Communications in Statistics—Theory and Methods, 44: 2957–2975, 2015
ISSN: 0361-0926 print / 1532-415X online
DOI: 10.1080/03610926.2013.805626

Asymptotically Efficient Estimation of a Bivariate Gaussian–Weibull Distribution and an Introduction to the Associated Pseudo-truncated Weibull

STEVE P. VERRILL, JAMES W. EVANS, DAVID E. KRETSCHMANN, AND CHERILYN A. HATFIELD

USDA Forest Service Forest Products Laboratory, Madison, Wisconsin, USA

Two important wood properties are stiffness (modulus of elasticity or MOE) and bending strength (modulus of rupture or MOR). In the past, MOE has often been modeled as a Gaussian and MOR as a lognormal or a two or three parameter Weibull. It is well known that MOE and MOR are positively correlated. To model the simultaneous behavior of MOE and MOR for the purposes of wood system reliability calculations, we introduce a bivariate Gaussian–Weibull distribution and the associated pseudo-truncated Weibull. We use asymptotically efficient likelihood methods to obtain an estimator of the parameter vector of the bivariate Gaussian–Weibull, and then obtain the asymptotic distribution of this estimator.

Keywords Bivariate Gaussian-Weibull; Gaussian copula; Likelihood methods; Modulus of rupture; Modulus of elasticity; Normal distribution; One-step Newton estimator; Reliability; Weibull distribution.

1. Introduction

Two important wood properties are stiffness (modulus of elasticity or MOE) and bending strength (modulus of rupture or MOR). In the past, MOE has often been modeled as a Gaussian and MOR as a lognormal or a two- or three-parameter Weibull; see, for example, ASTM, 2010a; Evans and Green, 1988; Green and Evans, 1988.

Design engineers must ensure that the loads to which wood systems are subjected rarely exceed the systems' strengths. To this end, ASTM D 2915 (ASTM, 2010a) and ASTM D 245 or ASTM D 1990 (ASTM 2010b,c) describe the manner in which "allowable properties" are assigned to populations of structural lumber. In essence, an allowable strength property is calculated by estimating a fifth percentile of a population (actually a 95% content, one-sided lower 75% tolerance bound) and then dividing that value by

Received October 19, 2012; Accepted May 8, 2013.

Address correspondence to Steve P. Verrill, USDA Forest Service Forest Products Laboratory, RWV 4851, 1 Gifford Pinchot Drive, Madison WI 53726; E-mail: sverrill@fs.fed.us
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

"duration of load" (aging) and safety factors. The intent is that the population can only be used in applications in which the load does not exceed the allowable property. Of course there are stochastic issues associated with variable loads, uncertainty in estimation, and the division of a percentile with no consideration of population variability. Thus, from a statistician's perspective, this is not an ideal approach to ensuring reliability of wood systems. However, it is the currently codified approach.

To apply this approach, one must obtain estimates of the fifth percentiles of MOR distributions. Currently, one method for obtaining estimates involves fitting a two-parameter Weibull distribution to a sample of MORs. To obtain this fit, either a maximum likelihood approach or a linear regression approach based on order statistics is permitted under ASTM D 5457 (ASTM, 2010d).

Unfortunately, these methods are often applied to populations that are not really distributed as two-parameter Weibulls. For example, in the United States, construction grade 2 by 4's are often classified into visual categories—select structural, number 1, number 2—or into machine stress-rated (MSR) grades. In the case of MSR grades, MOE boundaries are selected, MOE is measured nondestructively, and boards are placed into categories based upon the MOE bins into which the boards fall. Because MOE and MOR are correlated, bins with higher MOE boundaries also tend to contain board populations with higher MOR values. The fifth percentiles of these MOR populations are sometimes estimated by fitting Weibull distributions to these populations. Statisticians recognize that this poses a problem. Even if the full population of lumber strengths were distributed as a Weibull, we would not expect that subpopulations formed by visual grades or MOE binning would continue to be distributed as Weibulls.

In fact, such a subpopulation is not distributed as a Weibull. Instead, if the full joint MOE–MOR population were distributed as a bivariate Gaussian–Weibull, the subpopulation would be distributed as a "pseudo-truncated Weibull" (PTW). In this article, we obtain the distribution of a PTW and show how to obtain estimates of its parameters and its quantiles by fitting a bivariate Gaussian–Weibull to the full MOE–MOR distribution. To do this, we first define a particular form of a bivariate Gaussian–Weibull distribution. In Secs. 2 and 3 of this article, we describe this form and establish that it can be fit by asymptotically efficient likelihood methods in the full MOE–MOR case. In Secs. 4 and 5, we discuss the truncated case and derive the density of a PTW.

We note that the bivariate Gaussian–Weibull distribution has uses other than as a generator of pseudo-truncated Weibulls. For example, engineers who are interested in simulating the performance of wood systems must begin with a model for the joint stiffness, strength distribution of the members of the system; see, for example, Rosowsky and Yu (2004), Rosowsky et al. (2005), and Triche and Partain (2006). Provided that we are considering the *full* population, a Gaussian–Weibull is one possible model for this joint distribution.

Bivariate Gaussian–Weibull distributions have not yet appeared in the literature. However, Gumbel (1960), Freund (1961), Marshall and Olkin (1967), Block and Basu (1974), Clayton (1978), Lee (1979), Hougaard (1986), Sarker (1987), Lu and Bhattacharyya (1990), Patra and Dey (1999), Johnson et al. (1999), Quiroz Flores (2010), Lee et al. (2011), and others have previously investigated bivariate Weibulls.

We note that the bivariate Gaussian–Weibull distribution that we investigate in the current paper is not the only possible bivariate distribution with Gaussian and Weibull marginals. In essence we begin with a "Gaussian copula"—a bivariate uniform distribution generated by starting with a bivariate normal distribution and then applying normal cumulative distribution functions to its marginals. However, there is a large literature on alternative copulas (multivariate distributions with uniform marginals); see, for example, Nelsen (1999) and Jaworski et al. (2010). These alternatives would lead to alternative bivariate Gaussian–Weibulls. Ultimately, the test of the usefulness of our proposed version of a Gaussian–Weibull for a particular application will depend on the match between the theoretical distribution and data. Still, we believe that the analysis of our proposed version in the current paper represents a useful step in the construction and evaluation of bivariate Gaussian–Weibull distributions.

