Title
Hypernormal Densities

Authors
Giacomini, Raffaella
Haefke, Christian
White, Halbert
et al.

Publication Date
2002-09-07

Hypernormal Densities*

Raffaella Giacomini[†] **Andreas Gottschling**[‡]
Christian Haefke[§] **Halbert White**[¶]

September 7, 2002

Abstract

We propose a new family of density functions that possess both flexibility and closed form expressions for moments and anti-derivatives, making them particularly appealing for applications. We illustrate its usefulness by applying our new family to obtain density forecasts of U.S. inflation. Our methods generate forecasts that improve on standard methods based on AR-ARCH models relying on normal or Student's t-distributional assumptions.

JEL Classification: C63; C53, C45.
Keywords: ARMA-GARCH models, Neural Networks, Nonparametric Density Estimation, Forecast Accuracy.

*The authors wish to thank John Bilson, Manfred Deistler, Carl FitzGerald, Jens Jackwerth, Eugene Kandel, Christian Krattenthaler, Andrew Patton, Michael Wolf, and participants of the 2001 European Summer Symposium in Financial Markets in Gerzensee for helpful discussions and valuable suggestions. Haefke acknowledges financial support from EU grant HPMF-CT-2001-01252. White's participation was supported by NSF grants SBR 9811562 and SES 0111238.

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

‡ E-mail: agottsch@bancs.de

§ Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain, E-mail: christian.haefke@econ.upf.es

¶ Department of Economics, University of California, San Diego, La Jolla, CA 92093 - 0508, E-mail: hwhite@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. In duration analysis unobserved variables are integrated out to avoid spurious duration dependence. 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.

This paper introduces a new family of probability density functions, the *hypernormal*, that converge to the normal 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 even greater flexibility by extending our family through mixtures, which have the structure of a single hidden layer artificial neural network. Because of their flexibility and tractability, our new family of densities may be broadly useful for econometric analysis of economic and financial data. Furthermore our proposed mixture of densities integrable in closed form may be used as a substitute for quadrature methods in numerical integration, thus exploiting the approximation capabilities of artificial neural networks.

The practice of forecasting the entire density of a variable instead of simply focusing on its conditional mean is slowly becoming common practice in the fields of macroeconomic and financial forecasting. See Tay and Wallis (2000) for a survey. This is in part a response to the need to accommodate asymmetric loss functions. In this case the commonly reported point forecast (and the associated confidence interval) becomes unsatisfactory, since they rest on the assumption of quadratic loss. The availability of density forecasts is especially relevant in the context of policy analysis, which may benefit from a complete assessment of the uncertainty associated with forecasts of the variables of interest, as opposed to merely accompanying the point forecast with a confidence interval. Recently Diebold, Gunther and Tay (1998) introduced a framework for forecast evaluations which we apply to compare U.S. inflation forecasts based on normal and our new hypernormal densities.

The outline of the paper is as follows. Section 2 provides a motivating example and discusses the integrability results, Section 3 provides a brief discussion of artificial neural networks and extends the integrability results using mixtures and artificial neural nets. This has the significant further benefit of bringing conditional densities into our framework. Section 4 presents our application to maximum-likelihood-based U.S. inflation forecasts. Section 5 concludes. Appendices contain certain mathematical details and proofs of all results.

2 A New Family of Density Functions

In a seminal paper, Johnson (1949) discusses a class of transformations such that the transformed variables can be thought of as having a normal distribution. His focus is on testing and the simplifications brought about by normality rather than the creation of systems of frequency curves. To illustrate, with Z a standard normal, Johnson suggests a system of transformations such that

$$Z = \gamma + \delta f(Y).$$

We easily see that for $\gamma = 0$, $\delta = 1$, $f = \ln$ we obtain that the distribution of Y is lognormal.

Here we consider certain transformations that enable us to obtain frequency curves with desirable properties such as

1. closed form expressions for the cdf and moments
2. extra flexibility beyond the normal.

In fact Student's t-distribution fulfills both of these requirements and is closely related to the normal. As we show next, it is in fact a one parameter extension of the standard normal.

2.1 A Simple Example: Deriving the t-distribution

Consider the Box-Cox (1964)-like transformation

$$A_\lambda(\omega) = \frac{\omega^{\frac{\lambda}{1+\lambda}} - 1}{\lambda} \quad 0 < \lambda < 1, \tag{1}$$

which converges to $\ln(\omega)$ as $\lambda \to 0$ as can be seen immediately by applying l'Hôpital's rule. The associated inverse transformation is

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

which converges to the exponential function as $\lambda \to 0$. Taking the natural logarithm yields

$$T_\lambda(x) = \ln A_\lambda^{-1}(x) = \frac{1+\lambda}{\lambda} \ln(1 + \lambda x), \tag{2}$$

an identity transformation as $\lambda \to 0$.

Next consider the standard normal density function,

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

Applying T_λ to x^2 and neglecting the normalization term $1/\sqrt{2\pi}$ we obtain

$$\begin{aligned}\ddot{t}_\lambda(x) &= \exp\left\{-\frac{1}{2}T_\lambda(x^2)\right\} \\ &= \exp\left\{-\frac{1+\lambda}{2\lambda}\ln\left(\lambda x^2+1\right)\right\} \\ &= \left(1+\lambda x^2\right)^{-\frac{1+\lambda}{2\lambda}}.\end{aligned}$$

Now reparametrize by letting $\nu = 1/\lambda$ so that

$$\tilde{t}_\nu(x) = \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

and define a normalizing constant as the integral of $\tilde{t}_\nu$ over the real line[1]:

$$\begin{aligned}\kappa_\nu &= \left(\int_{-\infty}^{\infty}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right) \\ &= \frac{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}.\end{aligned}$$

We now obtain the density function for Student's (1908) t by normalizing $\tilde{t}$:

$$\begin{aligned}t_\nu(x) &= \kappa_\nu^{-1}\tilde{t}_\nu(x) \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.\end{aligned}$$

2.2 Further Extensions of the Normal Distribution

Extending our derivation of the t-distribution, we now apply a more general Box - Cox - type transformation, which we call the "power Box-Cox" transformation, given by

$$\mathcal{P}_{\lambda,\zeta}(\omega) = \frac{\omega^{\left(\frac{\lambda(1-\lambda)}{1-\lambda^{1+\zeta}}\right)}-1}{\lambda} \quad 0<\lambda<1, \quad \zeta\geq 0. \tag{3}$$

This reduces to the standard Box - Cox transformation for $\zeta = 0$ and gives the transformation of the previous section with $\zeta = 1$. The transformation is motivated by the observation that

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

[1] See the proof of theorem 2 with $\zeta = 1$.

Hence for integer ζ we can interpret the power Box - Cox as adding powers of λ in the denominator of the exponent. We thus have

$$\begin{aligned}
\mathcal{P}_{\lambda,0}(\omega) &= \frac{\omega^{\left(\frac{\lambda}{\lambda^0}\right)}-1}{\lambda} \\
\mathcal{P}_{\lambda,1}(\omega) &= \frac{\omega^{\left(\frac{\lambda}{\lambda^0+\lambda^1}\right)}-1}{\lambda} \\
\mathcal{P}_{\lambda,2}(\omega) &= \frac{\omega^{\left(\frac{\lambda}{\lambda^0+\lambda^1+\lambda^2}\right)}-1}{\lambda}
\end{aligned}$$

and so on. Also observe that for any given λ and $\omega < 1$ the closest approximation to the natural log is given by the standard Box-Cox ($\zeta = 0$) transformation. However, for $\omega > 1$ the approximation to the natural logarithm becomes better for increasing ζ, as we see for $\lambda = 1/3$ and $\lambda = 1/7$ in figures 1 and 2.

Lemma 1 *Let $\mathcal{P}_{\lambda,\zeta}(\omega)$ be as defined in (3). Then for all $\zeta \geq 0$*

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

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 desired extension of (2):

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

Although the normal distribution plays a central role in economics, econometrics, and finance, its cdf does not have a closed form expression. In fact, 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

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

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. As we saw in the preceeding section, for $\zeta = 1$ we obtain the well-known t distribution.

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 $\zeta \geq 0$ 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. Moreover, $\boldsymbol{h}_{\lambda,\zeta}$ 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 all $m > 0$:*

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

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

Furthermore, for these values of λ, closed form expressions for the integer 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 $\zeta \geq 0$ $\boldsymbol{h}_{\lambda,\zeta}$ converges to ϕ uniformly as $\lambda \to 0$. Accordingly we define $h_{0,\zeta} \equiv \phi$.*

Figure 3 presents a plot of the densities for $\zeta = 0$ and a few values of λ compared to the normal density and the t_2 $(= h_{1/2,1})$.

