99-24

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

CLOSED FORM INTEGRATION OF ARTIFICIAL NEURAL NETWORKS WITH SOME APPLICATIONS TO FINANCE

BY

ANDREAS GOTTSCHLING

CHRISTIAN HAEFKE

AND

HALBERT WHITE

DISCUSSION PAPER 99-24
NOVEMBER 1999

Closed Form Integration of Artificial Neural Networks With Some Applications to Finance

Andreas Gottschling* **Christian Haefke†**
Halbert White†

November 30, 1999

Abstract

Many economic and econometric applications require the integration of functions lacking a closed form antiderivative, which is therefore a task that can only be solved by numerical methods. We propose a new family of probability densities that can be used as substitutes and have the property of closed form integrability. This is especially advantageous in cases where either the complexity of a problem makes numerical function evaluations very costly, or fast information extraction is required for time-varying environments. Our approach allows generally for nonparametric maximum likelihood density estimation and may thus find a variety of applications, two of which are illustrated briefly:

- Estimation of *Value at Risk* based on approximations to the density of stock returns.
- Recovering risk neutral densities for the valuation of options from the option price – strike price relation.

JEL Classification: C45, G13, C63;
Keywords: Option Pricing, Neural Networks, Nonparametric Density Estimation;

* Deutsche Bank Research, Grosse Gallusstrasse 10-14, 60311 Frankfurt, Germany, E-mail: andreas.gottschling@db.com

† Department of Economics, University of California, San Diego, La Jolla, CA 92093 - 0508, E-mail: chaefke@ucsd.edu

1 Introduction

Integrals of particular functions play a central role in economics, econometrics, and finance. For example, the notion of Value at Risk used to assess portfolio risk exposure is defined in terms of an integral of the probability density function (pdf) of portfolio returns. As another example, the price of a European call option can be expressed in terms of an integral of the cumulative distribution function (cdf) of risk neutralized asset returns. For reasons of familiarity and theoretical convenience, the normal distribution (or distributions derived from the normal, such as the log-normal) plays a central role in such analyses. Nevertheless, the normal distribution does not provide an empirically plausible basis for describing asset or portfolio returns, nor is it analytically tractable; neither the normal probability density nor the normal cdf have closed form integrals.

Here we provide a family of probability density functions that contains the normal or log-normal densities as limiting cases, but which are both more plausible empirically because of their much greater flexibility and more tractable analytically, possessing closed form expressions for their integrals (cdf's), and for integrals of their cdfs. In special cases, the inverse cdf (quantile function) also has a closed form expression, especially convenient for analyzing Value at Risk. We gain flexibility by constructing our family as the output of a single hidden layer artificial neural network; upon normalization, the output is a cdf or pdf of a particular mixture distribution. Analytic tractability arises from careful choice of the hidden unit activation function. Because of their flexibility and tractability, our new family of densities may be broadly useful for econometric analysis of economic and financial data.

Section 2 provides a brief discussion of artificial neural networks; section 3 discusses the main results, which are then applied in section 4 to Value at Risk estimation and in section 5 to recovering risk neutral densities from call option price data. Several appendices contain certain mathematical details and proofs of all results.

2 Artificial Neural Networks

Artificial neural networks (ANNs) have emerged as a prominent class of flexible functional forms for nonlinear function approximation. A leading case is the single hidden layer feedforward neural network, written as:

$$f(\mathbf{x}, \beta, \gamma) = \sum_{j=1}^{q} \beta_j \cdot g\left(\tilde{\mathbf{x}}^T \gamma_j\right), \tag{1}$$

with $\tilde{\mathbf{x}} = (1, x_1, x_2, \ldots, x_r)$, $\gamma = \left(\gamma_1^T, \gamma_2^T, \ldots, \gamma_q^T\right)^T$, $\gamma_j \in R^{r+1}$ and $\beta = \left(\beta_1^T, \ldots, \beta_q^T\right)^T$. See Kuan and White (1994) for additional background. When, for some finite non-negative integer ℓ, g is ℓ-finite, that is, g is continuously differentiable of order ℓ and has Lebesgue integrable ℓ^{th} derivative, then functions of the form (1) are able to approximate large classes of functions (and their derivatives) arbitrarily well, as shown by Hornik, Stinchcombe, and White (1990) (HSW). A common choice for g is that it be a given cdf; the logistic cdf is the leading choice. We shall pay particular attention to the case in which g is a pdf, so that its integral is a cdf. Imposing the constraint $\sum_{j=1}^{q} \beta_j = 1, \beta_j \geq 0$ when g is a density permits us to interpret (1) as a mixture density with weights β_j. Such mixtures can approximate arbitrary densities as shown by White (1996, theorem 19.1). The additive form of (1) not only delivers flexibility, but it also provides the foundation for analytic tractability: the properties of the integral of f depend solely on the properties of the integral of g.

Note that we view g as a univariate pdf, but that its argument is the linear combination $\tilde{\mathbf{x}}^T \gamma_j$. For the moment suppose that $r = 1$, so $\tilde{\mathbf{x}}^T \gamma_j = \gamma_{j0} + \gamma_{j1} \cdot x_1$. We therefore allow x_1 to be scaled and shifted inside g so that $f(\mathbf{x}, \beta, \gamma)$ can be viewed as a mixture of univariate pdf's in the usual way. On the other hand, if $r > 1$ we can view $f(\mathbf{x}, \beta, \gamma)$ as a conditional density for one of the elements of $\mathbf{x}$, say x_1, given the rest: $x_2, \ldots, x_r$. The use of the linear transformation $\tilde{\mathbf{x}}^T \gamma_j$ can be seen as permitting scaling and shifting as before, but with the shift now incorporating conditioning effects of the form $\gamma_{j0} + \sum_{i=2}^{r} x_i \gamma_{ji}$. Thus, we view g and f as pdf's for a particular random variable, though possibly conditional on other random variables. Treatment of multivariate densities in a framework analogous to that proposed here is possible but is beyond our present scope and is accordingly deferred.

We now turn our attention to choosing g in a way that delivers the desired closed form expressions for the integral of g.

3 A Family of Density Functions

To motivate our new family of probability density functions consider the transformation:

$$T_\lambda(x) = \frac{1}{\lambda} \ln(\lambda x + 1) \quad x > 0, \quad 0 < \lambda \leq 1. \tag{2}$$

We have $T_\lambda(x) \to x$ as $\lambda \to 0$. $T_\lambda(x)$ is the logarithm of the inverse Box – Cox (1964) transformation. To see this, consider the standard Box - Cox transformation given by

$$B_\lambda(\omega) = \frac{\omega^\lambda - 1}{\lambda} \quad 0 < \lambda \leq 1,$$

which converges to the natural logarithm as $\lambda \to 0$. The inverse Box - Cox transformation is thus

$$B_\lambda^{-1}(x) = (\lambda x + 1)^{\frac{1}{\lambda}},$$

which converges to the exponential function as $\lambda \to 0$. Taking natural logarithms gives

$$T_\lambda(x) = \ln B_\lambda^{-1}(x) = \frac{1}{\lambda} \ln(\lambda x + 1).$$

Our new family is based on the logarithm of the inverse power Box - Cox transformation, given by

$$\mathcal{P}_{\lambda,\zeta}(\omega) = \frac{\omega^{\left(\frac{\lambda(1-\lambda)}{1-\lambda^{1+\zeta}}\right)} - 1}{\lambda} \tag{3}$$

for non-negative integer ζ. (We let Z_0^+ denote the set of non-negative integers.) This reduces to the standard Box - Cox transformation for $\zeta = 0$.

Lemma 1 *Let $\mathcal{P}_{\lambda,\zeta}(\omega)$ be as defined in (3). Then for all $\zeta \in Z_0^+$*

$$\lim_{\lambda \to 0} \mathcal{P}_{\lambda,\zeta}(\omega) = \ln(\omega)$$

and

$$\lim_{\lambda \to 1} \mathcal{P}_{\lambda,\zeta}(\omega) = \omega - 1 \, .$$

By inverting the power Box Cox and taking natural logarithms we obtain our extension of (2):

$$\mathcal{T}_{\lambda,\zeta}(x) = \frac{1-\lambda^{1+\zeta}}{\lambda(1-\lambda)} \ln\left(\lambda x + 1\right) \quad \lambda \in (0,1), \quad \zeta \in Z_0^+ \, . \tag{4}$$

3.1 Analogs of the Normal Distribution

The normal distribution with pdf

$$\phi(x) \equiv \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

plays a central role in economics, econometrics, and finance. Nevertheless, for any function of the form

$$f(x) = a \exp\left(bx^{2n} + c\right)$$

for real a, b, c, and natural number n, no closed form expressions for the antiderivatives exist (Magid 1994). Consider, however, the replacement of x^2 in the exponential component of the normal density by its log inverse power Box - Cox transform

$$\mathcal{T}_{\lambda,\zeta}(x^2) = \frac{1-\lambda^{1+\zeta}}{\lambda(1-\lambda)} \ln\left(\lambda x^2 + 1\right) \; ,$$

which yields

$$\begin{aligned} \tilde{h}_{\lambda,\zeta}(x) &= \exp\left\{-\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} \ln\left(\lambda x^2 + 1\right)\right\} \quad \text{for} \quad \lambda \in (0,1) \\ &= \left(\lambda x^2 + 1\right)^{-\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}} \; . \end{aligned}$$

As we will show, this provides the basis for a family of densities having closed form expressions for its antiderivatives. Clearly, $\tilde{h}_{\lambda,\zeta}$ is a symmetric function of x. Further, note that for $\zeta = 1$ we obtain the well-known t distribution (Student 1908) with $1/\lambda$ degrees of freedom, for which the pdf can be written

$$t(x) = \frac{\Gamma\left(\frac{\lambda+1}{\lambda}\right)}{\Gamma\left(\frac{1}{2\lambda}\right)} \sqrt{\frac{\lambda}{\pi}} \left(\lambda x^2 + 1\right)^{-\frac{1}{2}\left(1+\frac{1}{\lambda}\right)},$$

where Γ is the Euler gamma function.

Our first result provides conditions on λ under which $\tilde{h}_{\lambda,\zeta}$ is integrable, so that with suitable normalization, $\tilde{h}_{\lambda,\zeta}$ is a density.

Theorem 2 *Let $\tilde{h}_{\lambda,\zeta}$ be as defined above. Then for all $\zeta \in Z_0^+$ and all $0 < \lambda < 1$*

$$\kappa_{\lambda,\zeta} \equiv \int_{-\infty}^{\infty} \tilde{h}_{\lambda,\zeta}(x)dx = \frac{\Gamma\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \frac{1}{2}\right)}{\Gamma\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}\right)} \sqrt{\frac{\pi}{\lambda}} < \infty.$$

We can now define the density function

$$\boldsymbol{h}_{\lambda,\zeta} = \kappa_{\lambda,\zeta}{}^{-1} \tilde{h}_{\lambda,\zeta}. \tag{5}$$

As noted previously $\boldsymbol{h}_{\lambda,\zeta}$ contains Student's t-distribution. $\boldsymbol{h}_{\lambda,\zeta}$ is a Pearson distribution of Type VII (Kendall and Stuart 1977), i.e.

$$\frac{df}{dx} = \frac{(x-a)f}{b_0 + b_1 x + b_2 x^2},$$

with $a = 0$, $b_1 = 0$, $b_0 > 0$, and $b_2 > 0$. For general properties of Pearson distributions the reader is referred to Kendall and Stuart (1977), chapter 6. $\boldsymbol{h}_{\lambda,\zeta}$ is also a special case of the generalized beta distribution proposed by McDonald (1984).

Under further restrictions on λ, $\boldsymbol{h}_{\lambda,\zeta}$ has finite $m-th$ moment:

Theorem 3 *Let $\boldsymbol{h}_{\lambda,\zeta}$ be as in (5). Then for $m > 0$:*

$$\int_{-\infty}^{\infty} |x|^m \tilde{h}_{\lambda,\zeta}(x)dx < \infty$$

for all $0 < \lambda < \frac{1}{1+m}$, $\zeta \in Z_0^+$.