2. A Bivariate Gaussian–Weibull Distribution

To generate a bivariate Gaussian–Weibull distribution, we follow Johnson and Kotz (1972); see also Kotz et al. 2000. (Taylor and Bender, 1988, 1989, introduced this technique in a lumber context.) Let X_1, X_2 be distributed as independent N(0,1)'s. Define $X = \mu + \sigma X_1$ and $Y = \rho X_1 + \sqrt{1-\rho^2} X_2$. Then X is distributed as a $\mathrm{N}(\mu, \sigma^2)$, Y is distributed as a N(0,1), and their correlation is ρ. Now let $U = \Phi(Y)$. Then U is a Uniform(0,1) random variable that is correlated with X. Finally, let $W = (-\ln(1-U))^{1/\beta}/\gamma$. Then W is distributed as a Weibull with shape parameter β and scale parameter $1/\gamma$, and the pair X, W have our joint "bivariate Gaussian–Weibull" distribution. (Verrill and Kretschmann, 2010, Appendix B, performed simulations that indicate that the sample correlation between X and W will be very close to the generating bivariate normal correlation, ρ.) In this article, we require that $\beta > 1$. Given this generating process, it is straightforward to show (see Appendix A) that the joint density is given by

$$\begin{aligned}\text{gaussweib}(x, w; \mu, \sigma, \rho, \gamma, \beta) \equiv\ & \gamma^\beta \beta w^{\beta-1} \exp(-(\gamma w)^\beta) \\ & \times \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma\sqrt{1-\rho^2}} \exp\left(-((x-\mu)/\sigma\right. \\ & \left. -\ \rho y)^2 / (2(1-\rho^2))\right),\end{aligned} \tag{1}$$

where

$$y = \Phi^{-1}(1 - \exp(-(\gamma w)^\beta))$$

and Φ is the N(0,1) cumulative distribution function.

In Fig. 1, we provide a contour plot of the bivariate Gaussian–Weibull distribution for a coefficient of variation (CV) equal to 0.15 and a generating correlation equal to 0.7. Additional plots are provided in Verrill et al. (2012a). Note in these plots that as the CV declines from 0.35 to 0.25 to 0.15 (as the Weibull shape parameter increases from 3.13 to 4.54 to 7.91) the density contours become much less elliptical. That is, the distribution diverges from a bivariate normal. We would expect this as a Weibull is "like a normal" for shape near 3.6 (skewness equals 0.00056, excess kurtosis equals -0.28), and a Weibull becomes skewed to the left and leptokurtic as the shape increases.

3. Asymptotic Distribution of the Estimated Parameter Vector of the Bivariate Gaussian–Weibull Distribution

Assume that we have have n independent pairs of observations, $(x_1, w_1), \ldots, (x_n, w_n)$ from the bivariate Gaussian–Weibull distribution. Then we have the following theorem.

Theorem 3.1.

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, I(\boldsymbol{\theta})^{-1}), \tag{2}$$

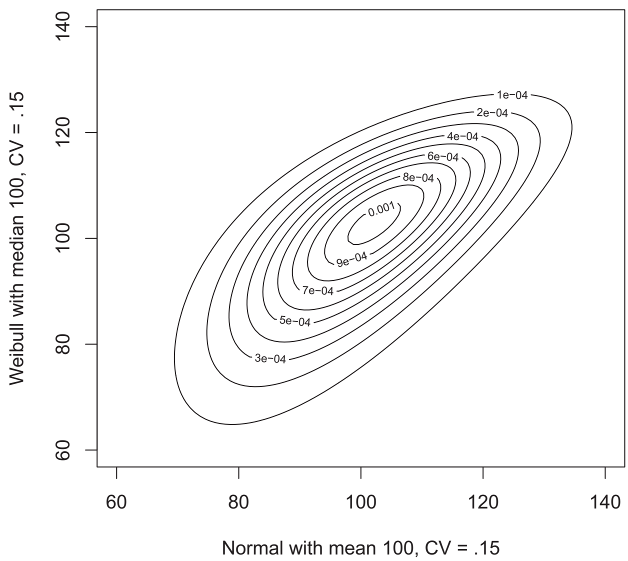

Figure 1. Contour plot of the bivariate Gaussian–Weibull density for Gaussian and Weibull coefficients of variation equal to 0.15 and a generating correlation of 0.7.

where $\boldsymbol{\theta} \equiv (\mu, \sigma, \rho, \gamma, \beta)^T$, $\hat{\boldsymbol{\theta}} \equiv (\hat{\mu}, \hat{\sigma}, \hat{\rho}, \hat{\gamma}, \hat{\beta})^T$ *is the one-step Newton estimator based on bivariate Gaussian–Weibull theory (the gradient and Hessian used to calculate the Newton step correspond to the first and second partials of the full Gaussian–Weibull likelihood), and the step is taken from the initial estimate* $\boldsymbol{\theta}_i \equiv (\mu_i, \sigma_i, \rho_i, \gamma_i, \beta_i)^T$ *where* μ_i, σ_i *are the usual univariate maximum likelihood estimators of the mean and standard deviation of a Gaussian* ($\bar{x}$ *and* $\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2/n}$), γ_i, β_i *are the usual univariate maximum likelihood estimators of* $1/$*scale and shape for a Weibull (see, for example, Johnson et al., 1994),* ρ_i *is the* $\sqrt{n}$*-consistent estimator of* ρ *introduced in Appendix B, and the elements of* $I(\boldsymbol{\theta})$ *are listed in Appendix C. That is,* $\hat{\boldsymbol{\theta}}$ *is given by the Newton step*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_i - \mathbf{H}_{|\boldsymbol{\theta}_i}^{-1}\mathbf{g}_{|\boldsymbol{\theta}_i} \tag{3}$$

where the j*th element of the gradient* $\mathbf{g}_{|\boldsymbol{\theta}_i}$ *is the first partial of the log likelihood with respect to the* j*th parameter at* $\boldsymbol{\theta}_i$ *and the* j, k*th element of the Hessian* $\mathbf{H}_{|\boldsymbol{\theta}_i}$ *is the second partial of the log likelihood with respect to the* j*th and* k*th parameters at* $\boldsymbol{\theta}_i$.

Proof. The proof is an application of Theorem 4.2 of Chapter 6 of Lehmann (1983). To invoke Lehmann's theorem, we must first establish that the ρ estimator introduced in Appendix B is indeed $\sqrt{n}$-consistent. The proof of this fact is outlined in Appendix B and is provided in full in Verrill et al. (2012a).

We must then establish Lehmann's conditions. That his conditions (A0)–(A2) and A hold is clear. Lehmann's condition (B)(8) is established in Appendix E1 of Verrill et al. (2012a). Lehmann's condition (B)(9) is established in Appendices E2 and E3 of Verrill et al. (2012a). The fact that the information matrix is positive definite (Lehmann's condition C) is established in Appendix D of the current paper. Lehmann's condition (D) is established in Appendix J of Verrill et al. (2012a). □

4. A Truncated Bivariate Gaussian–Weibull Distribution

In wood engineering applications, it is often the case that we do not have data from a full bivariate Gaussian–Weibull distribution. Instead, we have data from the subpopulation that is formed by considering lumber whose MOE values lie between two pre-determined limits, c_l and c_u (that is, we have machine stress-rated lumber). It is clear that the joint density in this case is

$$\text{gaussweib}(x, w; \mu, \sigma, \rho, \gamma, \beta)/\left(\Phi\left((c_u - \mu)/\sigma\right) - \Phi((c_l - \mu)/\sigma)\right) \quad (4)$$

for x between c_l and c_u and 0 elsewhere.