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 Slater (1966). More recently Abadir (1999) has carefully summarized several results about hypergeometric functions relevant for econometricians and economists.

Theorem 5 *Let $\boldsymbol{h}_{\lambda,\zeta}$ be as in (5). Then for all $x \in R$, $\zeta \geq 0$, 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 can reduce the hypergeometric series to a finite polynomial by choosing the appropriate λ. For nonnegative integers n such that

$$-n = \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}$$

the infinite sum breaks off after n terms and therefore the expression in (7) has closed form. Solving for λ taking the value of ζ as given yields the values reported in table 1. For general ζ the choice of λ_n is given by the solution of an algebraic equation of order ζ in λ, which does not necessarily possess

Table 1: Expressions for λ conditional on ζ that yield closed form antiderivatives.

$\zeta = 0$	$\lambda_n = \frac{1}{2n+3}$	$n = 0, 1, 2, \ldots$
$\zeta = 1$	$\lambda_n = \frac{1}{2n+2}$	$n = 0, 1, 2, \ldots$
$\zeta = 2$	$\lambda_n = 1 + n - \sqrt{2n + n^2}$	$n = 0, 1, 2, \ldots$

solutions that allow a convenient expression for λ as a function of its coefficients. Nevertheless, as can be seen from figure 4 the resulting values for λ are quite similar for higher n. We focus particular attention on the $\zeta = 0$ case for the sake of convenience and simplicity[2].

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 6 *For $\zeta = 0$ and appropriate λ as in table 1, 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. Some simple solutions with $\zeta = 0$ are given in table 2.

Table 2: Cdfs for simple choices of λ.

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

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 $\{\boldsymbol{h}_{\lambda,\zeta}, 0 \leq \lambda < 1,\ \zeta \geq 0\}$ the *"standard hypernormal"* family. We say that $\boldsymbol{h}_{\lambda,\zeta}$ is standard 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

[2] Recall that the t-distribution corresponds to $\zeta = 1$; our choice $\zeta = 0$ has fatter tails with the same number of existing moments for given λ.

density $\boldsymbol{h}_{\lambda,\zeta}$, then this conditional expectation has the form

$$\frac{\int_{-\infty}^{a} x\boldsymbol{h}_{\lambda,\zeta}(x)dx}{\int_{-\infty}^{a} \boldsymbol{h}_{\lambda,\zeta}(x)dx},$$

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

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

These functions also have a closed form expression for all λ of the form provided in table 1.

3 Mixture Distributions and Artificial Neural Networks

Further flexibility can be achieved by considering mixtures of hypernormals, that is, by taking a convex combination of densities of scaled and shifted standard hypernormals. Just as occurs with mixtures of normals, hypernormal mixtures can deliver skewed distributions, distributions with tail properties unachievable by a single hypernormal or distributions with two or more modes. In fact, under suitable conditions, such mixtures can approximate any distribution in large classes of probability distribution functions. A potential advantage of using the hypernormal instead of the normal to form a mixture is that, because of the inherently greater flexibility of the hypernormal, one may require fewer terms (mixing densities) in the convex combination to achieve a given accuracy of approximation to whatever may be the true density of interest.

We establish our result for mixtures of hypernormals by exploiting available results for artificial neural networks (ANNs). As we shall see next, this not only delivers results directly, but also permits us to accommodate the approximation of conditional distributions. Over the last two decades ANNs have emerged as a prominent class of flexible functional forms for function approximation. A leading case is the single hidden layer feedforward neural network, written as:

$$\psi\left(\mathbf{x}, \beta, \gamma\right) = \sum_{j=1}^{q} \beta_j \cdot g\left(\tilde{\mathbf{x}}^T \gamma_j\right), \tag{8}$$

$\tilde{\mathbf{x}} = (1, x_1, x_2, \ldots, x_r)$, $\gamma = \left(\gamma_1^T, \gamma_2^T, \ldots, \gamma_q^T\right)^T$, $\gamma_j \in R^{r+1}$ 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 (8) 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 delivers the mixture density with weights β_j. Such mixtures can approximate arbitrary densities (e.g., White 1996, theorem 19.1). The form of (8) delivers not only flexibility, but it also provides the foundation for analytic tractability: the properties of the integral of ψ 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 $\psi(\mathbf{x}, \beta, \gamma)$ can be viewed as a mixture of univariate pdf's in the usual way. On the other hand, if $r > 1$ we can view $\psi(\mathbf{x}, \beta, \gamma)$ 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 ψ 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 flexible closed form expressions for the integral of ψ. We do this by putting $g = h_{\lambda,\zeta}$. Our next result shows that these mixtures can deliver arbitrarily accurate approximations to a large class of densities under suitable conditions.

Theorem 8 *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*

$$||f||_{m,\infty,\chi} \equiv max_{n \leq m} sup_{x \in \chi} |D^n f(x)| < \infty \tag{9}$$

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 $S_{\infty}^{m}(\chi)$ equipped with metric (9) using a single hidden layer feedforward network of the form

$$\psi_{\lambda,\zeta}(x, \theta) = \sum_{j=1}^{q} \beta_j \cdot \boldsymbol{h}_{\lambda,\zeta}\left(\tilde{x}^T \gamma_j\right), \tag{10}$$

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

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

Corollary 9 *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 (10) 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 9 have desirable approximation properties:

Theorem 10 *Let f and $\boldsymbol{h}_{\lambda,\zeta}$ be as in Theorem 8, and let $H_{\lambda,\zeta}$ be as in Corollary 9. 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 (9) using a single hidden layer feedforward network of the form $\Psi_{\lambda,\zeta}(\cdot)$ given in Corollary 9.*

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 9 and Theorem 10 for the integral of $\Psi_{\lambda,\zeta}$

Corollary 11 *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 $l \le u$ 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 12 *Let* f *and* $\boldsymbol{h}_{\lambda,\zeta}$ *be as in Theorem 5, and let* $\Xi_{i,\lambda,\zeta}$ *be as in Corollary 11. 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 (9) using a single hidden layer feedforward network of the form* $\Lambda_{i,\lambda,\zeta}$ *given in Corollary 11.*

When f is the antiderivative of a cdf, $\Lambda_{i,\lambda,\zeta}$ can approximate it. 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. This property is useful for example in option pricing contexts, as risk neutral densities can be well approximated by fitting networks involving our Ξ's to the option price - moneyness relation and then differentiating twice.

4 Application

The literature in the field of density forecasting and evaluation is still very young, and far from agreeing on the methods and techniques to be used to first produce the density forecasts (analytical or numerical) and then to evaluate their performance. We apply a forecast evaluation technique introduced by Diebold et al. (1998) to one-period-ahead density forecasts of inflation, generated using a range of different model specifications.

In the attempt to more flexibly model the variables of interest, we make use of our new family of probability density functions introduced in section 2.

We begin with a standard approach by modeling the unconditional distribution of inflation under a normal distribution assumption. Then we proceed to considering more flexible models for the conditional density of inflation. Most models used in the literature (AR-ARCH-type models) assume a conditionally normal distribution for the error process (or a Student's t-distribution, as suggested by Bollerslev (1987)). Since the aim here is to achieve greater flexibility, we make use of models of the AR-ARCH class for the conditional mean and variance, and consider the situation where the models' disturbances are conditionally hypernormal. The estimated models are used to produce sequences of density forecasts, that are evaluated using the approach proposed by Diebold et al. (1998). The performance of density forecasts with normal, Student's t and hypernormal distributional assumptions is then compared and contrasted, in order to investigate whether the hypernormal distribution can provide any improvement over the more common normal and Student's t-distributions.

We work with scaled and shifted versions of the standard hypernormal. Letting Z be the standard hypernormal, $\mu \in R$, and $\sigma > 0$, define $X = \mu + \sigma Z$. It is straightforward to show that the density of X is given by

$$g_{\lambda,\zeta;\mu,\sigma}(x) \equiv \kappa_{\lambda,\zeta,\sigma}^{-1} \left(\lambda \left(\frac{x-\mu}{\sigma} \right)^2 + 1 \right)^{-\frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}} \qquad \text{for } x \in \mathbf{R}, \tag{11}$$

where

$$\kappa_{\lambda,\zeta,\sigma} \equiv \sigma \kappa_{\lambda,\zeta} < \infty.$$

Further, let $Y = a + bX$, with $a \in R$, $b \neq 0$. Then Y has density $g_{\lambda,\zeta;a+b\mu,|b|\sigma}$, and we write $Y \sim HN(a + b\mu, |b|\sigma; \lambda, \zeta)$.

