm{i}} \int_{c_2 - \mathrm{i}\infty}^{c_2 + \mathrm{i}\infty} e^{nh(t)} \left[\sum_{j=0}^{m-1} \frac{g_j(t)}{n^j} + \mathcal{O}(n^{-m}) \right] \frac{\mathrm{d}t}{t}.$$

Denote a typical term in the integral as

$$I_{2,j} \equiv \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{1}{2\pi \mathrm{i}} \int_{c_2 - \mathrm{i}\infty}^{c_2 + \mathrm{i}\infty} e^{nh(t)} g_j(t) \frac{\mathrm{d}t}{t}.$$

Using the result from Bleistein (1966) and Rice (1968, Appendix F, by setting $\lambda = 0$), $I_{2,j}$ can be approximated as

$$I_{2,j} = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{nh(0)} \left\{ \left[1_{c_2>0} - \Phi(\sqrt{n}\hat{w}) \right] g_j(0) + \frac{\phi(\sqrt{n}\hat{w})}{\sqrt{n}} \left[\sum_{k=0}^{m-1} \frac{p_{j,k}}{n^k} + \mathcal{O}(n^{-m}) \right] \right\},$$

where $\hat{t}_r \equiv \hat{t} + r\hat{s}$, $\hat{w} \equiv \mathrm{sgn}(\hat{t}_r)\sqrt{2[h(0) - h(\hat{t}_r)]}$, $\hat{u} \equiv \hat{t}_r \sqrt{h''(\hat{t}_r)}$,

$$p_{j,k} = \sum_{l=0}^{2k} \sum_{q=0}^{l} \frac{(-1)^l g_j^{(q)}(\hat{t}_r)}{q! h''(\hat{t}_r)^{\frac{q}{2}} \hat{u}^{2k+1-l}} a_{k,l-q} - \frac{g_j(0)(-2)^k \left(\frac{1}{2}\right)_k}{\hat{w}^{2k+1}}, \tag{36}$$

and the coefficients $a_{i,j}$ are given by

$$a_{i,j} = \sum_{k=0}^{j} d_{k,j} (-2)^{i+k} \left(\frac{1}{2}\right)_{i+k},$$

with $d_{i,j}$ obtained from the recurrence relation

$$d_{0,0} = 1, \qquad d_{0,j} = 0, \ j \geq 1, \qquad d_{i,j} = \frac{1}{i} \sum_{k=1}^{j-i+1} \theta_{k+2} d_{i-1,j-k}, \ j \geq i \geq 1.$$

Here $\theta_k = h^{(k)}(\hat{t}_r)/[k! h''(\hat{t}_r)^{\frac{k}{2}}]$. Details of the derivation of this formula are available upon request. Collecting the terms with like power of n, the mth order approximation for I_2 is

$$I_2 = \sqrt{n}\phi(\sqrt{n}\tilde{w}_0) \left\{ \left[1_{c_2>0} - \Phi(\sqrt{n}\hat{w}) \right] \sum_{j=0}^{m-1} \frac{g_j(0)}{n^j} + \frac{\phi(\sqrt{n}\hat{w})}{\sqrt{n}} \sum_{j=0}^{m-1} \frac{1}{n^j} \sum_{k=0}^{j} p_{j-k,k} + \mathcal{O}(n^{-m}) \right\}.$$

Similarly, the mth order approximation for I_1 is

$$\begin{aligned} I_1 &= \frac{n}{2\pi \mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} e^{n\mathbb{K}(s,-rs)} \mathbb{K}_2(s, -rs) \mathrm{d}s \\ &= \sqrt{n}\phi(\sqrt{n}\tilde{w}_0) \left[\sum_{j=0}^{m-1} \frac{g_j(0)}{n^j} + \mathcal{O}(n^{-m}) \right]. \end{aligned}$$

It follows that the mth order saddlepoint approximation for the pdf is

$$\hat{f}_n^{(m)}(r) = \sqrt{n}\phi(\sqrt{n}\tilde{w}_0)\left\{[1-2\Phi(\sqrt{n}\hat{w})]\sum_{j=0}^{m-1}\frac{g_j(0)}{n^j} + \frac{2\phi(\sqrt{n}\hat{w})}{\sqrt{n}}\sum_{j=0}^{m-1}\frac{A_j}{n^j}\right\},$$

where

$$\begin{aligned} A_j &= \sum_{k=0}^{j} p_{j-k,k} \\ &= \sum_{k=0}^{j}\sum_{l=0}^{2k}\sum_{q=0}^{l}\frac{(-1)^l g_{j-k}^{(q)}(\hat{t}_r)}{q!h''(\hat{t}_r)^{\frac{q}{2}}\hat{u}^{2k+1-l}}a_{k,l-q} - \sum_{k=0}^{j}\frac{g_{j-k}(0)(-2)^k\left(\frac{1}{2}\right)_k}{\hat{w}^{2k+1}} \\ &= \sum_{k=0}^{j}\sum_{l=0}^{2k}\frac{(-1)^l}{\hat{u}^{2k+1-l}}b_{j,k,l} - \sum_{k=0}^{j}\frac{g_{j-k}(0)(-2)^k\left(\frac{1}{2}\right)_k}{\hat{w}^{2k+1}}, \end{aligned}$$

and

$$b_{j,k,l} = \sum_{q=0}^{l}\frac{g_{j-k}^{(q)}(\hat{t}_r)a_{k,l-q}}{q!h''(\hat{t}_r)^{\frac{q}{2}}}.$$

The above expression for A_j is undefined when $\hat{t}_r = 0$ (i.e., when $\hat{t} = -r\hat{s}$). In order to obtain its limit as $\hat{t}_r \to 0$, expand h in a Taylor series about $\hat{t}_r$. This yields

$$h(0) - h(\hat{t}_r) = -\hat{t}_r h'(\hat{t}_r) + \frac{\hat{t}_r^2}{2}h''(\hat{t}_r) - \frac{\hat{t}_r^3}{3!}h'''(\hat{t}_r) + \frac{\hat{t}_r^4}{4!}h^{(4)}(\hat{t}_r) - \cdots .$$

Using the fact that $h'(\hat{t}_r) = 0$, one has

$$\begin{aligned} \hat{w}^2 &= 2[\mathbb{K}(\tilde{s}_0, -r\tilde{s}_0) - \mathbb{K}(\hat{s},\hat{t})] \\ &= 2[h(0) - h(\hat{t}_r)] \\ &= 2\left[\frac{\hat{t}_r^2}{2!}h''(\hat{t}_r) - \frac{\hat{t}_r^3}{3!}h'''(\hat{t}_r) + \frac{\hat{t}_r^4}{4!}h^{(4)}(\hat{t}_r) - \cdots\right] \\ &= 2\sum_{j=2}^{\infty}(-1)^j\theta_j\hat{u}^j. \end{aligned}$$