5. The Pseudo-Truncated Weibull Distribution

The pseudo-truncated Weibull distribution function at w is given by integrating the truncated bivariate Gaussian–Weibull density (4) over the region $[c_l, c_u] \times [0, w]$. That is, from Eq. (1),

$$F_{\text{PTW}}(w) = \int_0^w F_1(s) \times F_2(s)/\left(\Phi((c_u - \mu)/\sigma) - \Phi((c_l - \mu)/\sigma)\right) ds \quad (5)$$

where

$$F_1(s) \equiv \gamma^\beta \beta s^{\beta-1} \exp(-(\gamma s)^\beta) \quad (6)$$

and

$$\begin{aligned} F_2(s) &\equiv \int_{c_l}^{c_u} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma\sqrt{1-\rho^2}} \exp\left(-\left((x-\mu)/\sigma - \rho y\right)^2/(2(1-\rho^2))\right) dx \quad (7) \\ &= \Phi((c_u - \mu)/(\sigma\sqrt{1-\rho^2}) - \rho y/\sqrt{1-\rho^2}) \\ &\quad - \Phi((c_l - \mu)/(\sigma\sqrt{1-\rho^2}) - \rho y/\sqrt{1-\rho^2}), \end{aligned}$$

where

$$y = \Phi^{-1}(1 - \exp(-(\gamma s)^\beta))$$

From results (5)–(7), the pseudo-truncated Weibull density is given by

$$\begin{aligned} f_{\text{PTW}}(w) = \gamma^\beta \beta w^{\beta-1} &\exp(-(\gamma w)^\beta) \quad (8) \\ &\times \left(\Phi\left((c_u - \mu)/\left(\sigma\sqrt{1-\rho^2}\right) - \rho y/\sqrt{1-\rho^2}\right)\right. \\ &- \left.\Phi\left((c_l - \mu)/\left(\sigma\sqrt{1-\rho^2}\right) - \rho y/\sqrt{1-\rho^2}\right)\right) \\ &/(\Phi((c_u - \mu)/\sigma) - \Phi((c_l - \mu)/\sigma)), \end{aligned}$$

where

$$y = \Phi^{-1}\left(1 - \exp\left(-(\gamma w)^\beta\right)\right)$$

Thus, as we would expect, for $\rho = 0$, the pseudo-truncated Weibull density is simply the Weibull density, $\gamma^\beta \beta w^{\beta-1} \exp(-(\gamma w)^\beta)$.

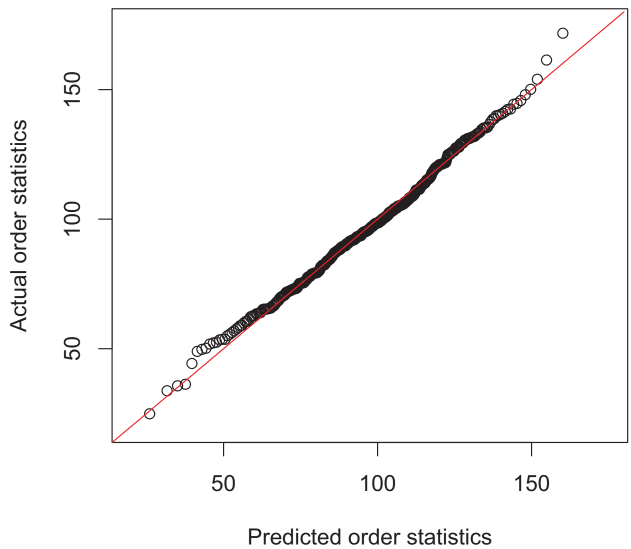

Figure 2. Weibull probability plot of a pseudo-truncated Weibull with generating coefficient of variation equal to 0.25 and generating correlation equal to 0.0. The straight line is the ordinate equals abscissa line.

In Appendix K of Verrill et al. (2012a), we show that as $\rho \to 1$, the density of a pseudo-truncated Weibull density converges to the density of a truncated Weibull.

Figures 2 and 3 are (one version of) Weibull probability plots of PTW data. We plot the ordered data from a PTW sample against the predicted ordered data from the best Weibull

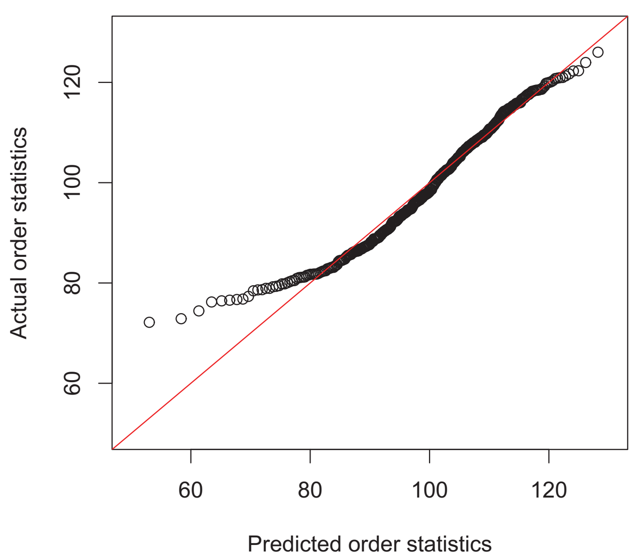

Figure 3. Weibull probability plot of a pseudo-truncated Weibull with generating coefficient of variation equal to 0.25 and generating correlation equal to 0.99. The straight line is the ordinate equals abscissa line.

fit to the data. If the data really were Weibull, then the plots would be approximately linear. In Fig. 2, the generating X, Y correlation was 0, so the data actually was Weibull and the plot is approximately linear. In Fig. 3, the generating X, Y correlation was 0.99, so the data was "far from Weibull" and the plot is quite nonlinear. For both data sets, the Weibull coefficient of variation was 0.25 and c_{l} and c_{u} corresponded to the 0.2 and 0.8 quantiles of the Gaussian distribution.

In Appendix L of Verrill et al. (2012a), we formally establish that for $\rho \neq 0$, pseudo-truncated Weibull distributions are not Weibull distributions.

6. Summary

In the context of wood strength modeling, we have introduced a bivariate Gaussian–Weibull distribution and the associated pseudo-truncated Weibull distribution. In this article, we have obtained the asymptotic distribution of the estimated parameter vector for a bivariate Gaussian–Weibull distribution. In Verrill et al. (2012b,c) we describe a Web-based program that obtains this asymptotically efficient estimate, simulations that investigate the small sample properties of this estimate, and additional simulations that establish that Weibull fits to PTW data can yield poor estimates of probabilities of failure.

References

ASTM (2010a). Standard practice for evaluating allowable properties for grades of structural lumber. D 2915-10. West Conshohocken, PA: ASTM International.

ASTM (2010b). Standard practice for establishing structural grades and related allowable properties for visually graded lumber. D 245-06. West Conshohocken, PA: ASTM International.

ASTM (2010c). Standard practice for establishing allowable properties for visually-graded dimension lumber from in-grade tests of full-size specimens. D 1990-07. West Conshohocken, PA: ASTM International.

ASTM (2010d). Standard specification for computing reference resistence of wood-based materials and structural connections for load and resistance factor design. D 5457-04a. West Conshohocken, PA: ASTM International.

Block, H. W., Basu, A. P. (1974). A Continuous bivariate exponential extension. *J. Amer. Statist. Assoc.* 69:1031–1037.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65:141–151.