When $Y \sim HN(\mu, \sigma; \lambda, \zeta)$, the parameters control the location, spread and tail behavior of the distribution. Whereas μ uniquely controls the center, all parameters interact in regulating spread, tail behavior, and higher moments of the distribution. In some special cases, however, it is easier to interpret the meaning of the individual parameters. For example, for $\mu = 0$, $\sigma = 1$, $\zeta = 1$, the density $g_{\lambda,\zeta,\mu,\sigma}$ is the familiar Student's t-density with $1/\lambda$ degrees of freedom. For this reason, in the following we will simply denote the hypernormal distribution corresponding to the parameterization $HN(0, 1; \lambda, 1)$ by $t(\nu)$, where $\nu = 1/\lambda$.

The forecast evaluation method used in the paper utilizes the c.d.f. of the variable of interest. The problem with some of the common distributions used in econometric modeling is the lack of closed form expressions for the c.d.f., as in the case of the normal. The c.d.f. for the hypernormal, in contrast, is computed easily.

4.1 Generating density forecasts

The focus of our application is to use the hypernormal distribution to generate one-period-ahead density forecasts of inflation. Density forecasts can be obtained by a range of techniques, including parametric and non-parametric methods. The density forecasts analyzed here are obtained from conditional parametric models. We consider several model specifications for inflation, assuming either conditionally normal, Student's t or hypernormal disturbances.

In the following, Y_t will denote the variable inflation at time t and I_{t-1} the information set available at time $t-1$. The models considered are the following.

Model 1 is a naive model that assumes the conditional distribution of inflation to be normal with mean μ and variance σ^2.

Model 1 : $Y_t|I_{t-1} \sim N(\mu, \sigma^2)$.

The inclusion of this simple model as a density forecast is mainly for illustrative purposes. As this model is most likely misspecified, we will be able to see whether and how the evaluation technique used in the paper reveals this misspecification.

The next set of models allows for conditional dynamics in the mean and the variance of the process. In line with Engle (1983), we focus on univariate models of the class AR(p)-ARCH(q).

Models 2, 3 and 4 are AR(p)-ARCH(q) with, respectively, normal, Student's t- and hypernormal disturbances:

Model 2 : $Y_t = \phi_0 + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t$,

$$u_t = \sqrt{h_t} v_t \tag{12}$$

$$h_t = k + \alpha_1 u_{t-1}^2 + ... + \alpha_q u_{t-q}^2 \tag{13}$$

and

$$v_t \mid I_{t-1} \sim N(0,1). \tag{14}$$

Model 3 : Same as Model 2 except

$$v_t \mid I_{t-1} \sim t(\nu). \tag{15}$$

Model 4 : Same as Model 2 except

$$v_t \mid I_{t-1} \sim HN(0,1;\lambda,0). \tag{16}$$

4.2 Estimation and Evaluation

A sample of forecast densities of inflation from the four models described in the previous section is generated using a recursive sampling scheme, as, e.g., in Clements and Smith (2000), to allow for the possibility of time-varying densities. The available sample is divided into two parts, 1959:1-1981:3 and 1981:4-1997:10 with the first part used for estimation and the second part for out-of-sample evaluation. Each model's parameters are initially estimated using the data in the first sample, and the model is used to generate a one-step-ahead density forecast. The sample is then increased by adding the following observation, the model's parameters are re-estimated and the second density forecast is produced. Continuing in this fashion until all observations from the second part of the sample are utilized results in a sequence of 200 density forecasts for each model of inflation. Notice that we do not re-specify the model at each iteration, but assume instead that the specification selected for the first estimation remains constant over time. For models of the AR(p)-ARCH(q) class the Schwartz BIC information criterion is used to choose lag lengths for the conditional mean and conditional variance equations and the parameters are estimated by maximum likelihood. For models assuming normal residuals, the density forecast of Y_t is normal with parameters depending on the chosen specification for the conditional mean and the conditional variance. The density forecast of Y_t for models that assume Student's t- and hypernormal disturbances will have parameters ν and λ that vary with time.

In line with Clements and Smith (2000) and Diebold et al. (1998), we ignore parameter estimation uncertainty and do not perform diagnostic tests on the estimated models when constructing our tests. This assumption is not uncommon in the forecast evaluation literature, where the forecasts are considered to be the primitives. In essence, we are sequentially conditioning on the information generating the parameters (and forecasts).

To evaluate the sample of density forecasts, we choose the method proposed by Diebold et al. (1998), which is based on the idea that a density forecast can be considered optimal if the model for the density is correctly specified. This approach allows one to evaluate forecasts without the need to specify a loss function, and in this sense it represents an improvement over most of the standard techniques for evaluating point forecasts, which typically assume a quadratic loss function.

The method consists of considering the sequence of probability integral transforms of inflation with respect to the density forecasts, that is

$$z_t = \int_{-\infty}^{y_t} p_t(u)du, \quad t = 1, ..., T \tag{17}$$

where y_t is the realization of inflation at time t and $p_t(y_t)$ the estimated density forecast.

Diebold et al. (1998) prove the following Proposition (adapted to fit our notation):

Proposition 13 *Let $\{y_t\}_{t=1}^T$ be the sequence of realizations of a process with true conditional densities $\{f_t(y_t|I_{t-1})\}_{t=1}^T$. If a sequence of density forecasts $\{p_t(y_t)\}_{t=1}^T$ coincides with $\{f_t(y_t|I_{t-1})\}_{t=1}^T$, then under the usual condition of a nonzero Jacobian with continuous partial derivatives, the sequence of probability integral transforms of $\{y_t\}_{t=1}^T$ with respect to $\{p_t(y_t)\}_{t=1}^T$ is i.i.d. U(0,1). That is,*

$$\{z_t\}_{t=1}^T \sim i.i.d.\, U(0,1).$$

In other words, if the sequence of density forecasts is correctly specified, the corresponding sequence of z_t's is i.i.d. $U(0,1)$. This result suggests evaluating the density forecasts $\{p_t(y_t)\}_{t=1}^T$ by testing the hypothesis of i.i.d. $U(0,1)$ on the sequence $\{z_t\}_{t=1}^T$.

As Diebold et al. (1998) point out, the fact that the i.i.d. $U(0,1)$ hypothesis on the z_t's is a joint hypothesis represents a drawback, since available tests such as the Kolmogorov-Smirnov cannot tell if rejection is due to violation of the uniform distributional assumption or of the i.i.d. assumption. We therefore follow the approach of the authors and consider a number of tests, supporting the Kolmogorov-Smirnov test with more informal, graphical tests of the i.i.d. $U(0,1)$ hypothesis.

Here, the Kolmogorov-Smirnov test measures the distance (maximum absolute difference) between the empirical c.d.f. of the sequence of z_t's and the theoretical c.d.f. of a uniform random variable. A non-trivial problem with the test is that it rests on the assumption of independence in the data and little is known about the impact of dependence on the behavior of the test statistic. Because of this, the outcome of the test could be unreliable if independence is rejected.

We test the hypothesis of i.i.d. behavior of the sequence $\{z_t\}_{t=1}^T$ using Breusch-Godfrey LM tests for serial correlation. Since there may be dependence in higher moments, the tests are performed on the series $(z_t - \overline{z})^i, i = 1, ..., 4$, to detect misspecifications in the conditional mean, variance, skewness and kurtosis.

To assess uniformity, Diebold et al. (1998) suggest considering an estimate of the p.d.f. for the z_t's, like the histogram plot of the z's, and evaluate its distance from the theoretical p.d.f. of a $U(0,1)$. We also consider the empirical c.d.f., and compare it to the c.d.f. of a $U(0,1)$, as in Diebold, Tay and Wallis (1999). The estimates of the p.d.f. and of the c.d.f. are accompanied by 95% confidence intervals. For the histogram plot, the derivation of confidence intervals is made possible by the fact that under the hypothesis of i.i.d. $U(0,1)$ the number of observations that fall into a given bin (against all other bins combined) is distributed as a Binomial$(T, \frac{1}{N})$, where T is the sample size and N the number of bins. For the c.d.f. the bounds are the critical values for the Kolmogorov-Smirnov test for the given sample size.

4.3 Data

As in Stock and Watson (1999) we use monthly U.S. Consumer Price Index (CPI) data from 1959:1 to 1997:10 from the DRI Basic Database (formerly known as CITIBASE). Inflation is calculated as the log-difference of CPI over the sample period, multiplied by a factor of 100. Figure 5 shows a time-series plot of inflation and figure 6 the histogram and summary statistics.