Letting

$$B_{j,k} \equiv \frac{g_j(0)(-2)^k\left(\frac{1}{2}\right)_k}{\hat{w}^{2k+1}},$$

and using (36),

$$B_{j,k} = \sum_{l=0}^{2k}\sum_{q=0}^{l}\frac{(-1)^l g_j^{(q)}(\hat{t}_r)}{q!h''(\hat{t}_r)^{\frac{q}{2}}\hat{u}^{2k+1-l}}a_{k,l-q} - p_{j,k},$$

$$B_{j,k+1} = \sum_{l=0}^{2k+2}\sum_{q=0}^{l} \frac{(-1)^l g_j^{(q)}(\hat{t}_r)}{q!h''(\hat{t}_r)^{\frac{q}{2}}\hat{u}^{2k+3-l}} a_{k+1,l-q} - p_{j,k+1}.$$

Using the fact that $\hat{w}^2 B_{j,k+1} = -(2k+1)B_{j,k}$, it follows that

$$\left[2\sum_{j=2}^{\infty}(-1)^j\theta_j\hat{w}^j\right]\left[\sum_{l=0}^{2k+2}\sum_{q=0}^{l}\frac{(-1)^l g_j^{(q)}(\hat{t}_r)}{q!h''(\hat{t}_r)^{\frac{q}{2}}\hat{u}^{2k+3-l}}a_{k+1,l-q} - p_{j,k+1}\right]$$
$$= -(2k+1)\left[\sum_{l=0}^{2k}\sum_{q=0}^{l}\frac{(-1)^r g_j^{(q)}(\hat{t}_r)}{q!h''(\hat{t}_r)^{\frac{q}{2}}\hat{u}^{2k+1-l}}a_{k,l-q} - p_{j,k}\right].$$

Comparing the constant term on both sides, $p_{j,k}$ can be expressed as

$$p_{j,k} = -\frac{2}{2k+1}\sum_{l=0}^{2k+1}\theta_{2k+3-l}\sum_{q=0}^{l}\frac{g_j^{(q)}(\hat{t}_r)a_{k+1,l-q}}{q!h''(\hat{t}_r)^{\frac{q}{2}}}.$$

Taking the limit,

$$\lim_{\hat{t}_r\to 0} p_{j,k} = -\frac{2}{2k+1}\sum_{l=0}^{2k+1}\bar{\theta}_{2k+3-l}\sum_{q=0}^{l}\frac{g_j^{(q)}(0)\bar{a}_{k+1,l-q}}{q!h''(0)^{\frac{q}{2}}},$$

where $\bar{\theta}_j$ and $\bar{a}_{i,j}$ are the values of θ_j and $a_{i,j}$ evaluated at $\hat{t}_r = 0$. It follows that at $r = r^* \equiv -\hat{t}/\hat{s}$ where $\hat{t}_r = 0$,

$$\hat{f}_n^{(m)}(r^*) = \sqrt{\frac{2}{\pi}}\phi(\sqrt{n}\tilde{w}_0)\sum_{j=0}^{m-1}\frac{\bar{A}_j}{n^j},$$

where

$$\begin{aligned}
\bar{A}_j &\equiv \lim_{\hat{t}_r\to 0} A_j \\
&= \lim_{\hat{t}_r\to 0}\sum_{k=0}^{j} p_{j-k,k} \\
&= -\sum_{k=0}^{j}\frac{2}{2k+1}\sum_{l=0}^{2k+1}\bar{\theta}_{2k+3-l}\sum_{q=0}^{l}\frac{g_{j-k}^{(q)}(0)\bar{a}_{k+1,l-q}}{q!h''(0)^{\frac{q}{2}}} \\
&= -\sum_{k=0}^{j}\frac{2}{2k+1}\sum_{l=0}^{2k+1}\bar{\theta}_{2k+3-l}\bar{b}_{j+1,k+1,l},
\end{aligned}$$

and $\bar{b}_{j+1,k+1,l}$ is the value of $b_{j+1,k+1,l}$ evaluated at $\hat{t}_r = 0$. The following subsections provide explicit expressions obtained by specializing these results to the first, second, and third order cases.

A.1. First Order Approximation

For $m = 1$ and $\hat{t}_r \neq 0$, using the fact that $g_0(\hat{t}_r) = 0$ yields

$$A_0 = p_{0,0} = \frac{g_0(\hat{t}_r)}{\hat{u}} - \frac{g_0(0)}{\hat{w}} = -\frac{g_0(0)}{\hat{w}}.$$

It follows that when $r \neq r^*$ (so that $\hat{t}_r \neq 0$),

$$\hat{f}_n^{(1)}(r) = \sqrt{n}\phi(\sqrt{n}\tilde{w}_0)g_0(0)\left[1 - 2\Phi(\sqrt{n}\hat{w}) - \frac{2\phi(\sqrt{n}\hat{w})}{\sqrt{n}\hat{w}}\right],$$

which yields the expression in Theorem 3. Regarding the limit at $\hat{t}_r = 0$,

$$\bar{A}_0 = -\bar{\theta}_3 g_0(0) + \frac{g_0'(0)}{h''(0)^{\frac{1}{2}}} = \frac{g_0'(0)}{h''(0)^{\frac{1}{2}}} = \frac{h''(0)^{\frac{1}{2}}}{\tilde{h}_2(0)^{\frac{1}{2}}} = \frac{|\mathbb{K}''(\hat{s},\hat{t})|^{\frac{1}{2}}}{\mathbf{c}_{r^*}'\mathbb{K}''(\hat{s},\hat{t})\mathbf{c}_{r^*}}.$$

It follows that at $r = r^*$,

$$\hat{f}_n^{(1)}(r^*) = \sqrt{\frac{2}{\pi}}\phi(\sqrt{n}\tilde{w}_0)\frac{|\mathbb{K}''(\hat{s},\hat{t})|^{\frac{1}{2}}}{\mathbf{c}_{r^*}'\mathbb{K}''(\hat{s},\hat{t})\mathbf{c}_{r^*}},$$

which proves (24).

A.2. Second Order Approximation

For $m = 2$ and $\hat{t}_r \neq 0$, we have

$$\begin{aligned}\hat{f}_n^{(2)}(r) = \hat{f}_n^{(1)}(r) &+ \frac{\phi(\sqrt{n}\tilde{w}_0)g_1(0)}{\sqrt{n}}\left[1 - 2\Phi(\sqrt{n}\hat{w})\right] \\ &+ \frac{2\phi(\sqrt{n}\tilde{w}_0)\phi(\sqrt{n}\hat{w})}{n}A_1,\end{aligned} \tag{37}$$

where

$$\begin{aligned}A_1 &= p_{0,1} + p_{1,0} \\ &= \frac{g_1(\hat{t}_r)}{\hat{u}} - \frac{g_0''(\hat{t}_r)}{2h''(\hat{t}_r)\hat{u}} + \frac{3\theta_3 g_0'(\hat{t}_r)}{h''(\hat{t}_r)^{\frac{1}{2}}\hat{u}} + \frac{g_0'(\hat{t}_r)}{h''(\hat{t}_r)^{\frac{1}{2}}\hat{u}^2} + \frac{g_0(0)}{\hat{w}^3} - \frac{g_1(0)}{\hat{w}}.\end{aligned}$$