Furthermore, for these values of λ, closed form expressions for the moments are given by:

$$\int_{-\infty}^{\infty} x^m \boldsymbol{h}_{\lambda,\zeta}(x)dx = \begin{cases} 0 & m \quad odd \\ \lambda^{-\frac{m}{2}} \frac{\Gamma\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \frac{m+1}{2}\right)\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \frac{1}{2}\right)\sqrt{\pi}} & m \quad even. \end{cases}$$

Thus, $\boldsymbol{h}_{\lambda,\zeta}$ has finite first moment for $\lambda < \frac{1}{2}$, finite second moment for $\lambda < \frac{1}{3}$, and so on.

As desired, $\boldsymbol{h}_{\lambda,\zeta}$ approaches the normal as a limiting case. In fact, the convergence is uniform.

Theorem 4 *Let $\boldsymbol{h}_{\lambda,\zeta}$ be as in (5). Then for each $\zeta \in Z_0^+$ $\boldsymbol{h}_{\lambda,\zeta}$ converges to ϕ uniformly as $\lambda \to 0$. Accordingly we define $h_{0,\zeta} \equiv \phi$.*

Figure (1) presents a plot of the densities for $\zeta = 0$ and various values of λ compared to the normal density.

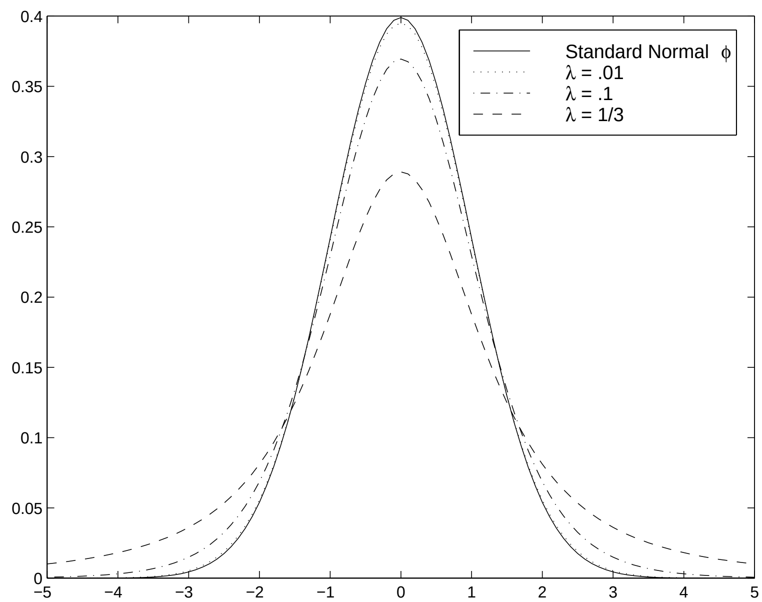

Figure 1: $\boldsymbol{h}_{\lambda,0}$ and the normal density

Now we consider the antiderivatives of $\boldsymbol{h}_{\lambda,\zeta}$. For a scalar function f of x, we write the first derivative as $Df = \frac{df}{dx}$. The antiderivative $D^{-1}f$ is such that $D(D^{-1}f) = f$. In forming the antiderivative, the "constant of integration" is here always taken to be zero. In the multivariate case, we denote partial derivatives as

$$D^{\alpha} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1}, \partial x_2^{\alpha_2}, \cdots, \partial x_r^{\alpha_r}},$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ is a multi-index, i.e. a vector of non-negative integers, and $|\alpha| = \sum_{i=1}^{r} |\alpha_i|$ is the magnitude of α. The corresponding antiderivative $D^{-\alpha} f$ is such that $D^{\alpha}(D^{-\alpha} f) = f$. In what follows, we often use the notation D^{-e_i}, which denotes the (first) antiderivative with respect to the i^{th} variable. Here e_i is the unit vector with a 1 in the i^{th} position and 0's elsewhere. As we are interested here only in derivatives with respect to x and not λ, we shall understand D, D^{-1}, D^{α}, $D^{-\alpha}$ to refer solely to derivatives or antiderivatives with respect to x.

In stating our result for the antiderivative of $\boldsymbol{h}_{\lambda,\zeta}$, we make use of the hypergeometric function ${}_2F_1$. This function is defined for complex a, b, c, and z as the analytic continuation in z of the hypergeometric series

$$ {}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)}{\Gamma(c+k)} \frac{z^k}{k!} \ . \tag{6} $$

The series converges absolutely for $|z| < 1$, as a ratio test shows. In our applications, we are interested in the hypergeometric function for any real x. For this, we make use of the transformation $z = \frac{\lambda x^2}{1+\lambda x^2}$ which yields $|z| < 1$ (the derivation is given in Appendix A). Additional useful background can be found in Bailey (1962) and Abramowitz and Stegun (1965).

Theorem 5 *Let $\boldsymbol{h}_{\lambda,\zeta}$ be as in (5). Then for all $x \in R$, $\zeta \in Z_0^+$, and $0 < \lambda < 1$*

$$ D^{-1}\boldsymbol{h}_{\lambda,\zeta}(x) = \frac{1}{2} + \frac{x}{\kappa_{\lambda,\zeta}\sqrt{(1+\lambda x^2)}} \cdot {}_2F_1\left(\frac{1}{2}, \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda\,(1-\lambda)}; \frac{3}{2}; \frac{\lambda x^2}{\lambda x^2+1}\right) \ . \tag{7} $$

To obtain a closed form solution one needs to reduce the hypergeometric series to a finite polynomial, which can be achieved for $\zeta = 0$ by choosing the appropriate λ:

Corollary 6 *For $\zeta = 0$, $D^{-1}\boldsymbol{h}_{\lambda,\zeta}(x)$ has a closed form expression for all λ of the form $\lambda_n = \frac{1}{2n+3}$, $n = 0, 1, 2, \ldots$.*

For general ζ the choice of λ is given by the solution of an algebraic equation of order ζ in λ, which for $\zeta > 4$ does not necessarily possess solutions that allow a convenient expression for λ as a function of its coefficients (Artin (1973), Theorem 45). Although the cases $0 \leq \zeta \leq 4$ can all be be handled straightforwardly, we focus particular attention on the $\zeta = 0$ case for the sake of convenience and simplicity.

Since the resulting expression depends on the normalization factor $\kappa_{\lambda,\zeta}$, which in turn is a function of λ, the following corollary provides a convenient method to calculate $\kappa_{\lambda,0}$.

Corollary 7 *For $\zeta = 0$ and λ as in Corollary 6 the normalization factor $\kappa_{\lambda,\zeta}$ is given by*

$$ \kappa_{\lambda_n} = \frac{n!2^{2n+1}}{(2n+1)!}\sqrt{2n+3} \ . $$

Note that no upper limit is imposed upon n; hence we can find arbitrarily close approximations to the normal pdf, all having closed form integrals. Since the polynomial expansion is proportional to the magnitude of n, some simple solutions with $\zeta = 0$ are given in Table 1.

Because of the central role played by the hypergeometric function in defining the properties of our family of analogs to the normal, we call the family

λ	$\kappa_{\lambda,0}$	$D^{-1}\boldsymbol{h}_{\lambda,0}$
$1/3$	$2\sqrt{3}$	$1/2+\frac{x}{2\cdot\sqrt{3+x^2}}$
$1/5$	$\frac{4}{3}\sqrt{5}$	$1/2+\frac{15x^3+2x^5}{4\sqrt{(5+x^2)}^3}$
$1/7$	$\frac{16}{15}\sqrt{7}$	$1/2+\frac{735x+140x^3+8x^5}{16\sqrt{(7+x^2)}^5}$

Table 1: Simple choices for λ.

$\{\boldsymbol{h}_{\lambda,\zeta}, 0 \leq \lambda < 1, \zeta \in Z_0^+\}$ the *"hypernormal"* family. We say that $\boldsymbol{h}_{\lambda,\zeta}$ is hypernormal with index λ, ζ.

The second antiderivative, $D^{-2}\boldsymbol{h}_{\lambda,\zeta}$, is also of interest. For example, suppose a risk manager requires to know the expected value of returns *given* that the portfolio value has fallen below the Value at Risk (VaR). If returns have the density $\boldsymbol{h}_{\lambda,\zeta}$, then this conditional expectation has the form $\int_{-\infty}^{a} x\boldsymbol{h}_{\lambda,\zeta}(x)dx / \int_{-\infty}^{a} \boldsymbol{h}_{\lambda,\zeta}(x)dx$, where a is an appropriate constant depending on the VaR. Applying integration by parts to the numerator, we obtain:

$$\begin{aligned}\int_{-\infty}^{a} x \cdot \boldsymbol{h}_{\lambda,\zeta}(x)dx &= a \cdot D^{-1}\boldsymbol{h}_{\lambda,\zeta}(a) - \int_{-\infty}^{a} D^{-1}\boldsymbol{h}_{\lambda,\zeta}(x)dx \\ &= a \cdot D^{-1}\boldsymbol{h}_{\lambda,\zeta}(a) - D^{-2}\boldsymbol{h}_{\lambda,\zeta}(a)\ ,\end{aligned}$$

where we use the fact that $D^{-1}\boldsymbol{h}_{\lambda,\zeta}(-\infty) = D^{-2}\boldsymbol{h}_{\lambda,\zeta}(-\infty) = 0$. The second antiderivative is given by our next result.

Theorem 8 *Let $\boldsymbol{h}_{\lambda,\zeta}$ be as in (5). Then for all $x \in R$, $\zeta \in Z_0^+$ and $0 < \lambda < 1$*

$$D^{-2}\boldsymbol{h}_{\lambda,\zeta}(x) \quad = \quad \frac{x}{2} + \frac{\sqrt{(1+\lambda x^2)}}{2\lambda\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}-1\right)\kappa_{\lambda,\zeta}} \cdot {}_2F_1\left(\frac{-1}{2}, \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}; \frac{1}{2}; \frac{\lambda x^2}{\lambda x^2+1}\right).$$

For $\zeta = 0$, these functions also have a closed form expression for all λ of the form $\lambda_n = \frac{1}{2n+3}$.

Now that we have an analytically tractable subfamily of densities, we add flexibility by forming artificial neural networks of the form (1). Our next result shows that these networks can deliver arbitrarily accurate approximations to a large class of densities, regardless of the value specified for λ.

Theorem 9 *Let f belong to the Sobolev space $S_\infty^m(\chi)$ where χ is an open, bounded subset of R^r. Elements of this space are functions with continuous derivatives of order m on the domain χ which satisfy*

$$||f||_{m,\infty,\chi} \equiv max_{n\leq m} sup_{x\in\chi} |D^n f(x)| < \infty \tag{8}$$

for some integer $m \geq 0$ (for further background see (Gallant and White 1992)). For integer $\ell < 1/\lambda - 1$, $\boldsymbol{h}_{\lambda,\zeta}$ is ℓ - finite. Then for all $m \leq \ell$, f can be *approximated as closely as desired in $S^m_\infty(\chi)$ equipped with metric (8) using a single hidden layer feedforward network of the form*

$$\psi_{\lambda,\zeta}\left(x,\theta\right)=\sum_{j=1}^{q}\beta_j\cdot\boldsymbol{h}_{\lambda,\zeta}\left(\tilde{x}^T\gamma_j\right), \tag{9}$$

where $\tilde{x}=(1,x)$, and q is sufficiently large.

Observe that $\boldsymbol{h}_{\lambda,\zeta}$ is always 0-finite by construction.

Corollary 10 *Let $H_{\lambda,\zeta}=D^{-e_i}\boldsymbol{h}_{\lambda,\zeta}$ denote the antiderivative of $\boldsymbol{h}_{\lambda,\zeta}$ with respect to the i-th variable, and let $l\le u$ be real numbers. Then the integral of the neural net (9) has the form*

$$\int_l^u \psi_{\lambda,\zeta}(x,\theta)dx_i=\Psi_{\lambda,\zeta}(x_{(i)}(u);\theta)-\Psi_{\lambda,\zeta}(x_{(i)}(l);\theta),$$

where $x_{(i)}(a)$ is the vector obtained by replacing the i^{th} element x_i from the vector x with a, and

$$\Psi_{\lambda,\zeta}(x_{(i)}(a);\theta)=\sum_{j=1}^{q}\beta_j\cdot H_{\lambda,\zeta}(a_{ij}(x_{(i)}(a),\gamma_{ij})),$$

where

$$a_{ij}(x_{(i)}(a),\gamma_{ij})=a\gamma_{ij}+\sum_{k=1,k\neq i}^{r+1}\tilde{x}_k\gamma_{kj}.$$

Furthermore, for $\zeta=0$, $\Psi_{\lambda,\zeta}(x_{(i)}(a);\theta)$ has a closed form expression for all λ of the form $\lambda_n=\frac{1}{2n+3}$, $n=0,1,2,\ldots$.