From the histogram we can see that the empirical distribution is skewed to the right, due to the rare occurrence of negative inflation. The Jarque-Bera test of normality leads to rejection of the null of normality. In order to test for non-stationarity in the data, we conduct an Augmented Dickey-Fuller Unit Root test. The ADF test statistic with 6 lags of the change in the dependent variable leads to rejection of the unit root hypothesis at the 5% level.

4.4 Results

In this section we evaluate density forecasts based on Models 1-4 described in section 4.1. We plot $\hat{\nu}$ and $\hat{\lambda}$ for the out-of-sample period in figure 7. For ease of

Table 3: Kolmogorov-Smirnov test of $H_0 : \{z_t\} \sim i.i.d.\, U(0,1)$

Model	Test Statistic	Critical value
M1: $N(\mu, \sigma^2)$	0.3067*	0.1027
M2: $AR(12) - ARCH(1) - nor$	0.1068*	0.1027
M3: $AR(12) - ARCH(1) - t$	0.1103*	0.1027
M4: $AR(12) - ARCH(1) - hyp$	0.0754	0.1027

Values of the test statistic and the critical value for the Kolmogorov-Smirnov test of the hypothesis of i.i.d. $U(0,1)$ of the z's from each model. A '*' indicates rejection of the null hypothesis at a 5% confidence level.

Table 4: p-values of LM test of no serial correlation in $(z_t - \overline{z})^i, i = 1, ..., 4$

	Series			
Model	$(z_t - \overline{z})$	$(z_t - \overline{z})^2$	$(z_t - \overline{z})^3$	$(z_t - \overline{z})^4$
M1: $N(\mu, \sigma^2)$	0.00	0.00	0.00	0.00
M2: $AR(12) - ARCH(1) - nor$	0.00	0.00	0.00	0.00
M3: $AR(12) - ARCH(1) - t$	0.01	0.04	0.16	0.29
M4: $AR(12) - ARCH(1) - hyp$	0.16	0.10	0.14	0.18

p-values for the Godfrey-Breusch LM test of no serial correlation in the first four powers of $(z_t - \bar{z})$. The null hypothesis is that there is no serial correlation in the relative series up to lag $p = 10$. The test statistic is computed as number of observations times the (uncentered) R^2 from a regression of the series on p of its lags. The LM test statistic is asymptotically distributed as a $\chi^2(p)$.

presentation, the estimation results for the individual forecasts are not discussed further. Subsequently, we focus on the behavior of the sequences of probability integral transforms obtained from the different models.

For each model of inflation, we test the null hypothesis of i.i.d. $U(0,1)$ for the sequence of probability integral transforms $\{z_t\}_{t=1}^{200}$ of the realizations of inflation with respect to the density forecasts generated by the model. The first test considered is a Kolmogorov-Smirnov test of the joint hypothesis of i.i.d. $U(0,1)$, followed by a test for i.i.d. behavior, in the form of the Breusch-Godfrey LM test for serial correlation up to 10 lags in the series $(z_t - \overline{z})^i$, $i = 1, ..., 4$. The correlograms of the various powers are also presented for visual inspection, as an addition to the formal test of i.i.d. Finally, to determine whether the z_t's are uniform, we plot their empirical p.d.f. and empirical c.d.f. against the theoretical p.d.f. and c.d.f. of a $U(0,1)$. The results for the Kolmogorov-Smirnov test and for the Breusch-Godfrey LM tests for all models are respectively reported in table 3 and table 4. Figures 8–11 plot the sample autocorrelograms of the z_t's. The empirical p.d.f. and c.d.f. are shown in figures 12 and 13.

4.4.1 Model 1

For the sequence of z_t's derived from the naive unconditionally normal model, the Kolmogorov-Smirnov test rejects the hypothesis of i.i.d. $U(0, 1)$ for the z_t's, with a value of the test statistic of 0.3067, which is greater than the 5% critical value of 0.1027.

The sequence of z_t's is affected by serial correlation in all powers of $(z_t - \overline{z})$, as revealed by the zero p-values for the LM test of no serial correlation in table 4. The same conclusion clearly emerges from the analysis of the sample autocorrelogram of all four powers of $(z_t - \overline{z})$, in figures 8–11. The presence of serial correlation in all powers of the z series suggests that the specified model fails to capture dynamics in the mean and the variance of inflation, as well as failing to correctly model higher moments. Uniformity is also rejected by the graphical analysis of the p.d.f. and c.d.f. of the z_t's, in figures 8 and 9. One can observe that many of the bin heights fall outside the confidence interval, with particularly high peaks between the 0.2-0.4 bins, compensated by too low bin heights in the last four bins. The empirical c.d.f. also reveals a strong deviation from uniformity, as the curve is outside the confidence interval in the second half of the domain. This means that too many observations fall into the first half of the forecast distributions and too few in the second half, relative to what we would observe if the data were really i.i.d. normal.

Overall, one can see how the evaluation technique allows the identification of some problems associated with the specified density forecast, also suggesting which features of the model should be modified.

4.4.2 Model 2

In all remaining three models, the conditional mean and variance are modeled as an $AR(12) - ARCH(1)$. Model 2 assumes the disturbances to be normal. We will thus denote Model 2 as $AR(12) - ARCH(1) - nor$. Using the recursively estimated density forecasts from Model 2, we generate a sequence of z_t's, for which we perform the battery of tests described above.

Tables 3 and 4 reveal that the Kolmogorov-Smirnov test leads to rejection of the hypothesis of i.i.d. $U(0, 1)$ for the series of z_t's and that the LM tests find serial correlation in all powers of $(z_t - \overline{z})$. The same conclusion is reached by examining the autocorrelograms of all four powers of $(z_t - \overline{z})$, in figures 8–11. For all powers, the correlograms are significantly different from the correlogram of an i.i.d. series. The histogram of the z's (figure 12) is closer to uniform than the one for Model 1, but some bins still fall outside the 95% confidence interval. Similarly, the empirical c.d.f. hits the boundaries of the confidence interval, seemingly falling outside (although only marginally) at some point in the second part of the interval, as can be observed in figure 13.

Overall, Model 2 fails on all counts, poorly capturing the dynamics of inflation and assuming a functional form which appears to be misspecified.

4.4.3 Model 3

The failure of the normal distribution to adequately describe economic variables is a well-established result, a fact that makes the failure of Model 2 unsurprising. In many cases, and particularly for financial data, the Student's t has been considered a viable alternative, given its ability to capture more complex tail behavior than the normal. We thus proceed to evaluate the forecasts produced by an $AR(12) - ARCH(1)$ model with Student's t-disturbances, which we indicate as $AR(12) - ARCH(1) - t$.

Table 3 shows that the Kolmogorov-Smirnov test rejects the null of i.i.d. $U(0,1)$ of the z's derived from the model, at the 5% confidence level. The null hypothesis of serial independence in the series $(z_t - \overline{z})$ and $(z_t - \overline{z})^2$ is also rejected at the 5% confidence level by the LM test, whose p-values are reported in table 4. However, serial correlation is not found in the series $(z_t - \overline{z})^3$ and $(z_t - \overline{z})^4$, as revealed by p-values for the LM test of magnitude, respectively, 0.16 and 0.29. The autocorrelograms of all four powers of $(z_t - \overline{z})$, in figures 8–11 also seem to indicate that the probability integral transforms of density forecasts from Model 3 are 'closer' to white noise, as suggested by sample autocorrelations which are at best marginally significant. However, the analysis of the histogram and of the empirical c.d.f. (figures 12 and 13) seems to suggest that the assumption of Student's t-disturbances does not induce significant improvements relative to the assumption of normality. The appearance of the p.d.f. and c.d.f. plots for Model 3 is very similar to those for Model 2: some of the bins in the histogram are outside the 95% confidence interval, and the empirical c.d.f. hits the boundaries of the confidence interval around the 45 degree line, marginally falling outside at some points.

Overall, Model 3 seems to improve on the performance of Model 2 only marginally, by eliminating or reducing the serial correlation in the powers of $(z_t - \overline{z})$. The Student's t-distributional assumption, however, appears overall to be inadequate.

4.4.4 Model 4

The final set of density forecasts is derived by assuming inflation to be described by an $AR(12) - ARCH(1)$ model with hypernormal disturbances having $\zeta = 0$ (denoted as $AR(12) - ARCH(1) - hyp$).