Evans, J. W., Green, D. W. (1988). *Mechanical Properties of Visually Graded Dimension Lumber: Vol 2. Douglas Fir–Larch*. Pub. PB-88-159-397. Springfield, VA: National Technical Information Service.

Freund, J. E. (1961). A bivariate extension of the exponential distribution. *J. Amer. Statist. Assoc.* 56:971–977.

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. *Ann. Mathemat. Statist.* 12:364–366.

Green, D. W., Evans, J. W. (1988). *Mechanical Properties of Visually Graded Dimension Lumber: Vol 4. Southern Pine*. Pub. PB-88-159-413. Springfield, VA: National Technical Information Service.

Gumbel, E. J. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc.* 55:698–707.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika* 73:671–678.

Jaworski, P., Durante, F., Härdle, W., Rychlik, T. Eds (2010). *Copula Theory and Its Applications. Lecture Notes in Statistics 198*. New York: Springer.

Johnson, N. L., Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley and Sons.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions. vol 1. 2nd ed*. New York: John Wiley and Sons.

Johnson, R. A., Evans, J. W., Green, D. W. (1999). Some bivariate distributions for modeling the strength properties of lumber. Research Paper FPL-RP-575. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory.

Kendall, M., Stuart, A. (1977). *The Advanced Theory of Statistics, vol. 1*. New York: MacMillan Publishing Co., Inc.

Kotz, S., Balakrishnan, N., Johnson, N. L. (2000). *Continuous Multivariate Distributions. Volume 1: Models and Applications. 2nd ed*. New York: John Wiley and Sons.

Lee, E., Kim, C., Lee, S. (2011). Life expectancy estimate with bivariate weibull distribution using archimedean copula. *Int. J. Biometrics Bioinform.* 5(3):149–161.

Lee, L. (1979). Multivariate distributions having Weibull properties. *J. Multivariate Anal.* 9:267–277.

Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: John Wiley and Sons.

Lu, J. C., Bhattacharyya, G. K. (1990). Some new constructions of bivariate Weibull models. *Ann. Inst. Statist. Math.* 42(3):543–559.

Marshall, A. W., Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* 62:30–44.

Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer.

Quiroz Flores, A. (2010). Testing copula functions as a method to derive bivariate Weibull distributions. Wilf Family Department of Politics, New York University working paper.

Rosowsky, D. V., Yu, G. (2004). Partial factor approach to repetitive-member system factors. *J. Struct. Eng.* 130(11):1829–1841.

Rosowsky, D. V., Yu, G., Bulleit, W. M. (2005). Reliability of light-frame wall systems subject to combined axial and transverse loads. *J. Struct. Eng.* 131(9):1444–1455.

Rudin, W. (1987). *Real and Complex Analysis*. New York: McGraw-Hill.

Patra, K., Dey, D. K. (1999). A multivariate mixture of Weibull distributions in reliability modeling. *Statist. Probab. Lett.* 45:225–235.

Sarkar, S. K. (1987). A continuous bivariate exponential distribution. *J. Amer. Statist. Assoc.* 82:667–675.

Taylor, S. E., Bender, D. A. (1988). Simulating correlated lumber properties using a modified multivariate normal approach. *Trans. Amer. Soc. Agri. Eng.* 31(1):182–186.

Taylor, S. E., Bender, D. A. (1989). A method for simulating multiple correlated lumber properties. *Forest Prod. J.* 39(7/8):71–74.

Triche, M. H., Partain, J. A. (2006). Effect of lumber variability on truss performance. Proc. 2006 World Conf. Timber Eng. Portland, Oregon.

Verrill, S., Durst, M. (2005). The decline and fall of Type II error rates. *Amer. Statist.* 59(4):287–291.

Verrill, S. P., Evans, J. W., Kretschman, D. E., Hatfield, C. A. (2012a). Asymptotically efficient estimation of a bivariate Gaussian–Weibull distribution and an introduction to the associated pseudo-truncated Weibull. USDA Forest Products Laboratory Research Paper 666. http://www1.fpl.fs.fed.us/fpl_rp666.pdf

Verrill, S. P., Evans, J. W., Kretschman, D. E., Hatfield, C. A. (2012b). Small sample properties of asymptotically efficient estimators of the parameters of a bivariate Gaussian–Weibull distribution. USDA Forest Products Laboratory Research Paper 667. http://www1.fpl.fs.fed.us/fpl_rp667.pdf

Verrill, S. P., Evans, J. W., Kretschman, D. E., Hatfield, C. A. (2012c). An evaluation of a proposed revision of the ASTM D 1990 grouping procedure. USDA Forest Products Laboratory Research Paper 671. http://www1.fpl.fs.fed.us/fpl_rp671.pdf

Verrill, S. P., Kretschman, D. E. (2010). Repetitive member factors for the allowable properties of wood products. USDA Forest Products Laboratory Research Paper 657. http://www1.fpl.fs.fed.us/fpl_rp657.pdf

Appendix A—Bivariate Gaussian–Weibull Density

Let X, Y have a joint bivariate normal distribution with

$$X \sim \mathrm{N}(\mu, \sigma^2)$$
$$Y \sim \mathrm{N}(0, 1)$$

and correlation$(X, Y) = \rho$.

Since $Y \sim \mathrm{N}(0, 1)$, we know that $\Phi(Y)$ is distributed as a Uniform (0,1). (Here, Φ denotes the N(0,1) cumulative distribution function.) Thus, we know that

$$W \equiv (-\ln(1 - \Phi(Y)))^{1/\beta}/\gamma \sim \mathrm{Weibull}(\gamma, \beta) \tag{A.9}$$

(a two-parameter Weibull distribution with scale parameter $1/\gamma$ and shape parameter β).

We then say that X, W have a bivariate Gaussian–Weibull distribution with parameters $\mu, \sigma, \rho, \gamma$, and β.

Using the multivariate form of the change-of-variables theorem (see, for example, Rudin 1987), we can calculate the joint density function of X, W. First, we invert Eq. (9) to obtain

$$Y = \Phi^{-1}\left(1 - \exp\left(-(\gamma W)^\beta\right)\right)$$

Thus, the transform that takes (x, w) to (x, y) is

$$\mathbf{T}(x, w) = \begin{pmatrix} T_1(x, w) \\ T_2(x, w) \end{pmatrix} = \begin{pmatrix} x \\ \Phi^{-1}\left(1 - \exp(-(\gamma w)^\beta)\right) \end{pmatrix}$$

The corresponding Jacobian matrix is

$$\begin{pmatrix} \partial T_1/\partial x & \partial T_1/\partial w \\ \partial T_2/\partial x & \partial T_2/\partial w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \gamma^\beta \beta w^{\beta-1} \exp(-(\gamma w)^\beta)/\phi(\Phi^{-1}(1 - \exp(-(\gamma w)^\beta))) \end{pmatrix}$$

and the absolute value of its determinant is

$$\det = \gamma^\beta \beta w^{\beta-1} \exp(-(\gamma w)^\beta)/\phi(\Phi^{-1}(1 - \exp(-(\gamma w)^\beta))).$$