Note that the transformed integration boundaries are different for each hidden unit because they depend on γ_{ij}.

The networks $\Psi_{\lambda,\zeta}$ of Corollary 10 have desirable approximation properties:

Theorem 11 *Let f and $\boldsymbol{h}_{\lambda,\zeta}$ be as in Theorem 9, and let $H_{\lambda,\zeta}$ be as in Corollary 10. Then for integer $\ell<1/\lambda$, $H_{\lambda,\zeta}$ is ℓ-finite and for all $m\le\ell$, f can be approximated as closely as desired in $S^m_\infty(\chi)$ equipped with metric (8) using a single hidden layer feedforward network of the form $\Psi_{\lambda,\zeta}(\cdot)$ given in Corollary 10.*

When f is a cdf, $\Psi_{\lambda,\zeta}$ can approximate it, and its derivative – the associated pdf – is approximated by the derivative $\psi_{\lambda,\zeta}$ of $\Psi_{\lambda,\zeta}$, due to the denseness in Sobolev norm and the fact that $\Psi_{\lambda,\zeta}$ is always 1-finite by construction.

We also have analogs of Corollary 10 and Theorem 11 for the integral of $\Psi_{\lambda,\zeta}$

Corollary 12 *Let $\Xi_{i,\lambda,\zeta} = D^{-2e_i} \boldsymbol{h}_{\lambda,\zeta}$ denote the second antiderivative of $\boldsymbol{h}_{\lambda,\zeta}$ with respect to the i-th variable. Let $l \leq u$ be real numbers. Then the integral*

$$\int_l^u \Psi_{\lambda,\zeta} \left(x_{(i)}(a); \theta\right) da$$

has the form

$$\begin{aligned} \int_l^u \Psi_{\lambda,\zeta} \left(x_{(i)}(a); \theta\right) da &= \Lambda_{i,\lambda,\zeta}(x_{(i)}(u); \theta) - \Lambda_{i,\lambda,\zeta}(x_{(i)}(l); \theta), \\ \textit{where} \quad \Lambda_{i,\lambda,\zeta}(x_{(i)}(b); \theta) &= \sum_{j=1}^{q} \Xi_{i,\lambda,\zeta}(b_{ij}(x_{(i)}(b); \gamma_{ij}) \\ \textit{with} \quad b_{ij}(x_{(i)}(b); \gamma_{ij}) &= b\gamma_{ij} + \sum_{k=1, k \neq i}^{r+1} \tilde{x}_k \gamma_{kj}. \end{aligned}$$

In addition, for $\zeta = 0$, $\Lambda_{i,\lambda,\zeta}$ has a closed form expression for all λ of the form $\lambda_n = \frac{1}{2n+3}$, $n = 0, 1, 2, \ldots$.

A similar result for $D^{-(e_i+e_j)} \boldsymbol{h}_{\lambda,\zeta}$ can be obtained, but as our focus here is on the univariate case, we omit that result.

Corollary 13 *Let f and $\boldsymbol{h}_{\lambda,\zeta}$ be as in Theorem 5, and let $\Xi_{i,\lambda,\zeta}$ be as in Corollary 12. Then for integer $\ell < 1/\lambda + 1$, $\Xi_{i,1\lambda,\zeta}$ is ℓ-finite and for all $m \leq \ell$, f can be approximated as closely as desired in $S_\infty^m(\chi)$ equipped with metric (8) using a single hidden layer feedforward network of the form $\Lambda_{i,\lambda,\zeta}$ given in Corollary 12.*

When f is the antiderivative of a cdf $\Lambda_{i,\lambda,\zeta}$ can approximate it, and its derivatives (the cdf and pdf) can be approximated by the derivatives of $\Lambda_{i,\lambda,\zeta}$ due to the denseness in Sobolev norm and the fact that the associated activation function is always 2-finite.

3.2 Analogs of the Lognormal Distribution

Economic theory may dictate nonnegativity: for example, prices are non-negative. In such cases the normal distribution is not necessarily a reasonable assumption for the data generating process. For example, to model asset prices the lognormal density is frequently used. The lognormal pdf is given by

$$\tilde{\phi}(\omega) = \frac{1}{\omega} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\ln \omega)^2}{2}\right).$$

The antiderivative of the lognormal is easily obtained. Just note that by substituting $x = \ln\omega$ we obtain

$$\begin{aligned}\int_l^u \tilde{\phi}(\omega)\,d\omega &= \int_l^u \frac{1}{\omega}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\ln\omega)^2}{2}\right)d\omega \\ &= \int_{\ln l}^{\ln u}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx \\ &= \Phi(\ln u) - \Phi(\ln l),\end{aligned}$$

which just involves the antiderivative Φ of the standard normal.

We can replace the normal density appearing here with our new hypernormal density to obtain results analogous to those of section 3.1. In particular, consider

$$\tilde{g}_{\lambda,\zeta}(\omega) = \frac{1}{\omega}\left(\lambda(\ln\omega)^2 + 1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}} = \frac{1}{\omega}\tilde{h}_{\lambda,\zeta}(\ln\omega).$$

This result allows us to obtain the following helpful analog of Theorem 2:

Theorem 14 *Let $\tilde{g}_{\lambda,\zeta}$ be defined as above. Then for all $\zeta \in Z_0^+$ and $0 < \lambda < 1$*

$$L_{\lambda,\zeta} \equiv \int_0^\infty \tilde{g}_{\lambda,\zeta}(\omega)d\omega = \kappa_{\lambda,\zeta} < \infty. \tag{10}$$

Hence we can now define the density function

$$\boldsymbol{g}_{\lambda,\zeta} = L_{\lambda,\zeta}{}^{-1}\tilde{g}_{\lambda,\zeta} \tag{11}$$

for $0 < \lambda < 1$ and $\zeta \in Z_0^+$.

We call this density the log - hypernormal by analogy to the log-normal. By substituting $x = \ln\omega$ we obtain for $0 < l < u < \infty$

$$\begin{aligned}\int_l^u \boldsymbol{g}_{\lambda,\zeta}(\omega)\,d\omega &= \int_{\ln l}^{\ln u}\boldsymbol{h}_{\lambda,\zeta}(x)\,dx \\ &= H_{\lambda,\zeta}(\ln u) - H_{\lambda,\zeta}(\ln l) \\ &\equiv G_{\lambda,\zeta}(u) - G_{\lambda,\zeta}(l).\end{aligned}$$

Thus whenever $H_{\lambda,\zeta}$ has a closed form expression, $G_{\lambda,\zeta}$ will also.

Corollary 15 *Let $\boldsymbol{g}_{\lambda,\zeta}$ be as in (12). Then for all $\omega \in R^+$, $\zeta \in Z_0^+$, and $0 < \lambda < 1$*

$$D^{-1}\boldsymbol{g}_{\lambda,\zeta}(\omega) = \frac{1}{2} + \frac{\ln\omega}{\kappa_{\lambda,\zeta}\sqrt{(1+\lambda(\ln\omega)^2}}\cdot {}_2F_1\left(\frac{1}{2}, \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}; \frac{3}{2}; \frac{\lambda(\ln\omega)^2}{\lambda(\ln\omega)^2+1}\right). \tag{12}$$

For $\zeta = 0$, $D^{-1}\boldsymbol{g}_{\lambda,\zeta}(\omega)$ has a closed form expression for all λ of the form $\lambda_n = \frac{1}{2n+3}$, $n = 0, 1, 2, \ldots$.

Hence a closed form analytic solution analogous to the normal case can be obtained for the log-hypernormal cumulative distribution ($\zeta = 0$) by choosing the appropriate λ. Unfortunately, no moments exist for the log-hypernormal nor can a convenient closed form be obtained for $D^{-1}G_{\lambda,\zeta}$.

Theorem 16 *Let $\boldsymbol{g}_{\lambda,\zeta}$ be as in (12). Then for all $\zeta \in Z_0^+$ and $\lambda = \frac{1}{2n+3}$*

$$\int_0^\infty |\omega|^m \, \boldsymbol{g}_{\lambda,\zeta}(\omega) d\omega = \infty$$

if and only if $m \geq 1$.

Theorem 17 *Let $\boldsymbol{g}_{\lambda,\zeta}$ be as in (12). Then for all $\omega \in R^+$, $\zeta \in Z_0^+$ and $\lambda = \frac{1}{2n+3}$*

$$D^{-2}\boldsymbol{g}_{\lambda,\zeta}(\omega) d\omega = \frac{\omega}{\kappa_{\lambda,\zeta}} \frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)})}$$

$$\left(\sum_{k=0}^{\infty} \frac{\Gamma(\frac{1}{2}+k)\Gamma(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}+k)}{\Gamma(\frac{3}{2}+k)} \frac{(-\lambda)^k}{k!} \sum_{i=0}^{2k+1} (-1)^{2k+1-i} \frac{(2k+1)!}{i!} (\ln x)^i \right).$$

Although the hypergeometric series ${}_2F_1$ terminates after n terms for the argument $\frac{\ln(\omega)^2}{1+\ln(\omega)^2}$ for specific values of λ as shown in Theorem 5, for the argument $-\ln(\omega)^2$ we do not get a similar termination. Intuitively (within the convergence radius and for appropriate choices of λ) the series of alternating terms of $-\ln(\omega)^2$ converges towards the same limit as the *finite* sum of the series with argument $\frac{\ln(\omega)^2}{1+\ln(\omega)^2}$ but the former series can not be reduced to a polynomial for some choice of λ. Hence, any true closed form solution for $D^{-1}G_{\lambda,\zeta}$ can not be obtained by these means.

Since closed form expressions for $D^{-1}G_{\lambda,\zeta}$ or $D^{-1}(1-G_{\lambda,\zeta})$ do not appear readily available, the previous results are not entirely satisfactory - for example, they do not deliver a closed form expression for pricing European call options.

To avoid excessive technicalities we only consider the case $\zeta = 0$ forthwith. Note, though, that the analysis carries through as in the hypernormal case for higher values of ζ subject to the solvability by radicals of algebraic equations in ζ of order 5 and above as previously noted.

The difficulty arises from the presence of $\ln\omega$ in the relation $G_\lambda(\omega) = H_{\lambda,\zeta}(\ln\omega)$. Consider, then, replacing $\ln\omega$ with the Box – Cox transform of $\omega > 0$,

$$B_\delta(\omega) = \frac{\omega^\delta - 1}{\delta} \quad 0 < \delta \leq 1,$$

which gives

$$G_{\lambda,\delta}(\omega) \simeq H_\lambda\left(B_\delta(\omega)\right) \equiv H_{\lambda,\delta}$$

(omitting the normalizing constant). For given $0 < \lambda < 1$, the associated derivative $h_{\lambda\delta} = DH_{\lambda\delta}$ is given by

$$\begin{aligned}\tilde{h}_{\lambda,\delta}(\omega) &= \omega^{\delta-1} h_{\lambda,\zeta}\left(B_\delta(a)\right) \\ &= \omega^{\delta-1}\left(\lambda\left(\frac{\omega^\delta-1}{\delta}\right)^2+1\right)^{\frac{-1}{2\lambda}}, \quad 0<\delta\leq 1. \end{aligned} \tag{13}$$

Our first result for $\tilde{h}_{\lambda,\delta}$ concerns the normalization factor, used to convert $\tilde{h}_{\lambda,\delta}$ into a density:

Theorem 18 *Let $\tilde{h}_{\lambda,\delta}$ be as defined above. If $0<\lambda<1$, then for all $0<\delta<1$ we have*

$$\kappa_{\lambda,\delta} \equiv \int_0^\infty \tilde{h}_{\lambda,\delta}(\omega)d\omega < \infty,$$

where $\kappa_{\lambda,\delta}$ can be calculated as:

$$\kappa_{\lambda,\delta} = \frac{1}{\sqrt{\lambda}}\cdot\frac{\Gamma\left(\frac{3}{2}\right)\cdot\Gamma\left(\frac{1-\lambda}{2\lambda}\right)}{\Gamma\left(\frac{1}{2\lambda}\right)} + \frac{\frac{1}{\delta}}{\sqrt{1+\frac{\lambda}{\delta^2}}}\cdot {}_2F_1\left(\frac{1}{2},\frac{3}{2}-\frac{1}{2\lambda};\frac{3}{2};\frac{\frac{\lambda}{\delta^2}}{1+\frac{\lambda}{\delta^2}}\right).$$

We can now define the density function

$$\boldsymbol{h}_{\lambda,\delta} = {\kappa_{\lambda,\delta}}^{-1}\tilde{h}_{\lambda,\delta}. \tag{14}$$

As our next result makes precise, for fixed λ and as δ ranges from 0 to 1 we have densities that range from analogs of the log-normal to analogs of the truncated normal (See Figures 2 – 5). When $\lambda = 0$ and $\delta = 0$ we have precisely the log normal, whereas $\lambda = 0$ and $\delta = 1$ give the normal truncated at 0 (the "half-normal"). Thus we can view δ as a parameter affecting the skewness of the associated distribution, whereas λ affects the kurtosis.