Writing $g_0(t) \equiv h'(t)/\tilde{h}_2(t)^{\frac{1}{2}}$, it is easy to see that

$$g_0'(\hat{t}_r) = \frac{h''(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} \quad \text{and}$$

$$g_0''(\hat{t}_r) = \frac{h'''(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} - \frac{h''(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} = \frac{6\theta_3 h''(\hat{t}_r)^{\frac{3}{2}}}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} - \frac{h''(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}}.$$

Using these expressions, one has that

$$\frac{g_0''(\hat{t}_r)}{2h''(\hat{t}_r)} - \frac{3\theta_3 g_0'(\hat{t}_r)}{h''(\hat{t}_r)^{\frac{1}{2}}} = -\frac{\tilde{h}_2'(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}}. \tag{38}$$

Using (35) and the fact that $\tilde{J}_0(t) = h'(t)$, $g_1(t)$ can be written as

$$\begin{aligned} g_1(t) &= \left[\frac{\tilde{h}_4(t)}{8\tilde{h}_2(t)^2} - \frac{5\tilde{h}_3(t)^2}{24\tilde{h}_2(t)^3}\right]\frac{h'(t)}{\tilde{h}_2(t)^{\frac{1}{2}}} + \frac{1}{2\tilde{h}_2(t)^{\frac{3}{2}}}\left[\frac{\tilde{h}_3(t)\tilde{J}_1(t)}{\tilde{h}_2(t)} - \tilde{J}_2(t)\right] \\ &= \left[\frac{\tilde{h}_4(t)}{8\tilde{h}_2(t)^2} - \frac{5\tilde{h}_3(t)^2}{24\tilde{h}_2(t)^3}\right]\frac{h'(t)}{\tilde{h}_2(t)^{\frac{1}{2}}} - \frac{\tilde{h}_2'(t)}{2\tilde{h}_2(t)^{\frac{3}{2}}}, \end{aligned} \tag{39}$$

where the last equality follows from the identity

$$\tilde{h}_k'(t) = \tilde{h}_{k+1}(t)\tilde{s}'(t) + \tilde{J}_k(t) = -\frac{\tilde{h}_{k+1}(t)\tilde{J}_1(t)}{\tilde{h}_2(t)} + \tilde{J}_k(t), \quad k \geq 1. \tag{40}$$

As $h'(\hat{t}_r) = 0$, $g_1(\hat{t}_r)$ reduces to

$$g_1(\hat{t}_r) = -\frac{\tilde{h}_2'(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}}. \tag{41}$$

Using (38) and (41), A_1 can be simplified to

$$A_1 = \frac{g_0'(\hat{t}_r)}{h''(\hat{t}_r)^{\frac{1}{2}}\hat{u}^2} + \frac{g_0(0)}{\hat{w}^3} - \frac{g_1(0)}{\hat{w}} = \frac{1}{\hat{t}_r^2|\mathbb{K}''(\hat{s},\hat{t})|^{\frac{1}{2}}} + \frac{g_0(0)}{\hat{w}^3} - \frac{g_1(0)}{\hat{w}}. \tag{42}$$

Regarding the limit at $\hat{t}_r = 0$,

$$\hat{f}_n^{(2)}(r^*) = \sqrt{\frac{2}{\pi}}\phi(\sqrt{n}\tilde{w}_0)\left[\frac{|\mathbb{K}(\hat{s},\hat{t})|^{\frac{1}{2}}}{\mathbf{c}_{r^*}'\mathbb{K}''(\hat{s},\hat{t})\mathbf{c}_{r^*}} + \frac{\bar{A}_1}{n}\right], \tag{43}$$

where

$$\bar{A}_1 = \left(3\bar{\theta}_4 - \frac{15\bar{\theta}_3^2}{2}\right)\frac{g_0'(0)}{h''(0)^{\frac{1}{2}}} + \frac{3\bar{\theta}_3 g_0''(0)}{2h''(0)} - \frac{g_0'''(0)}{6h''(0)^{\frac{3}{2}}} - \bar{\theta}_3 g_1(0) + \frac{g_1'(0)}{h''(0)^{\frac{1}{2}}}. \tag{44}$$

Evaluating $\bar{A}_1$ requires explicit expressions for $g'''(\hat{t}_r)$, $g_1'(\hat{t}_r)$, $h'''(\hat{t}_r)$ and $h^{(4)}(\hat{t}_r)$. It is straightforward to show that

$$\begin{aligned} g_0'''(t) = {} & \frac{h^{(4)}(t)}{\tilde{h}_2(t)^{\frac{1}{2}}} - \frac{3h'''(t)\tilde{h}_2'(t)}{2\tilde{h}_2(t)^{\frac{3}{2}}} + \frac{9h''(t)\tilde{h}_2'(t)^2}{4\tilde{h}_2(t)^{\frac{5}{2}}} - \frac{3h''(t)\tilde{h}_2''(t)}{2\tilde{h}_2(t)^{\frac{3}{2}}} \\ & + h'(t)\left[-\frac{15\tilde{h}_2'(t)^3}{8\tilde{h}_2(t)^{\frac{7}{2}}} + \frac{9\tilde{h}_2'(t)\tilde{h}_2''(t)}{4\tilde{h}_2(t)^{\frac{5}{2}}} - \frac{\tilde{h}_2'''(t)}{2\tilde{h}_2(t)^{\frac{3}{2}}}\right]. \end{aligned}$$

Thus, again using the fact that $h'(\hat{t}_r) = 0$,

$$g_0'''(\hat{t}_r) = \frac{h^{(4)}(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} - \frac{3h'''(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} + \frac{9h''(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)^2}{4\tilde{h}_2(\hat{t}_r)^{\frac{5}{2}}} - \frac{3h''(\hat{t}_r)\tilde{h}_2''(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}}$$
$$= \frac{24\theta_4 h''(\hat{t}_r)^2}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} - \frac{9\theta_3 h''(\hat{t}_r)^{\frac{3}{2}}\tilde{h}_2'(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} + \frac{9h''(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)^2}{4\tilde{h}_2(\hat{t}_r)^{\frac{5}{2}}} - \frac{3h''(\hat{t}_r)\tilde{h}_2''(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}},$$

where $\tilde{h}_2'(\hat{t}_r)$ and $\tilde{h}_2''(\hat{t}_r)$ are obtained from (40), and are given by

$$\tilde{h}_2'(\hat{t}_r) = \tilde{J}_2(\hat{t}_r) - \frac{\tilde{h}_3(\hat{t}_r)\tilde{J}_1(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)},$$
$$\tilde{h}_2''(\hat{t}_r) = \tilde{J}_2'(\hat{t}_r) - \frac{\tilde{h}_3'(\hat{t}_r)\tilde{J}_1(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)} - \frac{\tilde{h}_3(\hat{t}_r)\tilde{J}_1'(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)} + \frac{\tilde{h}_3(\hat{t}_r)\tilde{J}_1(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^2}.$$