Thus, the Gaussian–Weibull pdf at x, w is

$$\mathrm{bivnorm}(x, y, \mu, \sigma, \rho) \times \det, \tag{A.10}$$

where

$$y = \Phi^{-1}\left(1 - \exp\left(-(\gamma w)^\beta\right)\right) \tag{A.11}$$

and

$$\mathrm{bivnorm}(x, y, \mu, \sigma, \rho) = \frac{1}{2\pi} \times \frac{1}{\sigma\sqrt{1-\rho^2}} \times \exp(\mathrm{arg})$$

where

$$\begin{aligned} \mathrm{arg} &= -\left((x-\mu)^2/\sigma^2 - 2\rho(x-\mu)y/\sigma + y^2\right)/(2(1-\rho^2)) \\ &= -\left((x-\mu)^2/\sigma^2 - 2\rho(x-\mu)y/\sigma + \rho^2 y^2 + y^2 - \rho^2 y^2\right)/(2(1-\rho^2)) \end{aligned}$$

$$= -\left((x-\mu)/\sigma - \rho y\right)^2/(2(1-\rho^2)) - y^2/2$$

That is, the Gaussian–Weibull pdf at x, w is given by

$$\begin{aligned} \text{gaussweib}(x, w; \mu, \sigma, \rho, \gamma, \beta) \equiv \gamma^{\beta}\beta w^{\beta-1} \exp\left(-(\gamma w)^{\beta}\right) \\ \times \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma\sqrt{1-\rho^2}} \exp\left(-\left((x-\mu)/\sigma\right.\right. \\ \left.\left. -\rho y\right)^2/(2(1-\rho^2))\right) \end{aligned} \tag{A.12}$$

Appendix B—$\sqrt{n}$-consistent Initial Estimators of the Parameters

We first list a lemma that provides a useful fact about the tail behavior of normal distributions. Versions of this fact have appeared previously in the statistical literature. See, for example, the discussions of "Mills' ratio" in Kendall and Stuart (1977) and Johnson et al. (1994). The particular form of the fact described in Lemma 1 is due to Gordon (1941). A simple proof of Lemma 1 is given in Verrill and Durst (2005).

Lemma B.1. *For $x < 0$,*

$$x^2/(x^2+1) < \Phi(x)/\left(\phi(x)/(-x)\right) < 1 \tag{B.13}$$

and for $x > 0$,

$$x^2/(x^2+1) < (1-\Phi(x))/(\phi(x)/x) < 1, \tag{B.14}$$

where $\Phi(x)$ is the N(0,1) cumulative distribution function and $\phi(x)$ is the N(0,1) probability density function.

Now, to invoke Theorem 4.2 of Lehmann (1983) to establish that our final estimators of the parameters are asymptotically efficient, we need to establish that our initial estimates of the parameters are $\sqrt{n}$-consistent. ($\hat{a}_n$ is a $\sqrt{n}$-consistent estimator of a if $\sqrt{n}(\hat{a}_n - a) = O_p(1)$. A sequence of random variables $\{X_n\}$ is $O_p(1)$ if given any $\epsilon > 0$, we can find constants M_ϵ, N_ϵ such that $n > N_\epsilon$ implies that $\text{Prob}(|X_n| > M_\epsilon) < \epsilon$.) As our initial estimators of μ and σ, we take the univariate Gaussian maximum likelihood estimators $\bar{x} = \quad x_i/n$ and $s = \sqrt{\quad (x_i - \bar{x})^2/n}$. As our initial estimators of γ and β we take the univariate Weibull maximum likelihood estimators, $\hat{\gamma}$ and $\hat{\beta}$; see, for example, Johnson et al., 1994. Thus, our initial estimators of μ, σ, γ, and β are $\sqrt{n}$-consistent. Our initial estimator of ρ is given by

$$\hat{\rho} \equiv \hat{s}_{xy}/\sqrt{s_{xx}\hat{s}_{yy}} \tag{B.15}$$

where

$$\hat{s}_{xy} \equiv \sum_{i=1}^{n}(x_i - \bar{x})(\hat{y}_i - \hat{\bar{y}})$$

$$s_{xx} \equiv \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\hat{s}_{yy} \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2$$

$$\bar{\hat{y}} \equiv \sum_{i=1}^{n} \hat{y}_i / n$$

and

$$\hat{y}_i \equiv g(w_i; \hat{\gamma}, \hat{\beta}) \equiv \Phi^{-1}\left(1 - \exp\left(-(\hat{\gamma} w_i)^{\hat{\beta}}\right)\right) \tag{B.16}$$

Theorem B.1.

$$\sqrt{n}\,(\hat{\rho} - \rho) = O_p(1)$$

where $\hat{\rho}$ is defined in Eq. (15).

Proof. We only outline the proof here. Details can be found in Appendix B of Verrill et al. (2012a).

Define

$$s_{xy} \equiv \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$s_{yy} \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\bar{y} \equiv \sum_{i=1}^{n} y_i / n,$$

where

$$y_i \equiv g(w_i; \gamma, \beta) \equiv \Phi^{-1}\left(1 - \exp\left(-(\gamma w_i)^{\beta}\right)\right) \tag{B.17}$$

(The distinction between the "hatted" variables in definitions (16) and the "unhatted" variables in definitions (17) is that in the hatted case, γ, β are replaced by their estimates $\hat{\gamma}, \hat{\beta}$.)

We know that

$$r \equiv s_{xy} / \sqrt{s_{xx} s_{yy}}$$

is a $\sqrt{n}$-consistent estimator of ρ. (That is, we know that $\sqrt{n}(r - \rho) = O_p(1)$.) Thus, we will be done if we can show that

$$\sqrt{n}\ (r - \hat{\rho}) = O_p(1) \tag{B.18}$$

We have

$$\begin{aligned} r - \hat{\rho} &= s_{xy} / \sqrt{s_{xx} s_{yy}} - \hat{s}_{xy} / \sqrt{s_{xx} \hat{s}_{yy}} \\ &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{s_{xx} s_{yy}}} - \frac{\sum_{i=1}^{n} (x_i - \bar{x})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{s_{xx} s_{yy}}} \end{aligned}$$

$$+\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{s_{xx}s_{yy}}}-\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{s_{xx}\hat{s}_{yy}}}$$

$$\equiv D_1+D_2 \tag{B.19}$$

To show that $\sqrt{n}D_1 = O_p(1)$, we need to show that

$$\sqrt{n}\left(\sum_{i=1}^{n}(x_i-\bar{x})\left(y_i-\bar{y}-(\hat{y}_i-\bar{\hat{y}})\right)/n\right)=O_p(1) \tag{B.20}$$

By the Cauchy-Schwarz inequality and the fact that $\sum_{i=1}^{n}(x_i-\bar{x})^2/n \xrightarrow{p} \sigma^2$, we know that we can establish result (20) by establishing that

$$\sum_{i=1}^{n}\left(y_i-\bar{y}-(\hat{y}_i-\bar{\hat{y}})\right)^2=O_p(1) \tag{B.21}$$

and it is clear that result (21) follows if

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=O_p(1) \tag{B.22}$$

(This follows because $\sum_{i=1}^{n}(z_i-\bar{z})^2 \le \sum_{i=1}^{n} z_i^2$.)