Theorem 19 *Let $\boldsymbol{h}_{\lambda,\delta}$ be as in (14). Then $\boldsymbol{h}_{\lambda,\delta}$ converges to $\phi(\cdot)1_{[\cdot\geq 0]}$ uniformly as $\lambda\to 0$, $\delta\to 1$ and $\boldsymbol{h}_{\lambda,\delta}$ converges to the log-normal $\tilde{\phi}$ uniformly as $\lambda\to 0$, $\delta\to 0$.*

Further analogs of our previous results for $h_{\lambda,\zeta}$ are as follows.

Theorem 20 *Let $\boldsymbol{h}_{\lambda,\delta}$ be as in (14). Let λ be any odd nonnegative integer and $t=\frac{m}{\delta}$ be any non-negative integer. With the substitution*

$$y = \omega^\delta$$

and using the notation

$$\Upsilon_{\lambda,\delta}(y,t) \equiv \frac{y^t}{\sqrt{R}^{2u+1}},$$

where

$$R = \lambda y^2 - 2\lambda y + \lambda + \delta^2$$

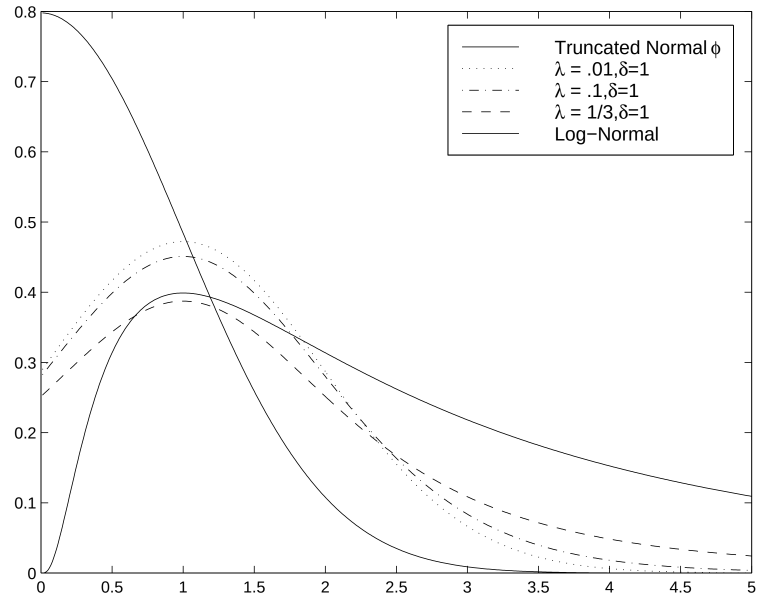

Figure 2: The densities of the log-normal, the truncated normal, and $g_{\lambda,\delta} = DG_{\lambda,\delta}$ for various λ, $\delta = 1$.

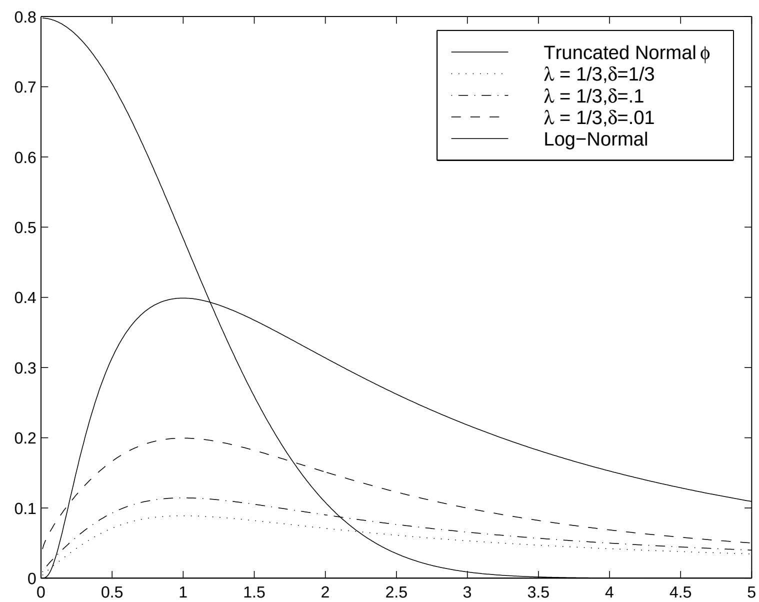

Figure 3: The densities of the log-normal, the truncated normal, and $g_{\lambda,\delta} = DG_{\lambda,\delta}$ for various δ, $\lambda = 1/3$.

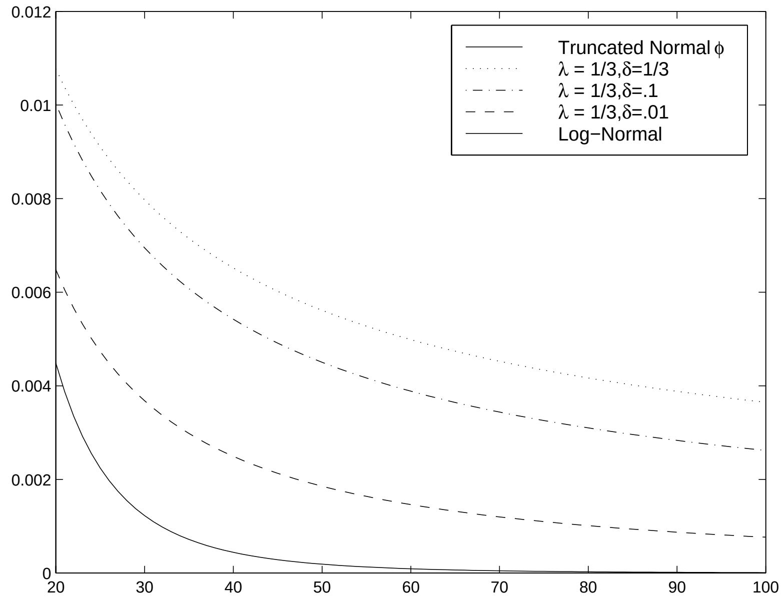

Figure 4: The extreme tail behavior of the log-normal, the truncated normal, and $g_{\lambda,\delta} = DG_{\lambda,\delta}$ for various δ, $\lambda = 1/3$.

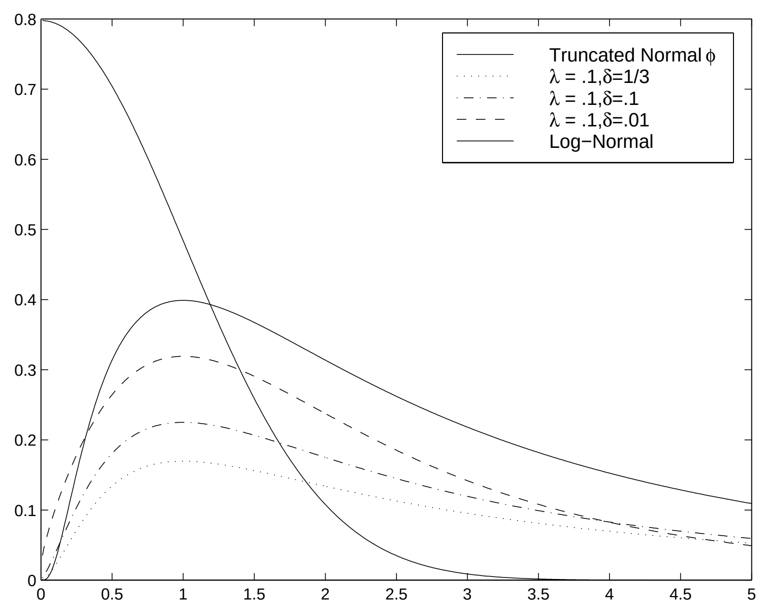

Figure 5: The densities of the log-normal, the truncated normal, and $g_{\lambda,\delta} = DG_{\lambda,\delta}$ for various δ, $\lambda = .1$.

for convenience, we have:

$$\int_0^\infty |\omega|^m \boldsymbol{h}_{\lambda,\delta}(\omega)d\omega = \frac{\delta^{\frac{1}{\lambda}-1}}{\kappa_{\lambda,\delta}} \int_0^\infty \Upsilon_{\lambda,\delta}(y,t)dy. \tag{15}$$

If $0 < \lambda < \frac{1}{1+m}$ *and* $\frac{\lambda \cdot m}{1-\lambda} < \delta < 1$ *then for all* $m > 0$*:*

$$\int_0^\infty |\omega|^m \boldsymbol{h}_{\lambda,\delta}(\omega)d\omega < \infty,$$

and the m-th moment can be calculated using the following recursion formula for the right-hand-side integral in 15:

$$\int_0^\infty \Upsilon_{\lambda,\delta}(y,t)dy = \xi_1(t)\int_0^\infty \Upsilon_{\lambda,\delta}(y,t-1)dy + \xi_2(t)\int_0^\infty \Upsilon_{\lambda,\delta}(y,t-2)dy,$$

where $\xi_1(t) = \frac{(2u-2t+1)}{(2u-t)}$ *and* $\xi_2(t) = \frac{(t-1)(\lambda+\delta^2)}{(2u-t)\lambda}$.
Each of the recursions terminates within t *steps yielding as the final term:*

$$\int_0^\infty \Upsilon_{\lambda,\delta}(y,0)dy = \frac{\left(1+\sum_{k=1}^{u-1}\prod_{i=1}^{k}\frac{(u-i)}{(2u-2i-1)}\cdot\left(\frac{2\lambda+2\delta^2}{\delta^2}\right)^k\right)}{(2u-1)\delta^2\left(\sqrt{\lambda+\delta^2}\right)^{2u-1}}.$$

Theorem 21 *Let* $\boldsymbol{h}_{\lambda,\delta}$ *be as in (14). Let* $0 < \lambda < \frac{1}{1+m}$*. For any* $0 < \delta < 1$ *and all* $x \in R^+$ *it holds that:*

$$D^{-1}\boldsymbol{h}_{\lambda,\delta}(\omega)d\omega = \frac{1}{2} + \frac{\frac{\omega^\delta-1}{\delta}}{\kappa_{\lambda,\delta}\sqrt{(1+\lambda\left(\frac{\omega^\delta-1}{\delta}\right)^2)}} \cdot {}_2F_1\left(\frac{1}{2},\frac{3}{2}-\frac{1}{2\lambda};\frac{3}{2};\frac{\lambda\left(\frac{\omega^\delta-1}{\delta}\right)^2}{\lambda\left(\frac{\omega^\delta-1}{\delta}\right)^2+1}\right). \tag{16}$$

Corollary 22 *For any* $0 < \delta < 1$, $D^{-1}\boldsymbol{h}_{\lambda,\delta}$ *has a closed form expression for all* λ *of the form* $\lambda_n = \frac{1}{2n+3}$, $n = 0, 1, 2, \ldots$.

The second antiderivative $D^{-2}\boldsymbol{h}_{\lambda,\delta}$ is also of interest because, for example, in option pricing the price of a European Call option with strike K and risk neutral cdf F can be shown to be given by $C_K = \int_K^\infty S(t)dt$, where $S(t) = 1 - F$.