$\tilde{h}_3'(\hat{t}_r)$ can be obtained from (40), and the general expression for $\tilde{J}_k'(t)$ is

$$\tilde{J}_k'(t) = -\frac{\tilde{J}_{k+1}(t)\tilde{J}_1(t)}{\tilde{h}_2(t)} + \sum_{j=0}^{k}\binom{k}{j}(-r)^j\mathbb{K}_{1^{k-j}2^{j+2}}(\tilde{s}, t - r\tilde{s}).$$

Differentiating (39) and using the fact that $h'(\hat{t}_r) = 0$ once more, it is found that

$$g_1'(\hat{t}_r) = \left[\frac{\tilde{h}_4(t)}{8\tilde{h}_2(\hat{t}_r)^2} - \frac{5\tilde{h}_3(\hat{t}_r)^2}{24\tilde{h}_2(\hat{t}_r)^3}\right]\frac{h''(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} - \frac{\tilde{h}_2''(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} + \frac{3\tilde{h}_2'(\hat{t}_r)^2}{4\tilde{h}_2(\hat{t}_r)^{\frac{5}{2}}}. \tag{45}$$

Regarding $h'''(t)$ and $h^{(4)}(t)$, using that $h''(t) = \mathbb{K}_{22}(\tilde{s}, t - r\tilde{s}) - [\tilde{J}_1(t)^2/\tilde{h}_2(t)]$ yields

$$h'''(t) = \mathbb{K}_{222}(\tilde{s}, t - r\tilde{s}) + [\mathbb{K}_{122}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{222}(\tilde{s}, t - r\tilde{s})]\tilde{s}'(t)$$
$$- \frac{2\tilde{J}_1(t)\tilde{J}_1'(t)}{\tilde{h}_2(t)} + \frac{\tilde{J}_1(t)^2\tilde{h}_2'(t)}{\tilde{h}_2(t)^2}$$
$$= \mathbb{K}_{222}(\tilde{s}, t - r\tilde{s}) - 3[\mathbb{K}_{122}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{222}(\tilde{s}, t - r\tilde{s})]\frac{\tilde{J}_1(t)}{\tilde{h}_2(t)}$$
$$+ \frac{3\tilde{J}_1(t)^2\tilde{J}_2(t)}{\tilde{h}_2(t)^2} - \frac{\tilde{J}_1(t)^3\tilde{h}_3(t)}{\tilde{h}_2(t)^3}$$

and

$$h^{(4)}(t) = \mathbb{K}_{2^4}(\tilde{s}, t - r\tilde{s}) - 4[\mathbb{K}_{12^3}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{2^4}(\tilde{s}, t - r\tilde{s})]\frac{\tilde{J}_1(t)}{\tilde{h}_2(t)}$$
$$+ 6[\mathbb{K}_{1^22^2}(\tilde{s}, t - r\tilde{s}) - 2r\mathbb{K}_{12^3}(\tilde{s}, t - r\tilde{s}) + r^2\mathbb{K}_{2^4}(\tilde{s}, t - r\tilde{s})]\frac{\tilde{J}_1(t)^2}{\tilde{h}_2(t)^2}$$

$$- \frac{3[\tilde{J}_1'(t)\tilde{h}_2(t) - \tilde{J}_1(t)\tilde{h}_2'(t)]^2}{\tilde{h}_2(t)^3} - \frac{4\tilde{J}_1(t)^3\tilde{J}_3(t)}{\tilde{h}_2(t)^3} + \frac{\tilde{J}_1(t)^4\tilde{h}_4(t)}{\tilde{h}_2(t)^4}.$$

Setting $\hat{t}_r = 0$ yields

$$\begin{aligned}\bar{A}_1 = &\left[\frac{3\bar{\theta}_3^2 - 2\bar{\theta}_4}{2} + \frac{\tilde{h}_4(0)}{8\tilde{h}_2(0)^2} - \frac{5\tilde{h}_3(0)^2}{24\tilde{h}_2(0)^3}\right] \frac{|\mathbb{K}(\hat{s},\hat{t})|^{\frac{1}{2}}}{\mathbf{c}_{r^*}'\mathbb{K}''(\hat{s},\hat{t})\mathbf{c}_{r^*}} + \frac{\bar{\theta}_3\tilde{h}_2'(0)}{2\tilde{h}_2(0)^{\frac{3}{2}}} \\ &- \frac{\tilde{h}_2''(0)}{4\tilde{h}_2(0)^{\frac{3}{2}}h''(0)^{\frac{1}{2}}} + \frac{3\tilde{h}_2'(0)^2}{8\tilde{h}_2(0)^{\frac{5}{2}}h''(0)^{\frac{1}{2}}}.\end{aligned} \tag{46}$$

A.3. Third Order Approximation

For $m = 3$ and $\hat{t}_r \neq 0$, we have

$$\begin{aligned}\hat{f}_n^{(3)}(r) = \hat{f}_n^{(2)}(r) &+ \frac{\phi(\sqrt{n}\tilde{w}_0)g_2(0)}{n^{\frac{3}{2}}}\left[1 - 2\Phi(\sqrt{n}\hat{w})\right] \\ &+ \frac{2\phi(\sqrt{n}\tilde{w}_0)\phi(\sqrt{n}\hat{w})}{n^2}A_2,\end{aligned} \tag{47}$$

where

$$\begin{aligned}A_2 &= p_{2,0} + p_{1,1} + p_{0,2} \\ &= \frac{g_2(\hat{t}_r)}{\hat{u}} - \frac{g_2(0)}{\hat{w}} + \left(\frac{a_{1,0}}{\hat{u}^3} - \frac{a_{1,1}}{\hat{u}^2} + \frac{a_{1,2}}{\hat{u}}\right) g_1(\hat{t}_r) \\ &\quad - \left(\frac{a_{1,0}}{\hat{u}^2} - \frac{a_{1,1}}{\hat{u}}\right)\frac{g_1'(\hat{t}_r)}{h''(\hat{t}_r)^{\frac{1}{2}}} + \frac{a_{1,0}}{\hat{u}}\frac{g_1''(\hat{t}_r)}{2h''(\hat{t}_r)} + \frac{g_1(0)}{\hat{w}^3} \\ &\quad - \left(\frac{a_{2,0}}{\hat{u}^4} - \frac{a_{2,1}}{\hat{u}^3} + \frac{a_{2,2}}{\hat{u}^2} - \frac{a_{2,3}}{\hat{u}}\right)\frac{g_0'(\hat{t}_r)}{h''(\hat{t}_r)^{\frac{1}{2}}} \\ &\quad + \left(\frac{a_{2,0}}{\hat{u}^3} - \frac{a_{2,1}}{\hat{u}^2} + \frac{a_{2,2}}{\hat{u}}\right)\frac{g_0''(\hat{t}_r)}{2h''(\hat{t}_r)} - \left(\frac{a_{2,0}}{\hat{u}^2} - \frac{a_{2,1}}{\hat{u}}\right)\frac{g_0'''(\hat{t}_r)}{6h''(\hat{t}_r)^{\frac{3}{2}}} \\ &\quad + \frac{a_{2,0}}{\hat{u}}\frac{g_0^{(4)}(\hat{t}_r)}{24h''(\hat{t}_r)^2} - \frac{3g_0(0)}{\hat{w}^5}.\end{aligned}$$