From definitions (16) and (17) we have

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}\left(g(w_i;\gamma,\beta)-g\left(w_i;\hat{\gamma},\hat{\beta}\right)\right)^2 \tag{B.23}$$

By Taylor's theorem this equals

$$\sum_{i=1}^{n}\left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial\gamma}|_{\boldsymbol{\theta}_{*,i}}(\hat{\gamma}-\gamma)+\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial\beta}|_{\boldsymbol{\theta}_{*,i}}\left(\hat{\beta}-\beta\right)\right)^2 \tag{B.24}$$

where $\boldsymbol{\theta}=(\gamma,\beta)^T$ and $\boldsymbol{\theta}_{*,i}\equiv(\gamma_{*,i},\beta_{*,i})^T$ lies on the line between $(\gamma,\beta)^T$ and $(\hat{\gamma},\hat{\beta})^T$.

Thus, given the Cauchy-Schwarz inequality, to establish result (22), it is sufficient to establish

$$\sum_{i=1}^{n}\left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial\gamma}|_{\boldsymbol{\theta}_{*,i}}\right)^2(\hat{\gamma}-\gamma)^2=O_p(1) \tag{B.25}$$

and

$$\sum_{i=1}^{n}\left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial\beta}|_{\boldsymbol{\theta}_{*,i}}\right)^2\left(\hat{\beta}-\beta\right)^2=O_p(1) \tag{B.26}$$

Because $\hat{\gamma}$ and $\hat{\beta}$ are the maximum likelihood estimates of γ and β, to establish results (25) and (26), it is sufficient to establish

$$\sum_{i=1}^{n}\left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial\gamma}|_{\boldsymbol{\theta}_{*,i}}\right)^2/n=O_p(1) \tag{B.27}$$

and

$$\sum_{i=1}^{n} \left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial \beta}|_{\boldsymbol{\theta}_{*,i}} \right)^2 /n = O_p(1) \tag{B.28}$$

Consider result (27). We have

$$\begin{aligned}
\sum_{i=1}^{n} \left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial \gamma}|_{\boldsymbol{\theta}_{*,i}} \right)^2 /n &= \sum_{w_i<w_{\text{low}}} \left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial \gamma}|_{\boldsymbol{\theta}_{*,i}} \right)^2 /n \\
&\quad + \sum_{w_{\text{low}} \le w_i \le w_{\text{up}}} \left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial \gamma}|_{\boldsymbol{\theta}_{*,i}} \right)^2 /n \\
&\quad + \sum_{w_{\text{up}}<w_i} \left(\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial \gamma}|_{\boldsymbol{\theta}_{*,i}} \right)^2 /n \\
&\equiv S_1 + S_2 + S_3,
\end{aligned}$$

where $0 < w_{\text{low}} < w_{\text{up}}$. Now we have

$$\frac{\partial g(w_i;\boldsymbol{\theta})}{\partial \gamma}|_{\boldsymbol{\theta}_{*,i}} = \beta_{*,i}\gamma_{*,i}^{\beta_{*,i}-1} w_i^{\beta_{*,i}} \exp\left(-(\gamma_{*,i}w_i)^{\beta_{*,i}}\right)/\phi\left(\Phi^{-1}\left(1-\exp\left(-(\gamma_{*,i}w_i)^{\beta_{*,i}}\right)\right)\right). \tag{B.29}$$

It is clear that this is "essentially" bounded for S_2. However, for S_1 and S_3 we have both numerators and denominators that are going to 0. The result is not immediately obvious. In Verrill et al. (2012a) we use Lemma 1 to show that S_1 and S_3 are $O_p(1)$.

This establishes result (27). Thus, to complete the proof of (22) we need to establish result (28). In general, the proof of result (28) is essentially the same as the proof of result (27); see Verrill et al. (2012a) for details.

As noted above, results (27) and (28) establish results (25) and (26) which establish result (22) which establishes

$$\sqrt{n}\, D_1 = O_p(1). \tag{B.30}$$

To complete the proof of the theorem we now need to show that

$$\sqrt{n}\, D_2 = O_p(1). \tag{B.31}$$

To establish (31), we first need to establish a few facts about y_i and $\hat{y}_i$. By the Cauchy-Schwarz inequality and result (22), we have

$$\begin{aligned}
\sqrt{n}\left|\bar{y} - \bar{\hat{y}}\right| &\le \sqrt{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|/n \\
&\le \sqrt{n}\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2/n\right)^{1/2} \\
&= \left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)^{1/2} = O_p(1).
\end{aligned}$$

Thus,

$$\sqrt{n}\left(\hat{\bar{y}}^2 - \bar{y}^2\right) = \sqrt{n}\left(\hat{\bar{y}} - \bar{y}\right)\left(\hat{\bar{y}} + \bar{y}\right) = O_p(1). \tag{B.32}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sum_{i=1}^{n}(\hat{y}_i + y_i)^2/n &= \sum_{i=1}^{n}(\hat{y}_i - y_i + 2y_i)^2/n \\
&= \left|\sum_{i=1}^{n}(\hat{y}_i - y_i)^2/n + 4\sum_{i=1}^{n}(\hat{y}_i - y_i)y_i/n + 4\sum_{i=1}^{n} y_i^2/n\right| \\
&\leq \sum_{i=1}^{n}(\hat{y}_i - y_i)^2/n + 4\left(\sum_{i=1}^{n}(\hat{y}_i - y_i)^2/n\right)^{1/2}\left(\sum_{i=1}^{n} y_i^2/n\right)^{1/2} \\
&\quad +4\sum_{i=1}^{n} y_i^2/n.
\end{aligned} \tag{B.33}$$

By results (22) and (33) and the fact that

$$\sum_{i=1}^{n} y_i^2/n \xrightarrow{p} E(Y^2)$$

we have

$$\sum_{i=1}^{n}(\hat{y}_i + y_i)^2/n = O_p(1) \tag{B.34}$$

By the Cauchy-Schwarz inequality and results (22) and (34) we have

$$\begin{aligned}
\sqrt{n}\left|\sum_{i=1}^{n}\left(\hat{y}_i^2 - y_i^2\right)/n\right| &= \sqrt{n}\left|\sum_{i=1}^{n}(\hat{y}_i - y_i)(\hat{y}_i + y_i)/n\right| \\
&\leq \sqrt{n}\left(\sum_{i=1}^{n}(\hat{y}_i - y_i)^2/n\right)^{1/2}\left(\sum_{i=1}^{n}(\hat{y}_i + y_i)^2/n\right)^{1/2} \\
&= O_p(1)
\end{aligned} \tag{B.35}$$

By results (32) and (35) we have

$$\begin{aligned}
\sqrt{n}\left(\hat{s}_{yy}/n - s_{yy}/n\right) &= \sqrt{n}\left(\sum_{i=1}^{n}\hat{y}_i^2/n - \hat{\bar{y}}^2 - \left(\sum_{i=1}^{n} y_i^2/n - \bar{y}^2\right)\right) \\
&= \sqrt{n}\left(\sum_{i=1}^{n}\left(\hat{y}_i^2 - y_i^2\right)/n - \left(\hat{\bar{y}}^2 - \bar{y}^2\right)\right) \\
&= O_p(1)
\end{aligned} \tag{B.36}$$