Theorem 23 *Let* $\boldsymbol{h}_{\lambda,\delta}$ *be as in (14),* λ *be any odd nonnegative integer and* $\frac{1}{\delta}$ *be any non-negative integer. Then*

$$D^{-2}\boldsymbol{h}_{\lambda,\delta}(\omega) = \omega \cdot D^{-1}\boldsymbol{h}_{\lambda,\delta}(\omega) - \Upsilon_{\lambda,\delta}, \tag{17}$$

where

$$\Upsilon_{\lambda,\delta} = \frac{\delta^{\frac{1}{\lambda}}}{\kappa_{\lambda,\delta}} \int \frac{\omega^\delta}{\sqrt{\lambda\omega^{2\delta} - 2\lambda\omega^\delta + \lambda + \delta^2}^{\frac{1}{\lambda}}} d\omega. \tag{18}$$

Corollary 24 *Let $D^{-2}\boldsymbol{h}_{\lambda,\delta}(\omega)$, λ and δ be defined as in Theorem 23. If $\lambda = \frac{1}{2n+3}, n = 0, 1, 2, \ldots$ and $\frac{\lambda}{1-\lambda} < \delta < 1$ it holds for all $\omega \in R^+$ that $D^{-2}\boldsymbol{h}_{\lambda,\delta}(\omega)$ is finite and has a closed form expression.*

While these results provide analytic tractability, the next set of results provides flexibility, analogously to the normal case.

Theorem 25 *Let f belong to the Sobolev space $S^m_\infty(\chi)$ as defined in Theorem 9 where χ is an open, bounded subset of $(R^+)^r$. For integer $\ell < 1/\lambda - 1$, $\boldsymbol{h}_{\lambda,\delta}$ is ℓ-finite. Then for all $m \leq \ell$, f can be approximated as closely as desired in $S^m_\infty(\chi)$ equipped with metric (8) using a single hidden layer feedforward network of the form*

$$\psi_{\lambda,\delta}(\omega, \theta) = \sum_{j=1}^{q} \beta_j \cdot \boldsymbol{h}_{\lambda,\delta}\left(\tilde{\omega}^T \gamma_j\right), \tag{19}$$

where $\tilde{\omega} = (1, \omega)$, and q is sufficiently large.

Corollary 26 *Let $H_{\lambda,\delta} = D^{-e_i}\boldsymbol{h}_{\lambda,\delta}$ denote the antiderivative of $\boldsymbol{h}_{\lambda,\zeta}$ with respect to the i-th variable, and let $l \leq u$ be real numbers. Then the integral of the neural net (19) has the form*

$$\int_l^u \psi_{\lambda,\delta}(\omega, \theta) d\omega_i = \Psi_{\lambda,\delta}(\omega_{(i)}(u); \theta) - \Psi_{\lambda,\delta}(\omega_{(i)}(l); \theta),$$

where $\omega_{(i)}(a)$ is the vector obtained by replacing the i^{th} element ω_i from the vector ω with a, and

$$\Psi_{\lambda,\delta}(\omega_{(i)}(a); \theta) = \sum_{j=1}^{q} \beta_j \cdot H_{\lambda,\delta}(a_{ij}(\omega_{(i)}(a), \gamma_i)),$$

where

$$a_{ij}(\omega_{(i)}(a), \gamma_i) = a\gamma_{ij} + \sum_{k=1, k\neq i}^{r+1} \tilde{\omega}_k \gamma_{kj}.$$

Furthermore, $\Psi_{\lambda,\delta}(\omega_{(i)}(a); \theta)$ has a closed form expression for $\lambda = \frac{1}{2n+3}, n = 0, 1, 2, \ldots$ and $\delta = \frac{1}{n}, n = 1, 2, \ldots$ as long as $\frac{\lambda}{1-\lambda} < \delta < 1$.

Note that analogously to the normal case the transformed integration boundaries are different for each hidden unit because they depend on β_{iq}.

Theorem 27 *Let f and $\boldsymbol{h}_{\lambda,\delta}$ be as in Theorem 25, and let $H_{\lambda,\delta}$ be as in Corollary 26. Then for integer $\ell < 1/\lambda$, $H_{\lambda,\delta}$ is ℓ-finite and for all $m \leq \ell$, f can be approximated as closely as desired in $S^m_\infty(\chi)$ equipped with metric (8) using a single hidden layer feedforward network of the form $\Psi_{\lambda,\delta}(\cdot)$ given in Corollary 26.*

When f is a cdf, $\Psi_{\lambda,\delta}$ can approximate it, and its derivative – the associated pdf – is approximated by the derivative $\psi_{\lambda,\delta}$ of $\Psi_{\lambda,\delta}$, due to the denseness in Sobolev norm. We also have analogs of Corollary 26 and Theorem 27 for the integral of $\Psi_{\lambda,\delta}$.

Corollary 28 *Let $\Xi_{i,\lambda,\delta} = D^{-2e_i}\boldsymbol{h}_{\lambda,\delta}$ denote the second antiderivative of $\boldsymbol{h}_{\lambda,\delta}$ Let $l \leq u$ be real numbers. Then the integral*

$$\int_l^u \Psi_{\lambda,\delta}\left(\omega_{(i)}(a);\theta\right) da$$

has the form

$$\begin{aligned}
\int_l^u \Psi_{\lambda,\delta}\left(\omega_{(i)}(a);\theta\right) da &= \Lambda_{i,\lambda,\delta}(\omega_{(i)}(u);\theta) - \Lambda_{i,\lambda,\delta}(\omega_{(i)}(l);\theta), \\
\textit{where} \quad \Lambda_{i,\lambda,\delta}(\omega_{(i)}(b);\theta) &= \sum_{j=1}^{q} \Xi_{i,\lambda,\delta}(b_{ij}(\omega_{(i)}(b);\gamma_{ij}) \\
\textit{with} \quad b_{ij}(\omega_{(i)}(b);\gamma_i) &= b\gamma_{ij} + \sum_{k=1,k\neq i}^{r+1} \tilde{\omega}_k\gamma_{kj}.
\end{aligned}$$

In addition, $\Lambda_{i,\lambda,\delta}(\omega_{(i)}(b);\theta)$ has a closed form expression for $\lambda = \frac{1}{2n+3}$, $\delta = \frac{1}{n+1}, n = 0, 1, 2, \ldots$ as long as $\frac{\lambda}{1-\lambda} < \delta < 1$.

Theorem 29 *Let f and $\boldsymbol{h}_{\lambda,\delta}$ be as in Theorem 25, and let $\Xi_{i,\lambda,\delta}$ be as in Corollary 28. Then for integer $\ell < 1/\lambda + 1$, $\Xi_{i,\lambda,\delta}$ is ℓ-finite and for all $m \leq \ell$, f can be approximated as closely as desired in $S_\infty^m(\chi)$ equipped with metric (8) using a single hidden layer feedforward network of the form $\Lambda_{i,\lambda,\delta}$ given in Corollary 28.*

Thus, when f is the antiderivative of a cdf, $\Lambda_{i,\lambda,\delta}$ can approximate it, and its derivatives (the cdf and pdf) can be approximated by the derivatives of $\Lambda_{i,\lambda,\delta}$, due to the denseness in Sobolev norm. This approximation property provides an appealing basis for attempts to recover risk neutral densities from the option price - strike price relation for European call options.

4 Value at Risk

We first apply our results to estimating Value at Risk (VaR). For theoretical background see Duffie and Pan (1997) or Jorion (1997). For alternative approaches to VaR estimation see Bertail, Haefke, Politis, and White (1999) or Danielsson and de Vries (1997). In general, we can model the density of returns as a mixture of densities that can be integrated in closed form to yield a cdf which can then be inverted to obtain VaR. Because returns can take both positive and negative values, we base our approximation on densities formed using $\boldsymbol{h}_{\lambda,\zeta}$. In this application, both integrability in closed form and adjustable tail fatness are convenient features of our approach.

Let V_t denote the value of a portfolio in period t and ρ_t the one period net return, defined as

$$\rho_t = \frac{V_t}{V_{t-1}} - 1.$$

For convenience, ρ_t is frequently assumed to be normally distributed. Many authors have noted that this assumption is implausible (e.g. Campbell, Lo and MacKinlay (1997)). To avoid this implausibility, we apply our specification from section 3.1, i.e. we take the density of ρ_t to be $\psi_\lambda(\cdot,\theta) = \psi_{\lambda,0}(\cdot,\theta)$, where λ is either given a priori or may be appended to θ and for convenience we set $\zeta = 0$. Let $VaR_{p,t}$ be the value that is "at risk" (i.e. can be lost) with a probability of p, i.e.

$$P(V_t < VaR_{p,t}) = p.$$

Although our notation does not reflect it explicitly, we understand P to be the probability conditional on information as of period $t-1$. Define $\nu_{p,t}$ such that

$$VaR_{p,t} = \nu_{p,t} V_{t-1},$$

so that $\nu_{p,t}$ can be interpreted as the gross return at which the corresponding value at risk is reached. We can then write

$$\begin{aligned} p &= P(V_t < VaR_{p,t}) \\ &= P(V_t < \nu_{p,t} V_{t-1}) \\ &= P\left(\frac{V_t}{V_{t-1}} - 1 < \nu_{p,t} - 1\right) \\ &= \Psi_\lambda(\nu_{p,t} - 1, \theta) \\ &\equiv \int_{-\infty}^{\nu_{p,t}-1} \psi_\lambda(\rho,\theta)\, d\rho. \end{aligned} \tag{20}$$

To obtain an estimate for θ we can approximate the density of returns using maximum likelihood, i.e.

$$\hat{\theta} = \mathrm{argmax}_\theta \sum_{t=1}^{T} \ln \psi_\lambda(\rho_t, \theta),$$

where we impose the appropriate restrictions on θ such that ψ_λ can be interpreted as a mixture of densities. If λ is estimated instead of fixed a priori, a final estimation setting $\hat{\lambda} = 1/(2n+3)$ for a suitable value of n can be performed. Using the estimator for θ (and λ, if estimated) we can then solve for $\nu_{p,t}$ using equation (20). The solution can be quickly and efficiently found numerically using Newton's method because we can calculate the first derivative of the objective function in closed form. Further, the solution to (20) is unique:

Proposition 30 *The function in equation (20) is continuous and strictly monotone and hence $\nu_{p,t}$ is unique.*

We illustrate this approach for the case with one hidden unit for $\zeta = 0$ and $\lambda = \frac{1}{3}$, which implies a not implausible degree of kurtosis for returns. Then

$$\begin{aligned} p &= \int_{-\infty}^{\nu_{p,t}-1} \boldsymbol{h}_{\frac{1}{3},0}(\rho)d\rho \\ &= \left[\frac{1}{2} + \frac{\rho}{2 \cdot \sqrt{3+\rho^2}}\right]_{-\infty}^{\nu_{p,t}-1}. \end{aligned}$$

Noting that $H_{1/3,0}(-\infty) = 0$ we have:

$$p = \frac{1}{2} + \frac{\nu_{p,t}-1}{2 \cdot \sqrt{3+(\nu_{p,t}-1)^2}}.$$

Rearranging, squaring both sides and solving for $\nu_{p,t}$ yields

$$\nu_{p,t} = \frac{2\sqrt{3}\left(p-\frac{1}{2}\right)}{\sqrt{1-2\left(p-\frac{1}{2}\right)} \cdot \sqrt{1+2\left(p-\frac{1}{2}\right)}} + 1.$$

For smaller values of λ (leading to larger exponents on $\nu_{p,t}$) a closed form solution might be impossible to obtain, but as previously remarked, Newton's method is readily available.

5 An Application to Option Pricing

Under standard assumptions (complete markets and the absence of arbitrage) the price of a European call option is given by

$$c(K, x, t, T) = \exp\{-\rho(T-t)\} \int_0^\infty \max(0, y-K)\, f_{T|t}(y|x)\, dy, \tag{21}$$

where K denotes the strike price, x the current price of the underlying asset, t denotes current time, T the expiration date, ρ a discount rate, y the price of the underlying asset at expiration, and $f_{T|t}(\cdot|\cdot)$ the unique risk neutral density of the underlying asset price (y) at expiration given the current price x. The first term can be interpreted as a discount factor and the integral is just the expected payoff under the risk neutral probability. For background see e.g. Lamberton and Lapeyre (1996).

The goal in this section is to derive a closed form expression for the expectation in equation (21) where we assume the risk neutral density to be one of our new mixture densities. Once we have a closed form expression, we can then readily estimate the free parameters using suitable nonlinear econometric methods.