The Kolmogorov-Smirnov test fails to reject the null hypothesis of i.i.d. $U(0,1)$ for the sequence of probability integral transforms derived from Model 4, at the 5% confidence level. Further, the LM test in table 4 fails to reject the null hypothesis of no serial correlation for all series $(z_t - \overline{z})^i, i = 1, ..., 4$ at typical confidence levels. The autocorrelograms in figures 8–11 also seem to confirm the conclusion that the hypothesis of i.i.d. behavior of the series $(z_t - \overline{z})$ can be reasonably accepted. The $U(0,1)$ behavior for the z's is also confirmed by an analysis of the histogram plot in figure 12, which displays all bins falling within the 95% confidence bounds. The empirical c.d.f. in figure 13 is also clearly within the confidence interval, making the probability integral transforms of the density forecasts generated by Model 4 pass all tests of the i.i.d. $U(0,1)$ hypothesis.

In conclusion, density forecasts obtained from an $AR(12)-ARCH(1)$ model with hypernormal disturbances having $\zeta = 0$ appear to provide the best approximation for the true density of inflation over the sample considered in the paper. The use of the hypernormal distributional assumption also constitutes an improvement over the more standard assumptions of normality (for which forecasts fail all evaluation tests) or Student's t-distribution (whose forecasting performance is only marginally superior to that assuming normality).

5 Conclusion

We propose a new family of density functions based upon the logarithm of the inverse power Box - Cox transform and the flexibility of mixture distributions. This yields densities capable of arbitrarily accurate approximation to large classes of functions whose antiderivatives have closed form expressions.

The closed form integrability property can also be used to substitute for numerical integration techniques like quadrature methods. Especially in higher dimensions our method may hold an advantage over the polynomials used with quadrature methods. The drawback of this approach, however, is that there is an underlying nonlinear numerical optimization problem that has to be solved to obtain the coefficients of the neural net approximation to the function that is to be integrated.

To illustrate the usefulness of the new class of densities we obtained parametric density forecasts of inflation using a range of different model specifications. The common cases of normal and Student's t-distributions were compared to the hypernormal distribution with $\zeta = 0$. We used a recursive sampling scheme to produce sequences of density forecasts from our models, and evaluated the forecasts using the techniques proposed by Diebold et al. (1998).

Considering models of the AR(p)-ARCH(q) class, we found that the $\zeta = 0$ hypernormal assumption for inflation greatly improved density forecast performance, giving a sequence of z's that passed all of our diagnostic tests, whereas the equivalent exercise on models assuming normality and Student's t-distributions failed on all or most counts. This suggests that the hypernormal could become a useful tool in econometric modeling, given its convenience, generality, and flexibility.

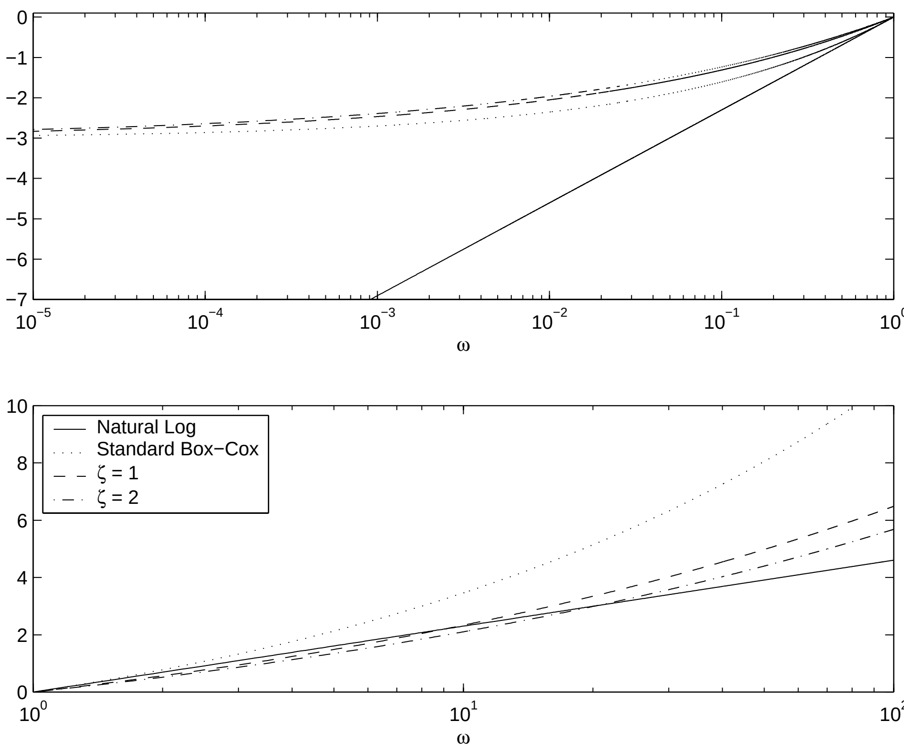

Figure 1: The Power Box-Cox transformation for $\lambda = 1/3$ and increasing values of ζ.

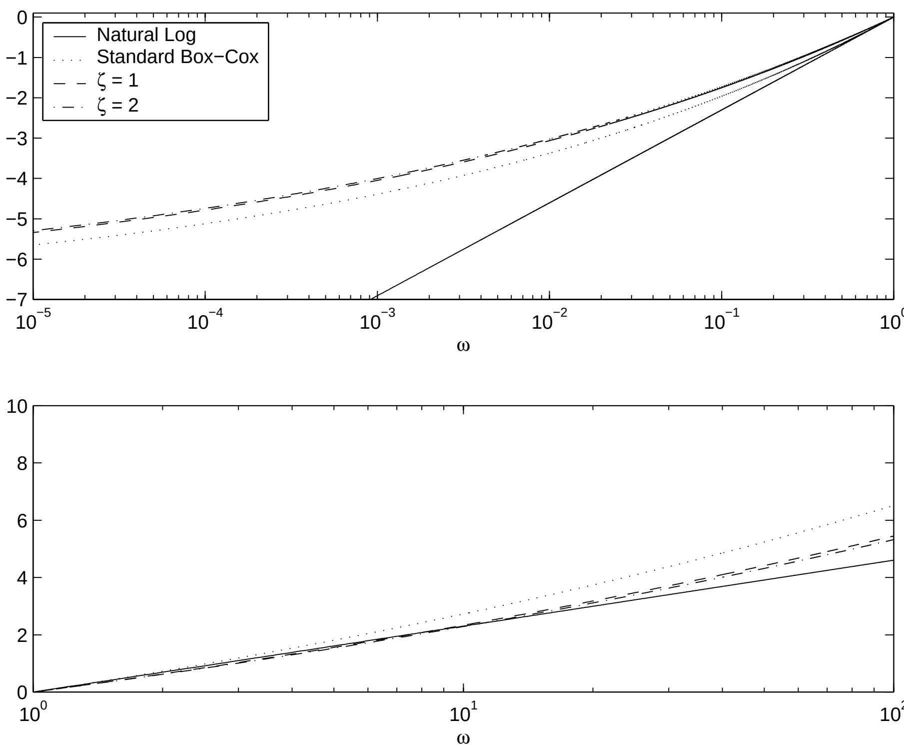

Figure 2: The Power Box-Cox transformation for $\lambda = 1/7$ and increasing values of ζ.