Using (40) and the identity

$$\sum_{k=1}^{2j}\frac{\tilde{h}_{k+1}(t)}{k!\tilde{h}_2(t)^{\frac{k+1}{2}}}\tilde{a}_{j,2j-k}(t) = 0,$$

$g_j(\hat{t}_r)$ can be written as

$$g_j(\hat{t}_r) = \sum_{k=2}^{2j}\frac{\tilde{h}_k'(\hat{t}_r)}{k!\tilde{h}_2(\hat{t}_r)^{\frac{k+1}{2}}}\tilde{a}_{j,2j-k}(\hat{t}_r).$$

Specifically,

$$\begin{aligned} g_2(\hat{t}_r) &= \frac{\tilde{a}_{2,2}(\hat{t}_r)\tilde{h}_2'(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} + \frac{\tilde{a}_{2,1}(\hat{t}_r)\tilde{h}_3'(\hat{t}_r)}{6\tilde{h}_2(\hat{t}_r)^2} + \frac{\tilde{a}_{2,0}(\hat{t}_r)\tilde{h}_4'(\hat{t}_r)}{24\tilde{h}_2(\hat{t}_r)^{\frac{5}{2}}} \\ &= \frac{35\tilde{h}_3(\hat{t}_r)^2\tilde{h}_2'(\hat{t}_r) - 20\tilde{h}_2(\hat{t}_r)h_3(\hat{t}_r)\tilde{h}_3'(\hat{t}_r)}{48\tilde{h}_2(\hat{t}_r)^{\frac{9}{2}}} \\ &\quad - \frac{15\tilde{h}_2(\hat{t}_r)\tilde{h}_4(\hat{t}_r)\tilde{h}_2'(\hat{t}_r) + 6\tilde{h}_2(\hat{t}_r)^2\tilde{h}_4'(\hat{t}_r)}{48\tilde{h}_2(\hat{t}_r)^{\frac{9}{2}}}. \end{aligned}$$

After some simplification, it can be verified that the coefficient associated with $1/\hat{u}$ in A_2 is zero. Therefore, A_2 can be written as

$$\begin{aligned} A_2 = \frac{1}{\hat{u}^2}&\Bigg(\left[3\theta_4 - \frac{15\theta_3^2}{2} + \frac{\tilde{h}_4(\hat{t}_r)}{8\tilde{h}_2(\hat{t}_r)^2} - \frac{5\tilde{h}_3(\hat{t}_r)^2}{24\tilde{h}_2(\hat{t}_r)^2}\right]\frac{h''(\hat{t}_r)^{\frac{1}{2}}}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} \\ &- \frac{3\theta_3\tilde{h}_2'(\hat{t}_r)}{2\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} + \frac{\tilde{h}_2''(\hat{t}_r)}{4\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}h''(\hat{t}_r)^{\frac{1}{2}}} - \frac{3\tilde{h}_2'(\hat{t}_r)^2}{8\tilde{h}_2(\hat{t}_r)^{\frac{5}{2}}h''(\hat{t}_r)^{\frac{1}{2}}}\Bigg) \\ &- \frac{1}{\hat{u}^3}\left[\frac{\tilde{h}_2'(\hat{t}_r)}{\tilde{h}_2(\hat{t}_r)^{\frac{3}{2}}} + \frac{6\theta_3 h''(\hat{t}_r)^{\frac{1}{2}}}{\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}}\right] - \frac{3\tilde{h}_2''(\hat{t}_r)^{\frac{1}{2}}}{\hat{u}^4\tilde{h}_2(\hat{t}_r)^{\frac{1}{2}}} \\ &- \frac{g_2(0)}{\hat{w}} + \frac{g_1(0)}{\hat{w}^3} - \frac{3g_0(0)}{\hat{w}^5}. \end{aligned} \tag{48}$$

Regarding the limit at $\hat{t}_r = 0$,

$$\hat{f}_n^{(3)}(r^*) = \sqrt{\frac{2}{\pi}}\phi(\sqrt{n}\tilde{w}_0)\left[\frac{|\mathbb{K}(\hat{s},\hat{t})|^{\frac{1}{2}}}{\mathbf{c}_{r^*}'\mathbb{K}''(\hat{s},\hat{t})\mathbf{c}_{r^*}} + \frac{\bar{A}_1}{n} + \frac{\bar{A}_2}{n^2}\right], \tag{49}$$

where

$$\begin{aligned} \bar{A}_2 = &\frac{g_0^{(5)}(0)}{40h''(0)^{\frac{5}{2}}} - \frac{5\bar{\theta}_3 g_0^{(4)}(0)}{8h''(0)^2} + \frac{5(7\bar{\theta}_3^2 - 2\bar{\theta}_4)g_0'''(0)}{4h''(0)^{\frac{3}{2}}} \\ &- \frac{15(21\bar{\theta}_3^3 - 14\bar{\theta}_3\bar{\theta}_4 + 2\bar{\theta}_5)g_0''(0)}{4h''(0)} \\ &+ \frac{15(231\bar{\theta}_3^4 - 252\bar{\theta}_3^2\bar{\theta}_4 + 28\bar{\theta}_4^2 + 56\bar{\theta}_3\bar{\theta}_5 - 8\bar{\theta}_6)g_0'(0)}{8h''(0)^{\frac{1}{2}}} \\ &+ \left(3\bar{\theta}_4 - \frac{15\bar{\theta}_3^2}{2}\right)\frac{g_1'(0)}{h''(0)^{\frac{1}{2}}} + \frac{3\bar{\theta}_3 g_1''(0)}{2h''(0)} - \frac{g_1'''(0)}{6h''(0)^{\frac{3}{2}}} - \bar{\theta}_3 g_2(0) + \frac{g_2'(0)}{h''(0)^{\frac{1}{2}}}. \end{aligned} \tag{50}$$

Evaluating $\bar{A}_2$ requires explicit expressions for $h^{(5)}(t)$ and $h^{(6)}(t)$. These expressions can be obtained by differentiating $h^{(4)}(t)$, but they are lengthy and hence omitted here.