From result (36) we have

$$\sqrt{n}\left(\sqrt{\hat{s}_{yy}/n}-\sqrt{s_{yy}/n}\right)=\sqrt{n}\left(\hat{s}_{yy}/n-s_{yy}/n\right)/\left(\sqrt{\hat{s}_{yy}/n}+\sqrt{s_{yy}/n}\right)$$
$$=O_p(1). \tag{B.37}$$

Now

$$D_2 \equiv \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{s_{xx}s_{yy}}}-\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{s_{xx}\hat{s}_{yy}}} \tag{B.38}$$
$$=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\hat{y}_i-\bar{\hat{y}})}{n}\times\frac{\sqrt{s_{xx}\hat{s}_{yy}/n^2}-\sqrt{s_{xx}s_{yy}/n^2}}{\sqrt{s_{xx}s_{yy}s_{xx}\hat{s}_{yy}/n^4}}$$
$$\equiv F_1\times F_2.$$

By the Cauchy-Schwarz inequality and (36)

$$|F_1|\le\left(\sum_{i=1}^{n}(x_i-\bar{x})^2/n\right)^{1/2}\left(\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2/n\right)^{1/2} \tag{B.39}$$
$$=\sqrt{s_{xx}/n}\sqrt{\hat{s}_{yy}/n}\xrightarrow{p}\sigma\times 1.$$

By results (37) and (38)

$$\sqrt{n}F_2=\frac{\sqrt{s_{xx}/n}}{\sqrt{s_{xx}s_{yy}s_{xx}\hat{s}_{yy}/n^4}}\times\sqrt{n}\left(\sqrt{\hat{s}_{yy}/n}-\sqrt{s_{yy}/n}\right)=O_p(1) \tag{B.40}$$

Results (38)–(40) imply that

$$\sqrt{n}\,D_2=O_p(1) \tag{B.41}$$

This completes the proof of the theorem. □

Appendix C—Elements of the Information Matrix

Denote the information by

$$I(\boldsymbol{\theta})\equiv\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15}\\ a_{12} & a_{22} & a_{23} & a_{24} & a_{25}\\ a_{13} & a_{23} & a_{33} & a_{34} & a_{35}\\ a_{14} & a_{24} & a_{34} & a_{44} & a_{45}\\ a_{15} & a_{25} & a_{35} & a_{45} & a_{55}\end{pmatrix}$$

Then, from Appendices D and E2 of Verrill et al. (2012a) we have

$$a_{11}=-E\left(\frac{\partial^2\ln(f(x,w))}{\partial\mu^2}\right)=\frac{1}{\sigma^2(1-\rho^2)} \tag{C.42}$$
$$a_{22}=-E\left(\frac{\partial^2\ln(f(x,w))}{\partial\sigma^2}\right)=\frac{2-\rho^2}{\sigma^2(1-\rho^2)} \tag{C.43}$$

$$a_{33} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \rho^2} \quad = \frac{(1+\rho^2)}{(1-\rho^2)^2} \tag{C.44}$$

$$a_{44} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \gamma^2} \quad = \frac{\rho^2}{1-\rho^2} E \quad \frac{\partial y}{\partial \gamma} \quad ^2 \quad + \frac{\beta^2}{\gamma^2}, \tag{C.45}$$

where y is given by (11) and

$$\frac{\partial y}{\partial \gamma} = \sqrt{2\pi} \times \beta \gamma^{\beta-1} \times w^{\beta} \times \exp(-(\gamma w)^{\beta}) \times \exp(y^2/2)$$

$$a_{55} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \beta^2} \quad = \frac{\rho^2}{1-\rho^2} E \quad \frac{\partial y}{\partial \beta} \quad ^2 \quad + \frac{1}{\beta^2} \tag{C.46}$$

$$+E\left((\ln(w))^2\right) + \frac{2}{\beta} E\,(\ln(w))$$

$$+2\ln(\gamma) E\,(\ln(w)) + \frac{2\ln(\gamma)}{\beta}$$

$$+(\ln(\gamma))^2\,,$$

where

$$\frac{\partial y}{\partial \beta} = \sqrt{2\pi} \times (\gamma w)^{\beta} \ln(\gamma w) \times \exp(-(\gamma w)^{\beta}) \times \exp(y^2/2)$$

$$a_{12} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \mu \partial \sigma} \quad = 0 \tag{C.47}$$

$$a_{13} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \mu \partial \rho} \quad = 0 \tag{C.48}$$

$$a_{14} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \mu \partial \gamma} \quad = \frac{\rho}{1-\rho^2} E \quad \frac{\partial y}{\partial \gamma} \quad \frac{1}{\sigma} \tag{C.49}$$

$$a_{15} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \mu \partial \beta} \quad = \frac{\rho}{1-\rho^2} E \quad \frac{\partial y}{\partial \beta} \quad \frac{1}{\sigma} \tag{C.50}$$

$$a_{23} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \sigma \partial \rho} \quad = \frac{-\rho}{\sigma(1-\rho^2)} \tag{C.51}$$

$$a_{24} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \sigma \partial \gamma} \quad = \frac{\rho^2}{\sigma(1-\rho^2)} E \quad y \frac{\partial y}{\partial \gamma} \tag{C.52}$$

$$a_{25} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \sigma \partial \beta} \quad = \frac{\rho^2}{\sigma(1-\rho^2)} E \quad y \frac{\partial y}{\partial \beta} \tag{C.53}$$

$$a_{34} = -E \quad \frac{\partial^2 \ln(f(x, w))}{\partial \rho \partial \gamma} \quad = \frac{\rho}{1-\rho^2} E \quad y \frac{\partial y}{\partial \gamma} \tag{C.54}$$

$$a_{35} = -E \quad \frac{\partial^2 \ln(f(x,w))}{\partial\rho\partial\beta} \quad = \frac{\rho}{1-\rho^2} E \quad y\frac{\partial y}{\partial\beta} \tag{C.55}$$

$$a_{45} = -E \quad \frac{\partial^2 \ln(f(x,w))}{\partial\gamma\partial\beta} \quad = \frac{\rho^2}{1-\rho^2} E \quad \frac{\partial y}{\partial\gamma}\frac{\partial y}{\partial\beta} \quad + \frac{\beta}{\gamma} E\left(\log(w)\right)$$
$$+\frac{1}{\gamma} + \frac{\beta\ln(\gamma)}{\gamma} \tag{C.56}$$

We know that the expectations above involving partial derivatives of y exist and are finite by work done in Appendix H of Verrill et al. (2012a). To calculate approximations to most of the expectations above—$E((\partial y/\partial\gamma)^2)$, $E\left((\partial y/\partial\beta)^2\right)$, $E\left((\partial y/\partial\gamma)(\partial y/\partial\beta)\right)$, $E\left(y\,(\partial y/\partial\gamma)\right)$, $E\left(y\,(\partial y/\partial\beta)\right)$, $E\left(\partial y/\partial\gamma\right)$, and $E\left(\partial y/\partial\beta\right)$—one can use, for example, the QUADPACK numerical integration routine dqags. $E\left((\ln(w))^2\right)$ and $E\left(\ln(w)\right)$ are related to the Euler–Mascheroni constant (see Eqs. (17)–(19) of Verrill et al. (2012a)) and can be calculated from it.