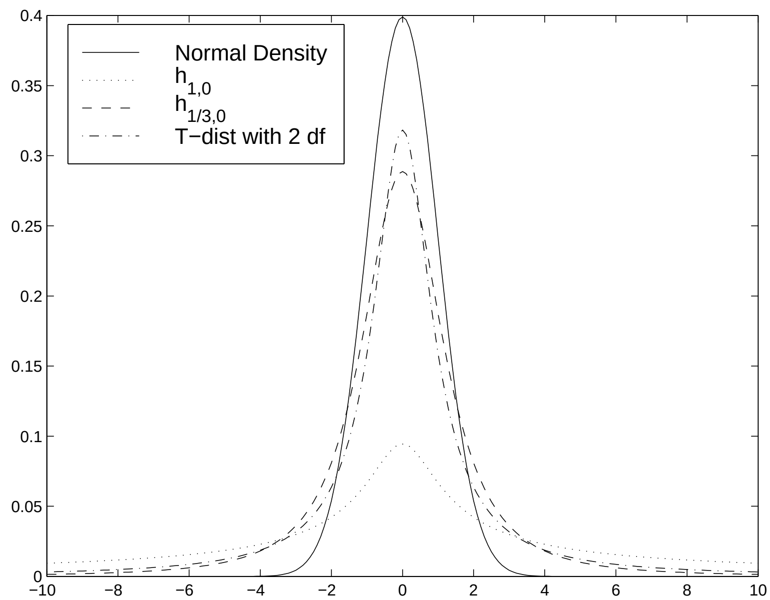

Figure 3: $\boldsymbol{h}_{\lambda,0}$ and the normal density

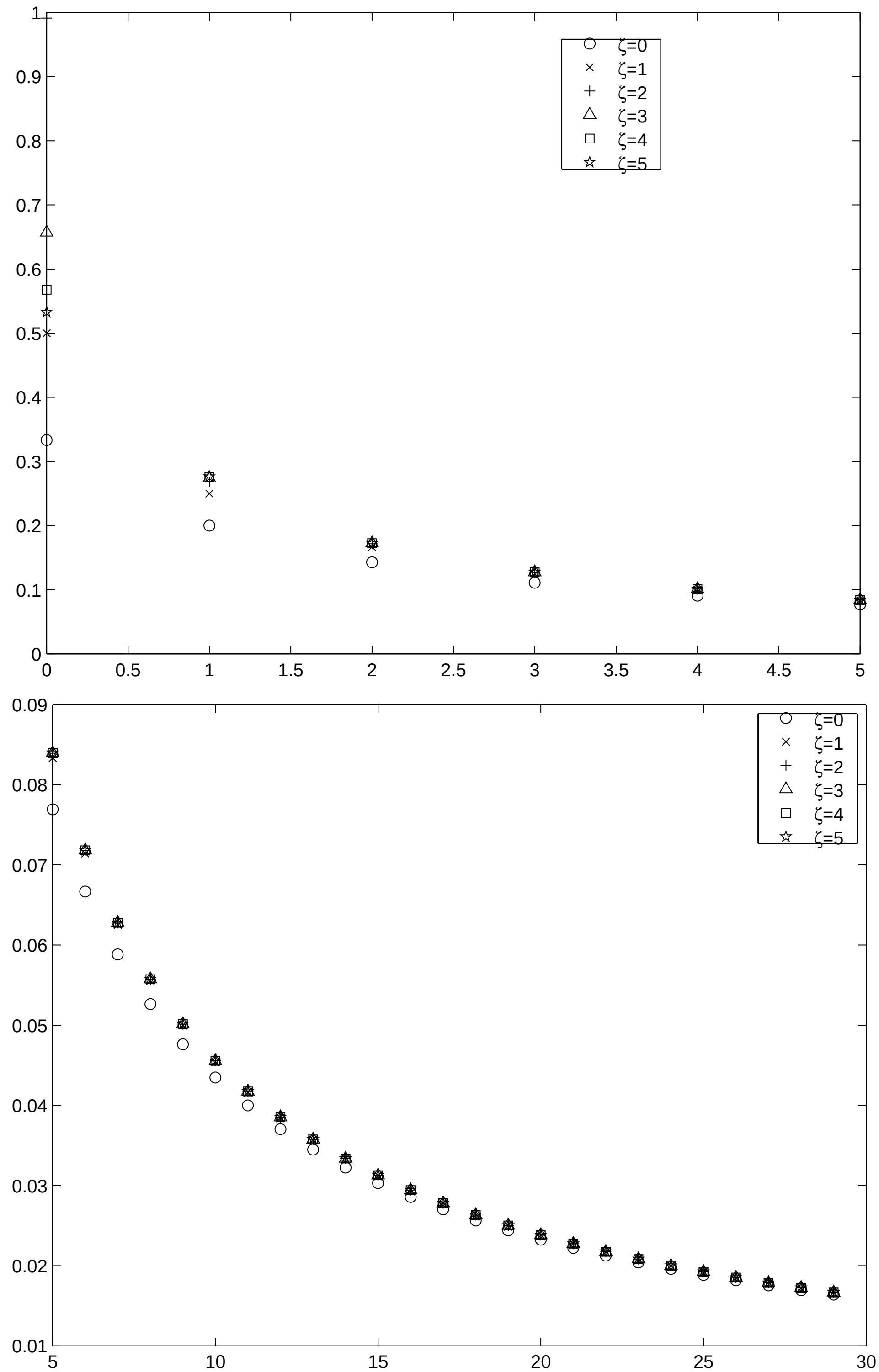

Figure 4: Values for $\lambda = 1/3$ as functions of n which yield closed form expressions for moments and cdf at different values of ζ.

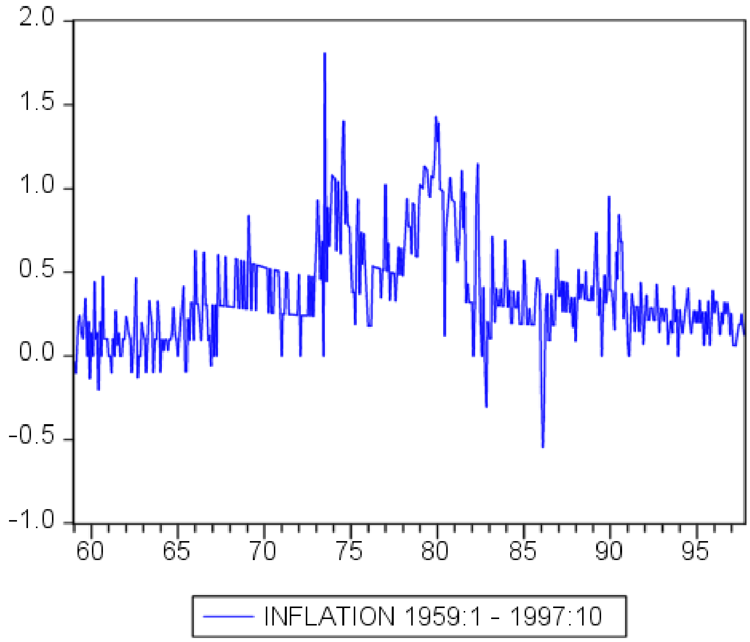

Figure 5: U.S. Inflation, 1959/1 - 1997/10.

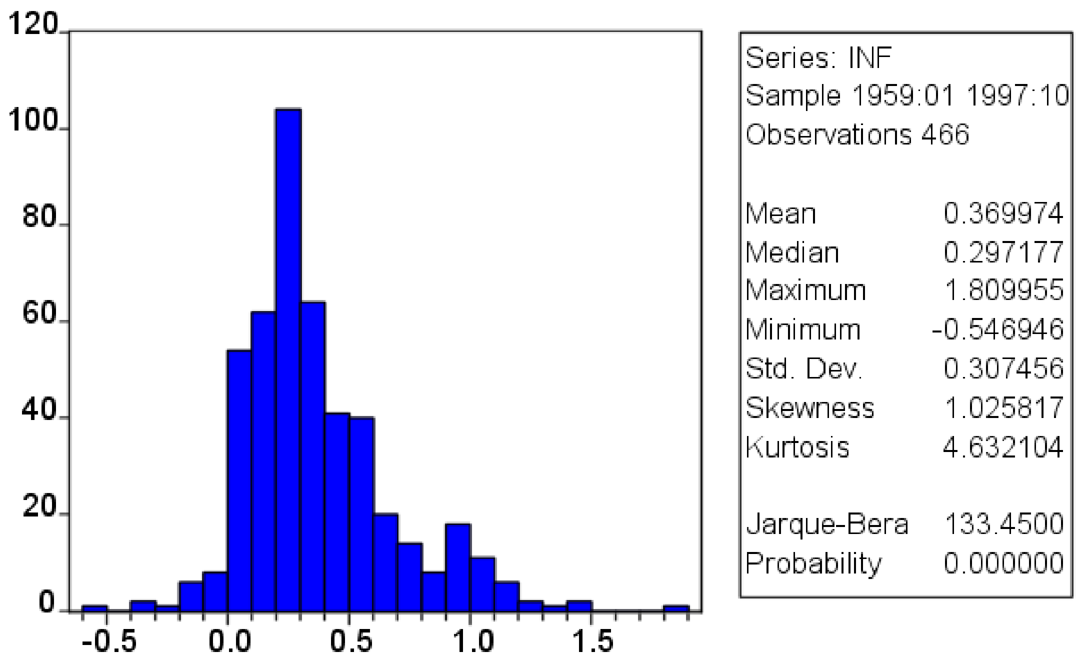

Figure 6: Histogram and descriptive statistics, U.S. Inflation, 1959/1 - 1997/10.