Letting $\tau = T - t$ be the time to expiration, suppressing dependence on ρ and denoting $p(y, x, \tau) \equiv f_{T|t}(y|x)$, we can write:

$$\begin{aligned} \mathcal{C}_0(K, x, \tau) &\equiv c(K, x, t, T)\exp\{\rho\tau\} \\ &= \int_K^\infty (y-K)\, p(y, x, \tau) dy \\ &= \int_K^\infty (1 - P(y, x, \tau))\, dy, \end{aligned}$$

where $P(y, x, \tau) \equiv D^{-e_1} p(y, x, \tau)$. We replace P with a neural net approximation $\Xi_{\lambda\delta}(\cdot;\theta) = D^{-e_1}\xi_{\lambda\delta}(\cdot,\theta)$ and define the input vector $\mathbf{x}$ to be $\mathbf{x} = (y, x, \tau)$. When P is replaced with a log-normal analog we obtain

$$\mathcal{C}_{\lambda\delta}(K, x, \tau; \theta) = \int_K^\infty (1 - \Xi_{\lambda\delta}(y, x, \tau; \theta))\, dy \ .$$

Corollary 28 ensures that for suitable choices of λ and δ, $\mathcal{C}_{\lambda\delta}$ has a closed form expression. For any given λ and δ we can compute the value of the call from a closed form expression and then estimate the vector of parameters, θ, by nonlinear regression. For example, the least squares estimator is obtained by solving

$$\min_\theta \sum_{d,i,j} \left(\exp\{\rho_d \tau_{d,j}\}\, \mathcal{C}_{d,i,j} - \mathcal{C}_{\lambda\delta}(K_i, x_d, \tau_{d,j}; \theta)\right)^2,$$

where d denotes various dates of observation, i an index over different strike prices, and j a time index to capture different expiration dates. Because the approximation $\mathcal{C}_{\lambda\delta}\left(K, x, \tau; \hat{\theta}_n\right)$ is, under regularity conditions, consistent for

$\mathcal{C}_0(K, x, \tau)$ in Sobolev norm (see e.g. Gallant and White (1992)), we can approximate the derivatives of $\mathcal{C}_0(K, x, \tau)$ with those of $\mathcal{C}_{\lambda\delta}\left(K, x, \tau; \hat{\theta}_n\right)$. In particular $-\frac{\partial^2 \mathcal{C}_{\lambda\delta}\left(K, x, \tau; \hat{\theta}_n\right)}{\partial K^2}$ can provide a good approximation (asymptotically) to $-\frac{\partial^2 \mathcal{C}_0(K, x, \tau)}{\partial K^2}$ which is the desired risk neutral density (Breeden and Litzenberger 1978). Further it follows that the "greeks", delta, gamma, and theta, corresponding to the first and second derivative of C_0 with respect to x and the first derivative of C_0 with respect to τ, respectively, can also be well approximated at the same time.

6 Concluding Remarks

We propose a new family of density functions based upon the logarithm of the inverse Box - Cox transform and the flexible structure of artificial neural networks. This yields mixtures of densities capable of arbitrarily accurate approximation to large classes of functions whose antiderivatives have closed form expressions. As examples we consider applications to estimation of Value at Risk and option pricing.

Acknowledgments

The authors wish to thank John Bilson, Manfred Deistler, Carl FitzGerald, Gabriele Guehring, Christian Krattenthaler, Christof Kreuter, Mico Mrkaic, Tymon Tatur, Pascal Vogt, and Georg Wegmann for helpful discussions and valuable suggestions. For all errors we blame each other. White's participation was supported by NSF Grant SBR-9811562.

References

Artin, E. (1973): *Galoissche Theorie*. Zuerich: Harri Deutsch.

Abramowitz, M., and I. Stegun (1965): *Handbook of Mathematical Functions*. New York: Dover.

Bailey, W. (1962): *Generalized Hypergeometric Series*. New York: Strechert-Hafner Service Agency.

Bertail, P., C. Haefke, D. Politis, and H. White (1999): "A Subsampling Approach to Estimating the Distribution of Diverging Statistics With Applications to Assessing Financial Market Risks," Mimeo.

Box, G., and D. Cox (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, 26, 211–252.

Breeden, D., and R. Litzenberger (1978): "Prices of State Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–657.

Campbell, J., Lo, A., and C. MacKinlay (1997): *The Econometrics of Financial Markets*. New York: Dover.

Danielsson, J., and C. de Vries (1997): "Extreme Returns, Tail Estimation, and Value-at-Risk," Mimeo.

Duffie, D., and J. Pan (1997): "An Overview of Value at Risk," *Journal of Derivatives*, Spring, 7–49.

Erdelyi, A., W. Magnus, F. Oberhettinger, and F. Tricomi (eds.) (1953): *Higher Transcendental Functions*, vol. 1. New York, NY: McGraw-Hill.

Gallant, A., and H. White (1992): "On Learning the Derivatives of an Unknown Mapping With Multilayer Feedforward Networks," *Neural Networks*, 5, 129–138.

Gradsteyn, I.S. and I.M. Ryzhik; A. Jeffrey (editor) (1994): *Table of Integrals, Series and Products*, 5th ed. San Diego, CA: Academic Press.

Hornik, K., M. Stinchcombe, and H. White (1990): "Universal Approximation of an Unknown Mapping and Its Derivatives," *Neural Networks*, 3, 551–560.

Jorion, P. (1997): *Value at Risk*. New York: Dow Jones Irvin.

Kendall, M., and A. Stuart (1977): *The Advanced Theory of Statistics*. London, UK: Charles Griffin & Company, 4 edn.

Kuan, C., and H. White (1994): "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, 1–91.

LAMBERTON, D., AND B. LAPEYRE (1996): *Introduction to Stochastic Calculus Applied to Finance*. London, UK: Chapman & Hall.

MAGID, A., (1994): "Lectures on Differential Galois Theory" *University Lecture Series 7*, Providence, RI, American Mathematical Society.

MCDONALD, J., (1984): "Some Generalized Functions for the Size Distribution of Income" *Econometrica*, Vol. 52, No.3, 647–663.

RUDIN, W. (1964): *Principles of Mathematical Analysis*. New York: McGraw-Hill.

STUDENT (1908): "The Probable Error of a Mean," *Biometrika*, 6, 1–25.

WHITE, H. (1996): "Parametric Statistical Estimation with Artifical Neural Networks," in *Mathematical Perspectives on Neural Networks*, ed. by P. Smolensky, M. Mozer, and D. Rumelhart. Mahwah, NJ: Lawrence Erlbaum, pp. 714–776.

WHITTAKER, E., AND G. WATSON (1962): *A Course of Modern Analysis*. Cambridge: Cambridge University Press.

Appendix A: The Integral Representation of the $_2F_1$ Function

In this appendix we show how to obtain the transformation of equation (7). We begin with Euler's transformation:

$$ {}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a} dt. $$

Now do a change of variable:

$$
\begin{aligned}
x &= 1-t \\
-dx &= dt \\
{}_2F_1(a,b;c;z) &= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}(-1)\int_1^0 (1-x)^{b-1}x^{c-b-1}(1-(1-x)z)^{-a}dx \\
&= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 (1-x)^{b-1}x^{c-b-1}(1-z+xz)^{-a}dx \\
&= \frac{\Gamma(c)}{\Gamma(c-b)\Gamma(b)}(1-z)^{-a}\int_0^1 x^{c-b-1}(1-x)^{b-1}(1+x\frac{z}{z-1})^{-a}dx \\
&= (1-z)^{-a} \cdot {}_2F_1\left(a, c-b, c, \frac{z}{z-1}\right).
\end{aligned}
$$

Appendix B: The Integral Recursion Formulas for Algebraic Functions of Higher Powers

In this appendix we give the formulas that allow the step-wise reduction of integer exponents of the numerator and denominator of a rational integrand. It can be shown (e.g. Gradshteyn, Ryzhik (1994) Chapter 2) that for integer m and positive integer n the following holds:

$$ \int \frac{y^m}{\sqrt{a+by+cy^2}^{2n+1}} dy = $$

$$
\begin{cases}
\frac{y^{m-1}}{(m-2n)c\sqrt{R}^{2n-1}} - \frac{(2m-2n-1)b}{2(m-2n)c}\int \frac{y^{m-1}}{\sqrt{R}^{2n+1}}dy - \frac{(m-1)a}{(m-2n)c}\int \frac{y^{m-2}}{\sqrt{R}^{2n+1}}dy & m \neq 2n \\
\frac{-y^{2n-1}}{(2n-1)c\sqrt{R}^{2n-1}} - \frac{b}{2c}\int \frac{y^{2n-1}}{\sqrt{R}^{2n+1}}dy + \frac{1}{c}\int \frac{y^{2n-2}}{\sqrt{R}^{2n-1}}dy & m = 2n \\
\frac{2(2cy+b)}{(2n-1)(4ac-b^2)\sqrt{R}^{2n-1}} + \frac{8(n-1)c}{(2n-1)(4ac-b^2)}\int \frac{1}{\sqrt{R}^{2n-1}}dy & m = 0 \\
\frac{2(2cy+b)}{(4ac-b^2)\sqrt{R}} & n = 1,
\end{cases}
$$

where $R = a + by + cy^2$.

Appendix C: Proofs

Proof of Lemma 1.
Note that for nonnegative integer ζ

$$\frac{\lambda(1-\lambda)}{\lambda^{1+\zeta}-1} = \frac{\lambda}{\sum_{i=0}^{\zeta} \lambda^i}$$

so that

$$\lim_{\lambda\to 0} \mathcal{P}_{\lambda,\zeta}(\omega) = \lim_{\lambda\to 0} \frac{\omega^{\left(\frac{\lambda}{\sum_{i=0}^{\zeta} \lambda^i}\right)} - 1}{\lambda} \ .$$

Apply L'Hopital's rule to obtain

$$\begin{aligned}
\lim_{\lambda\to 0} \frac{\omega^{\left(\frac{\lambda}{\sum_{i=0}^{\zeta} \lambda^i}\right)}}{\lambda} &= \lim_{\lambda\to 0} \ln(\omega) \frac{\sum_{i=0}^{\zeta} \lambda^i \sum_{i=0}^{\zeta} i\lambda^{i-1}}{\left(\sum_{i=0}^{\zeta} \lambda^i\right)^2} \omega^{\left(\frac{\lambda}{\sum_{i=0}^{\zeta} \lambda^i}\right)} \\
&= \lim_{\lambda\to 0} \ln(\omega) \frac{\sum_{i=0}^{\zeta} \lambda^i (1-i)}{\left(\sum_{i=0}^{\zeta} \lambda^i\right)^2} \omega^{\left(\frac{\lambda}{\sum_{i=0}^{\zeta} \lambda^i}\right)} \\
&= \ln(\omega).
\end{aligned}$$

■

Proof of Theorem 2.
Let us consider the general case of

$$f(t) = (1+\lambda t^2)^{-b} \ .$$

We exploit the symmetry property of f and note that

$$M_m = \int_{-\infty}^{\infty} |t|^m f(t)dt = 2\int_0^{\infty} t^m f(t)dt \ .$$

Now consider the substitution $u = 1/(1+\lambda t^2)$ and put $b = \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}$. Then we obtain

$$\begin{aligned}
t &= \lambda^{-1/2}(1-u)^{1/2}u^{-1/2} \\
dt &= -\frac{1}{2}\lambda^{-1/2}(1-u)^{-1/2}u^{-3/2} \\
M_m &= \lambda^{-\frac{m+1}{2}} \int_0^1 u^{b-\frac{m+1}{2}} (1-u)^{\frac{m-1}{2}} \ ,
\end{aligned}$$

a complete beta – integral, the solution of which is given by

$$M_m = \lambda^{-\frac{m+1}{2}} \frac{\Gamma\big(b - \frac{m+1}{2}\big)\Gamma\big(\frac{m+1}{2}\big)}{\Gamma(b)}.$$

For $m = 0$ this reduces to

$$M_0 = \frac{\Gamma\big(b - \frac{1}{2}\big)}{\Gamma(b)}\sqrt{\frac{\pi}{\lambda}}, \quad b > \frac{1}{2}.$$

It remains to be shown that $b > \frac{1}{2}$. For this, the following are equivalent:

$$\begin{aligned} \frac{1-\lambda^{1+\zeta}}{2\lambda\,(1-\lambda)} &> \frac{1}{2} \\ 1-\lambda^{1+\zeta} &> \lambda(1-\lambda) \quad \lambda \neq 1 \\ 1+\lambda+\lambda^2+\ldots+\lambda^{\zeta} &> \lambda, \end{aligned}$$

which clearly holds for $0 < \lambda < 1$.

■

Proof of Theorem 3.