Appendix D—Positive Definite Information Matrix

To invoke Lehmann's Theorem 4.2, we need to establish that the information matrix is positive definite. In Appendices E2 and E3 of Verrill et al. (2012a), we establish that

$$E \quad -\frac{\partial^2 \ln(f(x,w))}{\partial\theta_i\partial\theta_j} \quad = E \quad \frac{\partial f/\partial\theta_i}{f} \times \frac{\partial f/\partial\theta_j}{f} \tag{D.57}$$

Thus,

$$\begin{aligned}\mathbf{a}^T\mathbf{I}(\theta)\mathbf{a} &= \sum_{i=1}^{5}\sum_{j=1}^{5} a_i a_j E \quad \frac{\partial f/\partial\theta_i}{f} \times \frac{\partial f/\partial\theta_j}{f} \\ &= E\left(\quad \sum_{i=1}^{5} a_i \frac{\partial f/\partial\theta_i}{f} \quad ^2\right) \geq 0.\end{aligned} \tag{D.58}$$

To complete the proof that $\mathbf{I}(\theta)$ is positive definite we need to show that

$$\sum_{i=1}^{5} a_i \frac{\partial f/\partial\theta_i}{f} = 0 \text{ a.e.} \tag{D.59}$$

implies $\mathbf{a} = \mathbf{0}$. From result (172) of Verrill et al. (2012a) we have

$$\sum_{i=1}^{5} a_i \frac{\partial f/\partial\theta_i}{f} = a_1 \times \quad \frac{1}{\sigma}\frac{\left(\frac{x-\mu}{\sigma} - \rho y\right)}{1-\rho^2} \tag{D.60}$$

$$+ a_2 \times \quad \frac{-1}{\sigma} + \frac{1}{\sigma}\frac{\left(\frac{x-\mu}{\sigma} - \rho y\right)}{1-\rho^2} \quad \frac{x-\mu}{\sigma}$$

$$+ a_3 \times \quad \frac{\rho}{1-\rho^2} + \frac{\left(\frac{x-\mu}{\sigma} - \rho y\right) y}{1-\rho^2} - \frac{\left(\frac{x-\mu}{\sigma} - \rho y\right)^2 \rho}{(1-\rho^2)^2}$$

$$+ a_4 \times \quad \frac{\beta}{\gamma} - w^{\beta}\beta\gamma^{\beta-1} + \quad \frac{x-\mu}{\sigma} - \rho y \quad \frac{\rho}{1-\rho^2}\frac{\partial y}{\partial \gamma}$$

$$+ a_5 \times \quad \ln\gamma + \frac{1}{\beta} + \ln(w) - (\gamma w)^{\beta}\ln(\gamma w)$$

$$+ \quad \frac{x-\mu}{\sigma} - \rho y \quad \frac{\rho}{1-\rho^2}\frac{\partial y}{\partial \beta}$$

From result (137) of Verrill et al. (2012a), we have

$$\frac{\partial y}{\partial \gamma} = \beta\gamma^{\beta-1} \times w^{\beta} \times \exp(-(\gamma w)^{\beta})/\phi(y). \tag{D.61}$$

From result (138) of Verrill et al. (2012a), we have

$$\frac{\partial y}{\partial \beta} = (\gamma w)^{\beta}\ln(\gamma w) \times \exp(-(\gamma w)^{\beta})/\phi(y) \tag{D.62}$$

Recall that

$$y \equiv \Phi^{-1}(1 - \exp(-(\gamma w)^{\beta})$$

Now let $\epsilon > 0$ be given. Then results (59)–(62) imply that given any w_0, we can find an associated x, w rectangle chosen so that $(x-\mu)/\sigma - \rho y$ is small in the rectangle such that

$$\Bigg| a_2 \times \quad \frac{-1}{\sigma} \quad + a_3 \times \quad \frac{\rho}{1-\rho^2} \tag{D.63}$$

$$+ a_4 \times \quad \frac{\beta}{\gamma} - w^{\beta}\beta\gamma^{\beta-1}$$

$$+ a_5 \times \quad \ln\gamma + \frac{1}{\beta} + \ln(w) - (\gamma w)^{\beta}\ln(\gamma w) \quad \Bigg| < \epsilon/2$$

for some (x, w) in the rectangle.

A suitable rectangle can be written as $[x_0 - \delta, x_0 + \delta] \times [w_0 - \delta, w_0 + \delta]$ where δ can be made arbitrarily small, $(x_0 - \mu)/\sigma - \rho y_0 = 0$, and $y_0 = \Phi^{-1}(1 - \exp(-(\gamma w_0)^{\beta})$. By (59), there must be some (x, w) in the rectangle for which $\sum_{i=1}^{5} a_i \frac{\partial f/\partial\theta_i}{f} = 0$.

Taking w_0 large enough

$$|a_4 + a_5 K \ln(\gamma w)| < \epsilon \tag{D.64}$$

for K fixed and positive and w arbitrarily large. As ϵ was arbitrary, this implies that a_4 and a_5 equal 0.

Now, given results (59) and (60) and the fact that $a_4 = a_5 = 0$, given any $\epsilon > 0$, we can find an x, w region of positive measure (chosen so that y is large and $(x-\mu)/\sigma$ is bounded) such that (taking y large enough)

$$\Bigg| a_3 \times \quad \frac{-\rho}{1-\rho^2} - \frac{\rho^3}{(1-\rho^2)^2} \quad \Bigg| < \epsilon. \tag{D.65}$$

This implies that $a_3 = 0$ or $\rho = 0$. If $\rho = 0$, then (given that $a_4 = a_5 = 0$)

$$\sum_{i=1}^{5} a_i \frac{\partial f/\partial\theta_i}{f} = a_1 \times \left(\frac{1}{\sigma}\frac{x-\mu}{\sigma}\right) + a_2 \times \left(\frac{-1}{\sigma} + \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$+ a_3 \times \left(\frac{x-\mu}{\sigma} y\right). \tag{D.66}$$

Given results (59) and (66), given any $\epsilon > 0$, we can find an x, w region of positive measure (chosen so that y is large and $(x-\mu)/\sigma$ is bounded above and bounded below away from 0) such that (taking y large enough)

$$\left| a_3 \times \left(\frac{x-\mu}{\sigma}\right) \right| < \epsilon$$

for arbitrary $(x-\mu)/\sigma$ in the bounded region. Thus, $a_3 = 0$.

Next, given results (59) and (60) and the fact that $a_3 = a_4 = a_5 = 0$, given any $\epsilon > 0$, we can find an x, w region of positive measure (chosen so that $(x-\mu)/\sigma$ is large and y is bounded) such that (letting x get large enough)

$$\left| a_2 \times \frac{1}{\sigma(1-\rho^2)} \right| < \epsilon. \tag{D.11}$$

This implies that $a_2 = 0$.

Finally, results (59) and (60) and the fact that $a_2 = a_3 = a_4 = a_5 = 0$ imply that $a_1 = 0$, or $\mathbf{a} = 0$ as needed.