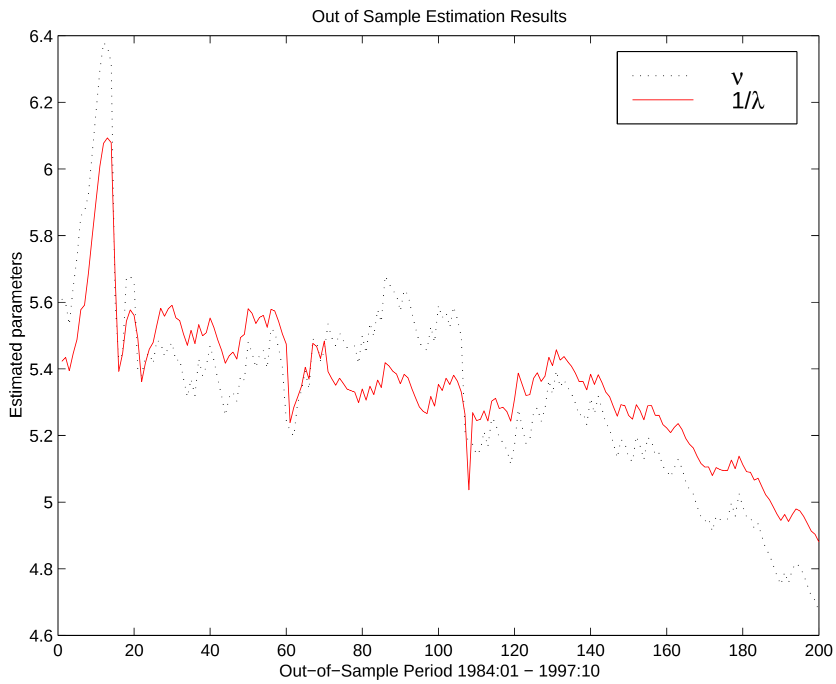

Figure 7: Estimated $\hat{\nu}$ and $1/\hat{\lambda}$ for the out-of-sample period.

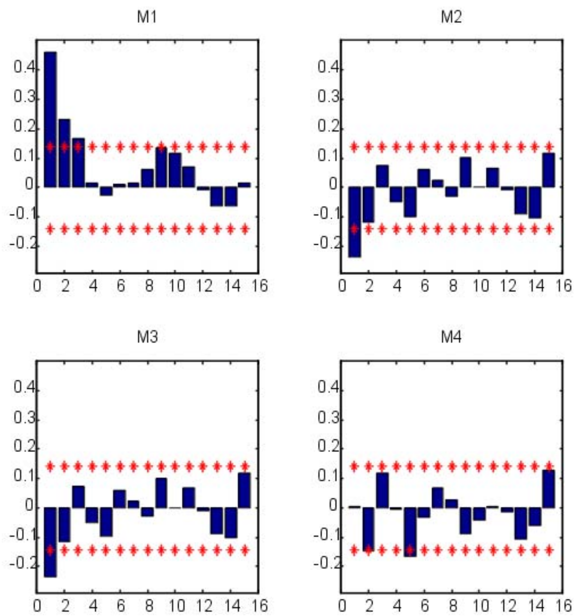

Figure 8: Sample autocorrelogram of $(z_t - \bar{z})$ from the density forecasts of the four models considered. Clockwise, the models are: M1: $\mathrm{N}(\mu, \sigma^2)$, M2: AR(12)-ARCH(1)-normal, M3: AR(12)-ARCH(1)-t, M4: AR(12)-ARCH(1)-hypernormal, $\zeta = 0$.

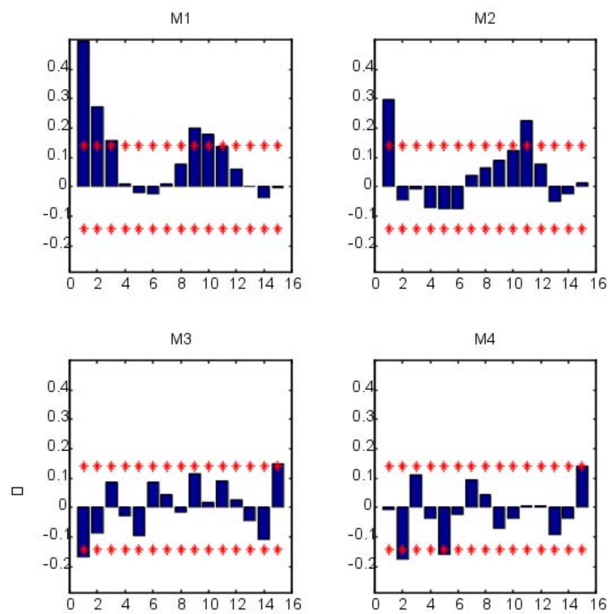

Figure 9: Sample autocorrelogram of $(z_t - \bar{z})^2$ from the density forecasts of the four models considered. Clockwise, the models are: M1: N(μ, σ^2), M2: AR(12)-ARCH(1)-normal, M3: AR(12)-ARCH(1)-t, M4: AR(12)-ARCH(1)-hypernormal, $\zeta = 0$.

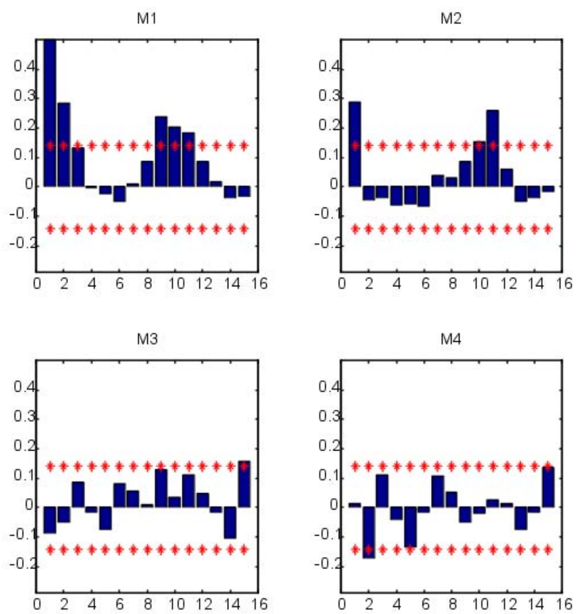

Figure 10: Sample autocorrelogram of $(z_t - \bar{z})^3$ from the density forecasts of the four models considered. Clockwise, the models are: M1: $N(\mu, \sigma^2)$, M2: AR(12)-ARCH(1)-normal, M3: AR(12)-ARCH(1)-t, M4: AR(12)-ARCH(1)-hypernormal, $\zeta = 0$.

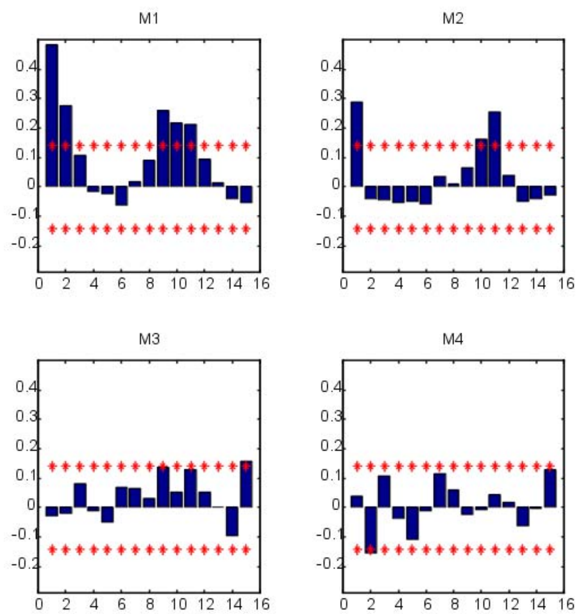

Figure 11: Sample autocorrelogram of $(z_t - \bar{z})^4$ from the density forecasts of the four models considered. Clockwise, the models are: M1: N(μ, σ^2), M2: AR(12)-ARCH(1)-normal, M3: AR(12)-ARCH(1)-t, M4: AR(12)-ARCH(1)-hypernormal, $\zeta = 0$.