B. Explicit Expressions for I in (27) for $d = 1$ and $d = 2$

The expressions below are valid if all elements of $\hat{\mathbf{t}}$ are nonnegative, which can always be achieved by an application of (28). For $d = 1$, $I(0, n\hat{\mathbb{K}}'', \hat{t}) = \Phi(-\hat{u}_n)$ and

$$I(3, n\hat{\mathbb{K}}'', \hat{t}) = (\hat{u}_n^2 - 1)\phi(\hat{u}_n) - \hat{u}_n^3 \Phi(-\hat{u}_n)(n\hat{\mathbb{K}}'')^{-3/2},$$

where $\hat{u}_n \equiv \hat{t}\sqrt{n\hat{\mathbb{K}}''}$. For $d = 2$, let $\rho \equiv \hat{\mathbb{K}}_{12}/\sqrt{\hat{\mathbb{K}}_{11}\hat{\mathbb{K}}_{22}}$, $\tilde{t}_1 \equiv \sqrt{n}\Big(\sqrt{\hat{\mathbb{K}}_{11}}\hat{t}_1 + \rho\sqrt{\hat{\mathbb{K}}_{22}}\hat{t}_2\Big)$, $\tilde{t}_2 \equiv \sqrt{n}\Big(\sqrt{\hat{\mathbb{K}}_{22}}\hat{t}_1 + \rho\sqrt{\hat{\mathbb{K}}_{11}}\hat{t}_1\Big)$, and define

$$\begin{aligned}
J_0 &\equiv \Phi_2(-\tilde{t}_1, -\tilde{t}_2; \rho), \\
J_1 &\equiv \frac{\phi(\tilde{t}_2)}{\sqrt{n\hat{\mathbb{K}}_{22}}}\Phi\left(-\frac{\tilde{t}_1 - \rho\tilde{t}_2}{\sqrt{1-\rho^2}}\right), \\
J_2 &\equiv \frac{\phi(\tilde{t}_1)}{\sqrt{n\hat{\mathbb{K}}_{11}}}\Phi\left(-\frac{\tilde{t}_2 - \rho\tilde{t}_1}{\sqrt{1-\rho^2}}\right), \\
J_3 &\equiv \frac{\phi_2(\tilde{t}_1, \tilde{t}_2; \rho)}{n\sqrt{\hat{\mathbb{K}}_{11}\hat{\mathbb{K}}_{22}}},
\end{aligned}$$

where $\phi_2(\cdot, \cdot; \rho)$ and $\Phi_2(\cdot, \cdot; \rho)$ denote, respectively, the pdf and cdf of a standard bivariate Gaussian with correlation ρ. The relevant functions can now be expressed as

$$\begin{aligned}
I([0,\ 0], n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) &= J_0, \\
I([3,\ 0], n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) &= \hat{t}_1^2(J_2 - \hat{t}_1 J_0) - (J_2 + n\hat{\mathbb{K}}_{12}k_1)(n\hat{\mathbb{K}}_{11})^{-1}, \\
I([2,\ 1], n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) &= k_1 + \hat{t}_1(J_1 - \hat{t}_2 J_0), \\
I([1,\ 2], n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) &= k_2 + \hat{t}_1(J_2 - \hat{t}_1 J_0), \\
I([0,\ 3], n\hat{\mathbb{K}}'', \hat{\mathbf{t}}) &= \hat{t}_2^2(J_1 - \hat{t}_2 J_0) - (J_1 + n\hat{\mathbb{K}}_{12}k_2)(n\hat{\mathbb{K}}_{22})^{-1},
\end{aligned}$$

where

$$k_1 \equiv \left(\hat{t}_1 - \frac{\hat{\mathbb{K}}_{12}}{\hat{\mathbb{K}}_{11}}\hat{t}_2\right)(\hat{t}_2 J_2 - J_4) \qquad \text{and} \qquad k_2 \equiv \left(\hat{t}_2 - \frac{\hat{\mathbb{K}}_{12}}{\hat{\mathbb{K}}_{22}}\hat{t}_1\right)(\hat{t}_1 J_1 - J_4).$$

C. Proofs

Proof (Lemma 1). Consider the case with $\mathbb{P}[X - \beta Y < 0] = 0$ first. The result is trivial if $\beta = \infty$ (i.e., Y is a positive or negative random variable). For the remainder of the proof, assume that β is finite. Writing

$$R = \frac{X}{Y} = R_1 + \beta,$$

where $R_1 = (X - \beta Y)/Y$, it is seen that $R < \beta$ (or $R_1 < 0$) if and only if $Y < 0$, as $X - \beta Y > 0$. If $r < \beta$, then $R < r$ (or $R_1 < r - \beta$) if and only if $(X - \beta Y)/(r - \beta) < Y < 0$. Thus

$$\mathbb{P}[R < r] = \mathbb{P}[Y < 0] - \mathbb{P}[Y < (X - \beta Y)/(r - \beta)] = \mathbb{P}[Y < 0] - \mathbb{P}[X - rY < 0].$$

Similarly, if $r > \beta$, then $R < r$ (or $R_1 < r - \beta$) if and only if $Y < 0$ or $Y > (X - \beta Y)/(r - \beta)$. Hence

$$\mathbb{P}[R < r] = \mathbb{P}[Y < 0] + \mathbb{P}[Y > (X - \beta Y)/(r - \beta)] = \mathbb{P}[Y < 0] + \mathbb{P}[X - rY < 0].$$

If $\mathbb{P}[X - \beta Y < 0] = 1$, define $R = (-X)/(-Y)$ and proceed as above. Combining the two cases gives the result. □

Proof ($\tilde{g}_0$ and $\hat{w}$ in (23) have opposite signs). First, note that as a function of r, both $\hat{w}$ and $\mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)$ switch sign only at $r = r^*$. It is immediate that $\hat{w}$ crosses the abscissa from below if $\hat{s} > 0$ and from above otherwise. Regarding $\mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)$, differentiate (22) with respect to r to obtain

$$\tilde{s}_0' \equiv \frac{\mathrm{d}}{\mathrm{d}r}\tilde{s}_0 = \frac{\tilde{s}_0[\mathbb{K}_{12}(\tilde{s}_0, -r\tilde{s}_0) - r\mathbb{K}_{22}(\tilde{s}_0, -r\tilde{s}_0)] + \mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)}{\mathbf{c}_r'\mathbb{K}''(\tilde{s}_0, -r\tilde{s}_0)\mathbf{c}_r}.$$

Using that $\lim_{r\to r^*} \tilde{s}_0' = \hat{s}[\hat{\mathbb{K}}_{12}(\hat{s}, \hat{t}) - r^*\hat{\mathbb{K}}_{22}(\hat{s}, \hat{t})]/\mathbf{c}_{r^*}'\hat{\mathbb{K}}''(\hat{s}, \hat{t})\mathbf{c}_{r^*}$ and simplifying,

$$\frac{\mathrm{d}}{\mathrm{d}r}\mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)\bigg|_{r=r^*} = -\hat{s}\frac{|\mathbb{K}''(\hat{s}, \hat{t})|}{\mathbf{c}_{r^*}'\hat{\mathbb{K}}''(\hat{s}, \hat{t})\mathbf{c}_{r^*}},$$

so that $\mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)$ crosses the axis in the opposite direction as $\hat{w}$ and consequently has the opposite sign. □