Use the proof of Theorem 2 to obtain:

$$M_m = \lambda^{-\frac{m+1}{2}} \frac{\Gamma\big(b - \frac{m+1}{2}\big)\Gamma\big(\frac{m+1}{2}\big)}{\Gamma(b)}.$$

Necessary and sufficient for the existence of M_m is that $b > \frac{m+1}{2}$. It is equivalent that

$$\begin{aligned} \frac{1-\lambda^{1+\zeta}}{2\lambda\,(1-\lambda)} &> \frac{m+1}{2} \\ 1-\lambda^{1+\zeta} &> \lambda(1-\lambda)(m+1) \quad \lambda \neq 1 \\ 1+\lambda+\lambda^2+\ldots+\lambda^{\zeta} &> \lambda(m+1). \end{aligned}$$

Since $\lambda^{1+\zeta} < \lambda$ follows from $0 < \lambda < 1$ it holds that:

$$\begin{aligned} \frac{1-\lambda^{1+\zeta}}{\lambda(1-\lambda)} > \frac{1}{\lambda} \quad &> \quad m+1 \\ \lambda \quad &< \quad \frac{1}{m+1}. \end{aligned}$$

Since $\lambda = \frac{1}{2n+3}$, a choice of $n > \frac{m-2}{2}$ always suffices.

■

Proof of Theorem 4.

First we establish that $\boldsymbol{h}_{\lambda,\zeta}(x)$ converges to ϕ pointwise. We have

$$\lim_{\lambda \to 0} h(\lambda, \zeta) \quad = \quad \lim_{\lambda \to 0} \frac{1}{\sqrt{2\pi}} \left(\lambda x^2 + 1\right)^{-\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \lim_{\lambda\to 0} \frac{1}{\left(\lambda x^2+1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}}} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{ \lim_{\lambda\to 0} -\frac{1-\lambda^{1+\zeta}}{2\lambda\left(1-\lambda\right)} \ln\left(\lambda x^2+1\right) \right\} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{ \lim_{\lambda\to 0} \frac{-x^2}{\left(\lambda x^2+1\right)} \frac{1}{\left(2-4\lambda\right)} \right\} \quad \text{by L'Hôpital's rule} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{x^2}{2} \right\}.
\end{aligned}
$$

Uniform convergence follows from pointwise convergence provided that $\sup_{x\in\Re}|\boldsymbol{h}_{\lambda,\zeta}(x)-\phi(x)|\to 0$ for $\lambda_n\to 0$ as $n\to\infty$ (e.g. Rudin (1964, theorem 7.9)). Since $\sup_{x\in\Re}|\boldsymbol{h}_{\lambda,\zeta}(x)-\phi(x)| = |\boldsymbol{h}_{\lambda,\zeta}(0)-\phi(0)|$ the uniform convergence follows.

■

Proof of Theorem 5.

To establish our result, we take $\lambda > 0$ so that

$$
\kappa_\lambda D^{-1} g\left(x,\lambda\right) \quad = \quad D^{-1}\left(\lambda x^2+1\right)^{-\frac{1}{2\lambda}}.
$$

Again consider the general case

$$
f(t) = (1+\lambda t^2)^{-b}.
$$

We have from Theorem 2 that

$$
\frac{\kappa_\lambda}{2} = \int_{-\infty}^{0} (1+\lambda t^2)^{-b} dt \ ,
$$

so that for $x < 0$ we can write

$$
F(x) = \frac{\kappa_\lambda}{2} - \int_0^{\infty} (1+\lambda t^2)^{-b} dt \ ,
$$

and for $x > 0$ we can write

$$
F(x) = \frac{\kappa_\lambda}{2} + \int_0^{\infty} (1+\lambda t^2)^{-b} dt \ .
$$

To obtain the integral, substitute as in Theorem 2 to obtain

$$
\int_0^x (1+\lambda t^2)^{-b} dt = -\frac{1}{2\sqrt{\lambda}} \int_1^{1/(1+\lambda x^2)} u^{b-3/2}(1-u)^{-1/2} du \ .
$$

Now substitute $v = 1-u$ to obtain

$$
\int_0^x (1+\lambda t^2)^{-b} dt = \frac{1}{2\sqrt{\lambda}} \int_0^{\frac{\lambda x^2}{1+\lambda x^2}} (1-v)^{b-3/2} v^{-1/2} dv \ .
$$

This has the form of an incomplete beta integral which can be expressed as a hypergeometric function (see Erdelyi, Magnus, Oberhettinger, and Tricomi (1953), section 2.5.3) and we obtain

$$\int_0^x (1+\lambda t^2)^{-b} dt = \frac{1}{\sqrt{\lambda}} \left(\frac{\lambda x^2}{1+\lambda x^2} \right)^{1/2} \cdot {}_2F_1 \left(\frac{1}{2}, \frac{3}{2} - b; \frac{3}{2}; \frac{\lambda x^2}{1+\lambda x^2} \right).$$

We can now write $F(x)$ as

$$\begin{aligned} F(x) &= \frac{\kappa_\lambda}{2} + \operatorname{sign}(x) \frac{1}{\sqrt{\lambda}} \left(\frac{\lambda x^2}{1+\lambda x^2} \right)^{1/2} \cdot {}_2F_1 \left(\frac{1}{2}, \frac{3}{2} - b; \frac{3}{2}; \frac{\lambda x^2}{1+\lambda x^2} \right) \\ &= \frac{\kappa_\lambda}{2} + \frac{x}{\sqrt{(1+\lambda x^2)}} \cdot {}_2F_1 \left(\frac{1}{2}, \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}; \frac{3}{2}; \frac{\lambda x^2}{\lambda x^2 + 1} \right). \end{aligned}$$

Normalizing by κ_λ now gives the desired result.

■

Proof of Corollary 6.

For nonnegative integers n such that $-n = 3/2 - 1/(2\lambda)$ the infinite sum breaks off after n terms. Solving for λ gives $\lambda_n = \frac{1}{2n+3}$ and the result follows.

■

Proof of Corollary 7.

Consider the normalizing constant κ_λ. For $\lambda = \frac{1}{2n+3}$ we can write it as

$$\begin{aligned} \kappa_\lambda &= \frac{\Gamma\left(\frac{1-\lambda}{2\lambda}\right)}{\Gamma\left(\frac{1}{2\lambda}\right)} \sqrt{\frac{\pi}{\lambda}} \\ &= \frac{\Gamma(n+1)}{\Gamma\left(n+\frac{3}{2}\right)} \sqrt{\frac{\pi}{\lambda}} \quad . \end{aligned}$$

Applying Legendre's duplication formula (Whittaker and Watson 1962, Corollary to 12.15)

$$2^{2z-1}\Gamma(z)\Gamma(z+\frac{1}{2}) = \Gamma(2z)\sqrt{\pi},$$

the above equation further simplifies to

$$\begin{aligned} \kappa_\lambda &= \frac{\Gamma(n+1)\, 2^{2n+1}}{\Gamma(2n+2)\sqrt{\pi}} \sqrt{\frac{\pi}{\lambda}} \\ &= 2^{2n+1} \frac{n!}{(2n+1)!} \sqrt{2n+3} \\ &= 2^{\frac{1-2\lambda}{\lambda}} \frac{\left(\frac{1-3\lambda}{2\lambda}\right)!}{\left(\frac{1-2\lambda}{\lambda}\right)! \sqrt{\lambda}} \quad . \end{aligned}$$

■

Proof of Theorem 8.
Multiplying with $\kappa_{\lambda,\zeta}$ and applying Euler's transformation yields:

$$\kappa_{\lambda,\zeta} D^{-1} \boldsymbol{h}_{\lambda,\zeta}(x) \quad = \quad \frac{\kappa_{\lambda,\zeta}}{2} + x \cdot {}_2F_1\left(\frac{1}{2}, \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}; \frac{3}{2}; -\lambda x^2\right).$$

Direct integration gives

$$\kappa_{\lambda,\zeta} D^{-2} \boldsymbol{h}_{\lambda,\zeta}(x) \quad = \quad \frac{x\kappa_{\lambda,\zeta}}{2} + \frac{1}{2\lambda\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - 1\right)} \cdot {}_2F_1\left(\frac{-1}{2}, \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - 1; \frac{1}{2}; -\lambda x^2\right),$$

and reapplying Euler's transformation gives

$$D^{-2} \boldsymbol{h}_{\lambda,\zeta}(x) \quad = \quad \frac{x}{2} + \frac{\sqrt{(1+\lambda x^2)}}{2\lambda\left(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - 1\right)\kappa_{\lambda,\zeta}} \cdot {}_2F_1\left(\frac{-1}{2}, \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}; \frac{1}{2}; \frac{\lambda x^2}{\lambda x^2+1}\right).$$

■

Proof of Theorem 9.
Theorem 3.1 of Gallant and White (1992) delivers the conclusion if

$$\psi_\lambda(x,\theta) = \sum_{j=1}^{q} \beta_j \boldsymbol{h}_{\lambda,\zeta}(\tilde{x}^T\gamma_j) \tag{22}$$

is ℓ-finite. Due to the finitely additive nature of (22) the result is not vacuous, if $\boldsymbol{h}_{\lambda,\zeta}$ is ℓ-finite for some ℓ. From the continuity of $\boldsymbol{h}_{\lambda,\zeta}$ and $\kappa_\lambda < \infty$ we have that $\boldsymbol{h}_{\lambda,\zeta}$ is ℓ-finite for $\ell = 0$. We proceed to verify that $\boldsymbol{h}_{\lambda,\zeta}$ is also ℓ-finite for $\ell < \frac{1}{\lambda} - 1$. Omitting the normalising factor κ_λ for clarity we have the following:

1. Continuity of $D^\ell \boldsymbol{h}_{\lambda,\zeta}(\cdot)$ follows from

$$D^\ell \boldsymbol{h}_{\lambda,\zeta}(x,\lambda) \quad = \quad (-1)^\ell \lambda^\ell x^\ell \left(\lambda x^2+1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \ell}$$

which is continuous as long as $\lambda x^2 + 1 > 0$ which always holds for $\lambda \in (0,1]$.

2. $\int_{-\infty}^{\infty} \left|D^\ell \boldsymbol{h}_{\lambda,\zeta}(x)\right| dx < \infty$ follows from

$$\begin{aligned}
\int_{-\infty}^{\infty} \left|D^\ell \boldsymbol{h}_{\lambda,\zeta}(x)\right| dx \quad &= \quad \int_{-\infty}^{\infty} \left|\lambda^\ell x^\ell \left(\lambda x^2+1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \ell}\right| dx \\
&\leq \quad \int_{-\infty}^{\infty} \left|x^\ell \left(\lambda x^2+1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \ell}\right| dx \\
&\leq \quad \int_{-\infty}^{\infty} |x|^\ell \left(\lambda x^2+1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)} - \ell} dx \\
&\leq \quad \int_{-\infty}^{\infty} |x|^\ell \left(\lambda x^2+1\right)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}} dx \\
&= \quad \int_{-\infty}^{\infty} |x|^\ell \tilde{h}_{\lambda,\zeta}(x) dx < \infty
\end{aligned}$$

by Theorem 3, provided $\lambda < 1/(1+\ell)$ or $\ell < 1/\lambda - 1$.