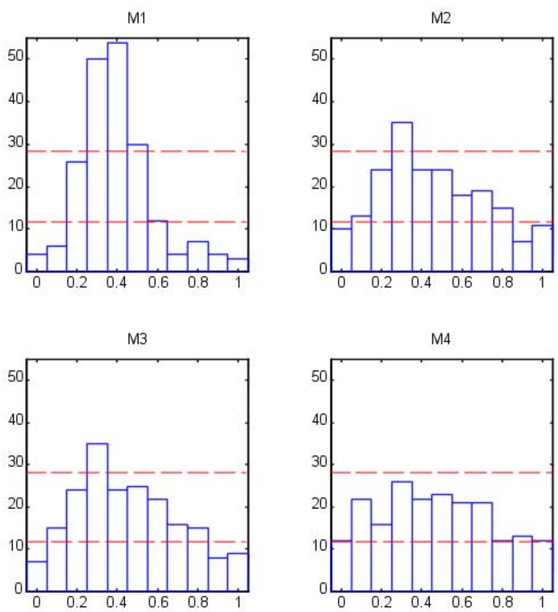

Figure 12: Histograms of the z_t's constructed for the density forecasts of the four models considered. Clockwise, the models are: M1: N(μ, σ^2), M2: AR(12)-ARCH(1)-normal, M3: AR(12)-ARCH(1)-t, M4: AR(12)-ARCH(1)-hypernormal, $\zeta = 0$. The dashed lines indicate 95% confidence intervals computed under the hypothesis that $z_t \sim i.i.d.\ U(0,1)$. See text for details.

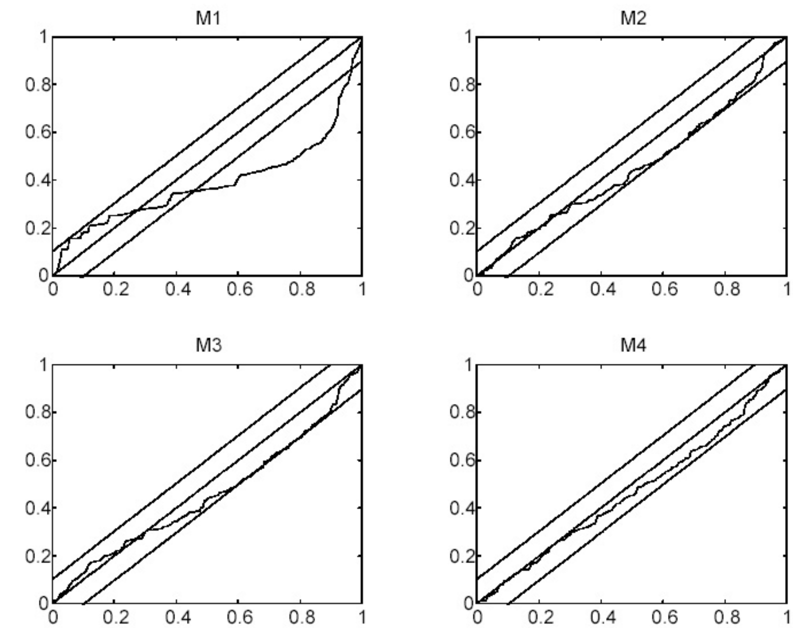

Figure 13: Empirical cdf of the z_t's constructed for the density forecasts of the four models considered. Clockwise, the models are: M1: N(μ, σ^2), M2: AR(12)-ARCH(1)-normal, M3: AR(12)-ARCH(1)-t, M4: AR(12)-ARCH(1)-hypernormal, $\zeta = 0$. The dashed lines indicate 95% confidence intervals are derived using the critical values from the Kolmogorov-Smirnov test for the given sample size.

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 Proofs

Proof of Lemma 1.

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

Apply L'Hopital's rule to obtain

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

■

Proof of Theorem 2.
We consider the general case

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

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

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

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

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

For $m = 0$ this reduces to

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

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

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

We know that $\lambda < 1$ so that from the set of all $\zeta \geq 0$ the left hand side is smallest for $\zeta = 0$. Setting $\zeta = 0$ we obtain

$$1-\lambda > \lambda(1-\lambda),$$

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

Proof of Theorem 3.
Use the proof of Theorem 2 to obtain:

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

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

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

As for theorem 2 we note that the left hand side attains its smallest possible value for $\zeta = 0$ and we obtain:

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

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

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

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

Proof of Theorem 5.

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

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

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

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

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, eds (1953), section 2.5.3) and we obtain

$$\int_0^x (1+\lambda t^2)^{-b} dt = \frac{1}{\sqrt{\lambda}} \left(\frac{\lambda x^2}{1+\lambda x^2} \right)^{1/2} \cdot_2 F_1 \left(\frac{1}{2}, \frac{3}{2} - b; \frac{3}{2}; \frac{\lambda x^2}{1+\lambda x^2} \right) .$$

We can now write $F(x)$ as

$$\begin{aligned} F(x) &= \frac{\kappa_\lambda}{2} + \text{sign}(x) \frac{1}{\sqrt{\lambda}} \left(\frac{\lambda x^2}{1+\lambda x^2} \right)^{1/2} \cdot_2 F_1 \left(\frac{1}{2}, \frac{3}{2} - b; \frac{3}{2}; \frac{\lambda x^2}{1+\lambda x^2} \right) \\ &= \frac{\kappa_\lambda}{2} + \frac{x}{\sqrt{(1+\lambda x^2)}} \cdot {}_2F_1 \left(\frac{1}{2}, \frac{3}{2} - \frac{1-\lambda^{1+\zeta}}{2\lambda(1-\lambda)}; \frac{3}{2}; \frac{\lambda x^2}{\lambda x^2 + 1} \right) . \end{aligned}$$

Normalizing by κ_λ now gives the desired result.

■

Proof of Corollary 6.
Consider the normalizing constant κ_λ. For $\lambda = \frac{1}{2n+3}$ we can write

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

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 7.
Multiplying with $\kappa_{\lambda,\zeta}$ and applying Euler's transformation (Snow (1952), equation II(2)) yields:

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

Direct integration gives

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

and reapplying Euler's transformation gives

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

■

Proof of Theorem 8.
Theorem 3.1 of Gallant and White (1992) delivers the conclusion if

$$\psi_\lambda(x,\theta) = \sum_{j=1}^{q} \beta_j \boldsymbol{h}_{\lambda,\zeta}(\tilde{x}^T\gamma_j) \tag{18}$$

is ℓ-finite. Due to the finitely additive nature of (18) 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 normalizing factor κ_λ for clarity we have the following:

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

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

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

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

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

■

Proof of Corollary 9.
By definition

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

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

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

■

Proof of Theorem 10.
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 (8) are sufficient for this purpose if the activation function g is ℓ-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 11.
This follows directly from Corollary 9 by substituting the functions from Theorem 7.
■
Proof of Corollary 12.
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 Proposition 13 see Diebold et al. (1998), page 868. ■

References

Abadir, K., "An Introduction to Hypergeometric Functions for Economists," *Econometric Reviews*, 1999, *18*, 287–330.

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

Bollerslev, T., "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 1987, *69*, 542–547.

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

Clements, M.P. and J. Smith, "Evaluating the Forecast Densities of Linear and Non-linear Models: Applications to Output Growth and Unemployment," *Journal of Forecasting*, 2000, *19*, 255–276.

Diebold, R.X., A.S. Tay, and K.F. Wallis, "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters," in R.F. Engle and H. White, eds., *Cointegration, Causality, and Forecasting: A Festschrift in Honor of C.W.J. Granger*, Oxford University Press, 1999, pp. 76–90.

____ , T.A. Gunther, and A.S. Tay, "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 1998, *39*, 863–883.

Engle, R.F., "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model," *Journal of Money, Credit, and Banking*, 1983, *25*, 286–301.

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

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

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

Johnson, N.L., "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 1949, *36*, 149–176.

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

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

Magid, A., *Lectures on Differential Galois Theory* University Lecture Series, 7, Provicence, RI: American Mathematical Society, 1994.

McDonald, J., "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 1984, *52*, 647–663.

Rudin, W., *Principles of Mathematical Analysis*, New York: McGraw-Hill, 1964.

Slater, L.J., *Generalized Hypergeometric Functions*, Cambrigde: Cambridge University Press, 1966.

Snow, C., *Hypergeometric and Legendre Functions With Applications to Integral Equations of Potential Theory* National Bureau of Standards Applied Mathematics Series 19, Washington D.C.: United States Department of Commerce, 1952.

Stock, J.H. and M.W. Watson, "Forecasting Inflation," *NBER Working Paper*, 1999, *7023.*

Student, "The Probable Error of a Mean," *Biometrika*, 1908, *6*, 1–25.

Tay, A.S. and K.F. Wallis, "Density Forecasting: A Survey," *Journal of Forecasting*, 2000, *19*, 235–254.

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

Whittaker, E.T. and G.N. Watson, *A Course of Modern Analysis*, Cambridge: Cambridge University Press, 1962.