Proof (Proof of Theorem 5). We begin by approximating $F_{\bar{W}}^n(0)$. The cgf of W is $\mathbb{K}(s, -rs)$. Define $\tilde{s}_0$ as in (22) and let

$$\tilde{\kappa}_0^{(j)} \equiv \frac{\partial^j}{\partial s^j}\mathbb{K}(s, -rs)\bigg|_{s=\tilde{s}_0} = \sum_{k=0}^{j}\binom{j}{k}(-r)^k\mathbb{K}_{1^{j-k}2^k}(\tilde{s}_0, -r\tilde{s}_0).$$

By Theorem 4, if $\tilde{s}_0 > 0$, then $\mathbb{P}[\bar{W} > 0] = P_1 + \mathcal{O}\left(n^{-1}\right)$. If $\tilde{s}_0 < 0$, the approximation is applied to $-\bar{W}$. The cgf of $-\bar{W}$ is $\mathbb{K}(-s, rs)$, so that the signs on $\tilde{s}_0$ and $\tilde{\kappa}_0^{(3)}$ are reversed, whereas $\tilde{\kappa}_0^{(0)}$ and $\tilde{\kappa}_0^{(2)}$ remain unaltered. Thus $\mathbb{P}[-\bar{W} > 0] = \mathbb{P}[\bar{W} < 0] = -P_1 + \mathcal{O}\left(n^{-1}\right)$. Combining the two approximations yields

$$\mathbb{P}[\bar{W} < 0] = H^*(\tilde{s}_0) - P_1 + \mathcal{O}\left(n^{-1}\right), \tag{51}$$

which can be verified to remain valid if $\tilde{s}_0 = 0$. A similar derivation shows that

$$\mathbb{P}[\bar{Y} < 0] = H^*(\check{t}_0) - P_2 + \mathcal{O}\left(n^{-1}\right), \tag{52}$$

where $\check{t}_0$ solves $\mathbb{K}_2(0, \check{t}_0) = 0$. Finally, define $(\hat{s}, \hat{t})$ as in (20) and let $\hat{t}_r \equiv \hat{t} + r\hat{s}$ as before. Assume for the moment that $\hat{s} > 0$ and $\hat{t}_r > 0$. The joint cgf of (W, Y) is $\mathbb{K}(s, t - rs)$. The saddlepoint is $\hat{\mathbf{t}}_r \equiv (\hat{s}, \hat{t}_r)$. Let $\hat{\mathbf{t}} \equiv (\hat{s}, \hat{t})$, $\widehat{\mathbb{K}}'' \equiv \mathbb{K}''(\hat{s}, \hat{t})$,

$$\hat{\kappa}^{(i,j)} \equiv \frac{\partial^{i+j}}{\partial s^i \partial t^j} \mathbb{K}(s, t - rs)\bigg|_{s=\hat{s},\, t=\hat{t}_r} = \sum_{k=0}^{k} \binom{i+j}{k} (-r)^k \mathbb{K}_{1^{i+j-k}2^k}(\hat{s}, \hat{t}),$$

and $\hat{\mathcal{K}} \equiv \mathbb{K}''_{W,Y}(\hat{s}, \hat{t}_r) = \left[\begin{smallmatrix} \hat{\kappa}^{(1,1)} & \hat{\kappa}^{(1,2)} \\ \hat{\kappa}^{(1,2)} & \hat{\kappa}^{(2,2)} \end{smallmatrix}\right]$. Simplification shows that $\hat{\mathbf{t}}_r' \hat{\mathcal{K}} \hat{\mathbf{t}}_r = \hat{\mathbf{t}}' \widehat{\mathbb{K}}'' \hat{\mathbf{t}}$. By Theorem 4, $\mathbb{P}[\bar{W} > 0, \bar{Y} > 0] = P_3 + \mathcal{O}\left(n^{-1}\right)$, so that

$$\mathbb{P}[\bar{W} < 0, \bar{Y} < 0] = \mathbb{P}[\bar{W} < 0] + \mathbb{P}[\bar{Y} < 0] + P_3 - 1 + \mathcal{O}\left(n^{-1}\right), \quad \hat{s}, \hat{t}_r > 0. \tag{53}$$

Next, assume that $\hat{s} > 0$ and $\hat{t}_r < 0$. Applying Theorem 4 to $(W, -Y)$ switches the sign on $\hat{t}_r$, the off-diagonal elements of $\mathcal{K}$, $\hat{\kappa}^{(3-j,j)}$ for odd j, and, by equation (28), on $\hat{I}_{i,j}$ for even j, but leaves $\hat{\kappa}^{(0,0)}$ and $\hat{\mathbf{t}}' \widehat{\mathbb{K}}'' \hat{\mathbf{t}}$ unaltered. Consequently, $\mathbb{P}[\bar{W} > 0, -\bar{Y} > 0] = \mathbb{P}[\bar{W} > 0, \bar{Y} < 0] = -P_3 + \mathcal{O}\left(n^{-1}\right)$, so that

$$\mathbb{P}[\bar{W} < 0, \bar{Y} < 0] = \mathbb{P}[\bar{Y} < 0] + P_3, \quad \hat{s} > 0, \quad \hat{t}_r < 0. \tag{54}$$

Now assume that $\hat{s} < 0$ and $\hat{t}_r > 0$. Applying Theorem 4 to $(-W, Y)$ switches the sign on $\hat{s}$, the off-diagonal elements of $\mathcal{K}$, $\hat{\kappa}^{(3-j,j)}$ for even j, and $\hat{I}_{i,j}$ for even i, but leaves $\hat{\kappa}^{(0,0)}$ and $\hat{\mathbf{t}}' \widehat{\mathbb{K}}'' \hat{\mathbf{t}}$ unaltered. Consequently, $\mathbb{P}[-\bar{W} > 0, \bar{Y} > 0] = \mathbb{P}[\bar{W} < 0, \bar{Y} > 0] = -P_3 + \mathcal{O}\left(n^{-1}\right)$, so that

$$\mathbb{P}[\bar{W} < 0, \bar{Y} < 0] = \mathbb{P}[\bar{W} < 0] + P_3, \quad \hat{s} < 0, \quad \hat{t}_r > 0. \tag{55}$$

Finally, if both $\hat{s} < 0$ and $\hat{t}_r < 0$, applying Theorem 4 to $(-W, -Y)$ switches the sign on both $\hat{s}$ and $\hat{t}_r$, $\hat{\kappa}^{(3-j,j)}$ for all j, and $\hat{I}_{i,j}$ if $i+j$ is odd, but leaves $\hat{\kappa}^{(0,0)}$, $\mathcal{K}$, and $\hat{\mathbf{t}}'\widehat{\mathbb{K}}''\hat{\mathbf{t}}$ unaltered. Thus

$$\mathbb{P}[-\bar{W} > 0, -\bar{Y} > 0] = \mathbb{P}[\bar{W} < 0, \bar{Y} < 0] = P_3 + \mathcal{O}\left(n^{-1}\right), \quad \hat{s}, \hat{t}_r < 0. \tag{56}$$

Combining (53) to (56) yields

$$\mathbb{P}[\bar{W} < 0, \bar{Y} < 0] = H^*(\hat{t}_r)\mathbb{P}[\bar{W} < 0] + H^*(\hat{s})\mathbb{P}[\bar{Y} < 0] \\ + P_3 - H^*(\hat{t}_r)H^*(\hat{s}) + \mathcal{O}\left(n^{-1}\right),$$

which, together with (11) and upon replacing $\mathbb{P}(\bar{W} < 0)$ and $\mathbb{P}(\bar{Y} < 0)$ with their respective approximations in (51) and (52), gives the result. We remark that it is tempting to replace $1 - 2H^*(\cdot)$ with $-\operatorname{sgn}(\cdot)$, but a careful analysis shows that only as stated is the result valid if any of $\hat{s}$ or $\hat{t}_r$ is zero. □

References

Anderson, T. W. and T. Sawa (1973). Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions. *Econometrica 41*, 683–714.