■

Proof of Corollary 10.
By definition

$$\int_l^u \psi_\lambda\,(x,\theta)\,dx_i \quad = \quad \sum_{j=1}^q \beta_j \int_l^u \boldsymbol{h}_{\lambda,\zeta}\left(\tilde{x}^T\gamma_j\right)dx_i.$$

Let us define

$$\begin{aligned} x &:= \tilde{x}^T\gamma \\ a_{ij}(a, x_{(i)}, \gamma_{ij}) &= a\gamma_{ij} + \sum_{k=1,k\neq i}^{r+1} \tilde{x}_k\gamma_{kj} \\ u_{ij} &:= a_{ij}(u, x_{(i)}, \gamma_{ij}) \\ l_{ij} &:= a_{ij}(l, x_{(i)}, \gamma_{ij}) \\ H_{\lambda,\zeta}(a_{ij}(a, x_{(i)}, \gamma_{ij})) &:= D^{-e_i}\boldsymbol{h}_{\lambda,\zeta}(x), \end{aligned}$$

which allows us to write

$$\begin{aligned} \int_l^u \boldsymbol{h}_{\lambda,\zeta}\left(x^T\gamma\right)dx_i &= \frac{1}{\kappa_{\lambda,\zeta}}\int_l^u \left(\lambda(\tilde{x}^T\gamma)^2+1\right)^{-\frac{1}{2\lambda}}dx_i \\ &= \frac{1}{\beta_i\sqrt{2\pi}}\int_{l_i}^{u_i}\left(\lambda x^2+1\right)^{-\frac{1}{2\lambda}}dx. \end{aligned}$$

Defining

$$\Psi_\lambda(x_{(i)}; a; \theta) = \sum_{j=1}^q \beta_j \cdot H_{\lambda,\zeta}(a_{ij}(a, x_{(i)}, \gamma_{ij}))\,,$$

we may consequently write the integral of the neural net as

$$\int_l^u \psi_\lambda(x,\theta)dx_i \quad = \quad \sum_{j=1}^q \beta_j\left[\Psi(x_{(i)}; u_{ij}; \theta) - \Psi(x_{(i)}; l_{ij}; \theta)\right].$$

■

Proof of Theorem 11.
Theorem 3.1, 3.2 and 3.3 of Gallant and White (1992) give sufficient conditions for uniform convergence of function approximators in Sobolev spaces. Single hidden layer feedforward neural networks given by (1) are sufficient for this purpose if the activation function g is *ell*-finite. This is shown in Theorem 9 for $\boldsymbol{h}_{\lambda,\zeta}$ and since the ℓ-finiteness of any non-negative function implies the $(\ell+1)$-finiteness of its antiderivative, the result follows for $H_{\lambda,\zeta}$.
■

Proof of Corollary 12.
This follows directly from Corollary 10 by substituting the functions from Theorem 8.

■

Proof of Corollary 13.
This result follows from Theorems 9 and 11 by applying the recursive ℓ-finiteness argument given in the proof of Theorem 11 one more time.

■

Proof of Theorem 14.
The substitution $x = \ln\omega$ immediately yields

$$\int_0^\infty \tilde{g}_{\lambda,\zeta}(\omega)d\omega = \int_{-\infty}^\infty \tilde{h}_{\lambda,\zeta}(x)dx = \kappa_{\lambda,\zeta}. \tag{23}$$

■

Proof of Corollary 15.
Follows from Corollary 6 and Theorem 14 via substitution.

■

Proof of Theorem 16.
For $m = 0$

$$\int_0^\infty |\omega|^0 \, \boldsymbol{g}_{\lambda,\zeta}(\omega)d\omega = \kappa_{\lambda,\zeta}$$

follows trivially from Theorem 14. For $m = 1$

$$\begin{aligned}\int_0^\infty |\omega| \, \boldsymbol{g}_{\lambda,\zeta}(\omega)d\omega &= \int_0^\infty \boldsymbol{h}_{\lambda,\zeta}(\ln\omega)d\omega &(24)\\ &= \frac{1}{\kappa_{\lambda,\zeta}}\int_0^\infty \left(\lambda\left(\ln\omega\right)^2+1\right)^{-\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}} d\omega &(25)\\ &= \frac{1}{\kappa_{\lambda,\zeta}}\int_{-\infty}^\infty \frac{e^x}{(\lambda x^2+1)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}}}dx. &(26)\end{aligned}$$

Since for any value of $\lambda > 0$ the denominator is a polynomial in x there exists $K \in Z_0^+$ such that

$$\frac{1}{\kappa_{\lambda,\zeta}}\int_{-\infty}^\infty \frac{e^x}{x^K}dx < \frac{1}{\kappa_{\lambda,\zeta}}\int_{-\infty}^\infty \frac{e^x}{(\lambda x^2+1)^{\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}}}dx. \tag{27}$$

and the left hand side of the inequality does not converge since $e^x > x^K$ for large x. Hence the right hand side integral cannot be finite either.

■

Proof of Theorem 17.
Given Corollary 15 we can integrate the infinite series associated with $G_{\lambda,\zeta}$ term by term with the argument $(\ln\omega)^{2n+1}$ (n=0,1,...). Integration by parts of the argument yields the recursive expression:

$$\int (\ln\omega)^s d\omega = \omega(\ln\omega)^s - \int (\ln\omega)^{s-1} d\omega. \tag{28}$$

Integrating each argument thus gives

$$\int (\ln \omega)^s d\omega = \omega \sum_{i=0}^{s} (-1)^{s-i} \frac{s!}{i!} (\ln \omega)^i \tag{29}$$

for *each* term. Summing up and noting that only odd powers of $\ln \omega$ occur in the integrated series yields:

$$\begin{aligned} \int \boldsymbol{g}_{\lambda,\zeta}(\omega) d\omega &= \frac{\omega}{\kappa_{\lambda,\zeta}} \frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)})} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{1}{2}+k)\Gamma(\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}+k)}{\Gamma(\frac{3}{2}+k)} \frac{(-\lambda)^k}{k!} \\ & \sum_{i=0}^{2k+1} (-1)^{2k+1-i} \frac{(2k+1)!}{i!} (\ln x)^i . \end{aligned}$$

■

Proof of Theorem 18. We have

$$\kappa_{\lambda,\delta} \equiv \int_0^\infty \tilde{h}_{\lambda,\delta}(\omega) d\omega = \int_0^\infty \omega^{\delta-1} \left(\lambda \left(\frac{\omega^\delta - 1}{\delta} \right)^2 + 1 \right)^{\frac{-1}{2\lambda}} d\omega \, .$$

Substitution of $y = \omega^\delta$ yields

$$\kappa_{\lambda,\delta} = \frac{1}{\delta} \int_0^\infty \left(\lambda \left(\frac{y-1}{\delta} \right)^2 + 1 \right)^{\frac{-1}{2\lambda}} dy \, ,$$

and further substitution of $x = \frac{y-1}{\delta}$ gives:

$$\kappa_{\lambda,\delta} = \int_{-\frac{1}{\delta}}^{\infty} \left(\lambda x^2 + 1 \right)^{\frac{-1}{2\lambda}} dx \, ,$$

which by the application of Theorem 5 can be calculated as

$$\kappa_{\lambda,\delta} = \left[\frac{1}{2} + \frac{x}{\sqrt{(1+\lambda x^2)}} \cdot {}_2F_1 \left(\frac{1}{2}, \frac{3}{2} - \frac{1}{2\lambda}; \frac{3}{2}; \frac{\lambda x^2}{\lambda x^2 + 1} \right) \right]_{-\frac{1}{\delta}}^{\infty} .$$

The result now follows by some simple algebra.

■

Proof of Theorem 19.

$$\begin{aligned} \lim_{\lambda \to 0} \lim_{\delta \to 1} \tilde{h}_{\lambda,\delta}(\omega) &= \lim_{\lambda \to 0} \lim_{\delta \to 1} \omega^{\delta-1} \left(\lambda \left(\frac{\omega^\delta - 1}{\delta} \right)^2 + 1 \right)^{-\frac{1}{2\lambda}} \\ &= \lim_{\lambda \to 0} \left(\lambda (\omega - 1)^2 + 1 \right)^{-\frac{1}{2\lambda}} , \end{aligned}$$

and through the substitution $y = \omega - 1$ the problem is reduced to the proof of Theorem 4. Similarly

$$\begin{aligned}
\lim_{\lambda \to 0} \lim_{\delta \to 0} \tilde{h}_{\lambda,\delta}(\omega) &= \lim_{\lambda \to 0} \lim_{\delta \to 0} \omega^{\delta - 1} \left(\lambda \left(\frac{\omega^{\delta} - 1}{\delta} \right)^2 + 1 \right)^{-\frac{1}{2\lambda}} \\
&= \omega^{-1} \lim_{\lambda \to 0} \left(\lambda (\ln \omega)^2 + 1 \right)^{-\frac{1}{2\lambda}} \\
&= \omega^{-1} \exp \left\{ -\frac{(\ln \omega)^2}{2} \right\} \text{ by Theorem 4} \\
&= \tilde{\phi}.
\end{aligned}$$

■

Proof of Theorem 20.
This follows from the sequential application of the recursion formulae for indefinite integrals of algebraic functions of higher powers given in Appendix B. Noting that $t < 2u$ by assumption and evaluating at the limits of integration yields the result.

■

Proof of Theorem 21 After the substitutions of the Proof of Theorem 18 and by the application of Theorem 5, the results follows.

■

Proof of Corollary 22 Noting that the coefficient sequence of the ${}_2F_1$ Theorem 5 and Theorem 21 are identical, the result follows immediately from Corollary 6.

■

Proof of Theorem 23 For any probability density f on R and its associated distribution function F it follows from integration by parts that:

$$\int_{-\infty}^{\omega} F(x)dx = \omega \cdot F(\omega) - \int_{-\infty}^{\omega} x \cdot f(x)\, dx.$$

Defining $f \equiv \boldsymbol{h}_{\lambda,\delta}$ and $F \equiv D^{-1}\boldsymbol{h}_{\lambda,\delta}$ it follows that $\Upsilon = D^{-1}(\omega \cdot \boldsymbol{h}_{\lambda,\delta})$.
Hence inserting the indefinite integral of the first moment ($m = 1$) from Theorem 20 in place of $D^{-1}(\omega \cdot \boldsymbol{h}_{\lambda,\delta})$ gives the result.

■

Proof of Corollary 24 The closed form of $D^{-2}\boldsymbol{h}_{\lambda,\delta}(\omega)$ only depends on $\Upsilon_{\lambda,\delta}$ since $D^{-1}\boldsymbol{h}_{\lambda,\delta}(\omega)$ is of closed form based upon Corollary 22. Using the general recursion formula for integrals of this type, given in Appendix B and noting that $t < 2u$ always holds under the conditions on λ and δ one gets:

$$\int \frac{y^t}{\sqrt{R}^{2u+1}} dy = \begin{cases} \frac{y^{t-1}}{(t-2u)\lambda\sqrt{R}^{2u-1}} + \frac{(2t-2u-1)}{(t-2u)} \int \frac{y^{t-1}}{\sqrt{R}^{2u+1}} dy - \frac{(t-1)(\lambda+\delta^2)}{(t-2u)\lambda} \int \frac{y^{t-2}}{\sqrt{R}^{2u+1}} dy & t < 2u \\ \frac{(y-1)\left(1+\sum_{k=1}^{u-1} \prod_{i=1}^{k} \frac{(u-i)}{(2u-2i-1)} \cdot \left(\frac{2R}{\delta^2}\right)^k\right)}{(2u-1)\delta^2\sqrt{R}^{2u-1}} & t = 0. \end{cases}$$

Since $t = \frac{1}{\delta}$ is an integer by assumption the repeated recursion of the first line yields the closed form of the second line in at most $n-2$ steps. The sum over all the resulting recursion paths, which terminate in the same closed-form final expression - albeit with different coefficients - is hence also of closed form; finiteness of $D^{-2}\boldsymbol{h}_{\lambda,\delta}$ follows from Theorems 20, 21 (for $m=1$), Corollary 22, and the fact that the integral is a monotonic linear form.

■

Proof of Theorem 25.

The result follows from the argument of Theorem 9 by using the transformation from Theorem 18.

■

Proof of Corollary 26.

The result follows from the proof of Theorem 10 by substituting $\boldsymbol{h}_{\lambda,\delta}$ for $\boldsymbol{h}_{\lambda,\zeta}$.

■

Proof of Theorem 27.

The result follows from the proof of Theorem 11 by substituting $\boldsymbol{h}_{\lambda,\delta}$ for $\boldsymbol{h}_{\lambda,\zeta}$.

■

Proof of Corollary 28.

The result follows from the proof of Corollary 12 by substituting $\boldsymbol{h}_{\lambda,\delta}$ for $\boldsymbol{h}_{\lambda,\zeta}$.

■

Proof of Theorem 29.

The result follows from the proof of Theorem 13 by substituting $\boldsymbol{h}_{\lambda,\delta}$ for $\boldsymbol{h}_{\lambda,\zeta}$.

■

Proof of Proposition 30.

Due to the additive nature of the integrand $\psi_\lambda(\rho,\theta)$ it suffices to consider the individual mixture of densities components $\boldsymbol{h}_{\lambda,\zeta}$. Since these have already been shown to be densities on R, strict monotonicity follows as long as no $x \in R$ exists for which $\boldsymbol{h}_{\lambda,\zeta}(x) = 0$. This is clearly impossible for any finite $x \in R$, therefore the result follows.

■