Bleistein, N. (1966). Uniform expansions of integrals with stationary points near algebraic singularity. *Communications on Pure and Applied Mathematics 19*, 353–370.

Broda, S. A. (2012). The expected shortfall of quadratic portfolios with heavy-tailed risk factors. *Mathematical Finance 22*, 710–728.

Broda, S. A. (2013). Tail probabilities and partial moments for quadratic forms in multivariate generalized hyperbolic random vectors. Tinbergen Institute Discussion Paper 13-001/III, Amsterdam, The Netherlands.

Broda, S. A. and R. Kan (2013). Higher order expansions for multivariate distribution functions. Working paper, University of Toronto.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications.* Cambridge: Cambridge University Press.

Chester, C., B. Friedman, and F. Ursell (1957). An extension of the method of steepest descents. *Proceedings of the Cambridge Philosophical Society 53*, 599–611.

Curtiss, J. H. (1941). On the distribution of the quotient of two chance variables. *The Annals of Mathematical Statistics 12*, 409–421.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics 25*, 631–650.

Daniels, H. E. (1987). Tail probability approximations. *International Statistical Review 55*, 37–48.

Daniels, H. E. and G. A. Young (1991). Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika 78*, 69–179.

Davidson, R. and J. G. MacKinnon (2008). Bootstrap inference in a linear equation estimated by instrumental variables. *Econometrics Journal 11*, 443–477.

Davidson, R. and J. G. MacKinnon (2010). Wild bootstrap tests for iv regression. *Journal of Business & Economic Statistics 28*, 128–144.

Davison, A. C. and D. V. Hinkley (1988). Saddlepoint approximations in resampling methods. *Biometrika 75*, 417–431.

DiCiccio, T. J., M. A. Martin, and G. A. Young (1994). Analytical approximations to bootstrap distribution functions using saddlepoint methods. *Statistica Sinica 4*, 281–295.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika 24*, pp. 428–440.

Flores-Lagunes, A. (2007). Finite sample evidence of iv estimators under weak instruments. *Journal of Applied Econometrics 22*, 677–694.

Forchini, G. (2006). On the bimodality of the exact distribution of the tsls estimator. *Econometric Theory 22*, 932–946.

Fuller, W. A. (1990). *Sampling Statistics*. Hoboken: Wiley.

Geary, R. C. (1944). Extension of a theorem by harald cramér on the frequency distribution of the quotient of two variables. *Journal of the Royal Statistical Society 107*, 56–57.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika 38*, 481–82.

Gordon, R. D. (1941). Values of mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Annals of Mathematical Statistics 12*, 364–366.

Gurland, J. (1948). Inversion formulae for the distribution of ratios. *Annals of Mathematical Statistics 19*, 228–237.

Hauschildt, R. (1969). Saddlepoint approximations to distribution functions. M.Sc. Thesis, University of British Columbia.

Hillier, G. (2006). Yet more on the exact properties of iv estimators. *Econometric Theory 22*, 913–931.

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika 56*, pp. 635–639.

Holly, A. and P. C. B. Phillips (1979). A saddlepoint approximation to the distribution of the k-class estimator of a coefficient in a simultaneous system. *Econometrica 47*, 1527–1547.

Jing, B.-Y., A. Feuerverger, and J. Robinson (1994). On the bootstrap saddlepoint approximations. *Biometrika 81*, 211–215.

Knight, J. L. (1986). Non-normal errors and the distribution of ols and 2sls structural estimators. *Econometric Theory 2*, 75–106.

Kolassa, J. E. (2003). Multivariate saddlepoint tail probability approximations. *The Annals of Statistics 31*(1), 274–286.

Kolassa, J. E. and J. Li (2010). Multivariate saddlepoint approximations in tail probability and conditional inference. *Bernoulli 16*, 1191–1207.

Li, J. (2009). Saddle point approximation. Ph.D. Thesis, Rutgers University.

Lugannani, R. and S. O. Rice (1980). Saddlepoint approximations for the distribution of sums of independent random variables. *Advances in Applied Probability 12*, 475–90.

Maddala, G. S. and J. Jeong (1992). On the exact small sample distribution of the instrumental variable estimator. *Econometrica 60*, 181–183.

Nelson, C. R. and R. Startz (1990a). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *The Journal of Business 63 (Part 2)*, S125–S140.

Nelson, C. R. and R. Startz (1990b). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica 58*, 967–976.

Phillips, P. C. B. (2006). A remark on bimodality and weak instrumentation in structural equation estimation. *Econometric Theory 22*, 947–960.

Rice, S. O. (1968). Uniform asymptotic expansions for saddle point integrals – applications to a probability distribution occuring in noise theory. *Bell System Technical Journal 47*(9), 1971–2013.

Richardson, D. H. (1968). The exact distribution of a structural coefficient estimator. *Journal of the American Statistical Association 63*, 1214–1225.

Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *Journal of the Royal Statistical Society, Series B 44*, 91–101.

Rosén, B. (1961). On the asymptotic distribution of sums of independent identically distributed random variables. *Arkiv för Matematik 4*(24), 323–332.

Sawa, T. (1969). The exact sampling distributions of ordinary least squares and two stage least squares estimates. *Journal of the American Statistical Association 64*, 923–980.

Shephard, N. G. (1991). From characteristic function to distribution function: A simple framework for the theory. *Econometric Theory 7*, 519–529.

Skovgaard, I. M. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability 24*, 875–887.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Wang, S. (1990a). Saddlepoint approximations for bivariate distributions. *Journal of Applied Probability 27*, 586–597.

Wang, S. (1990b). Saddlepoint approximations in resampling analysis. *Annals of the Institute of Statistical Mathematics 42*, 115–131.

Wendel, J. G. (1961). The nonabsolute convergence of Gil-Pelaez' inversion integral. *Annals of Mathematical Statistics 32*, 338–339.

Woglom, G. (2001). More results on the exact small sample properties of the instrumental variable estimator. *Econometrica 69*, 1381–1389.

Zhan, Z. (2010). Detecting weak identification by bootstrap. Working paper, Brown University.