

Mixtures of t -distributions for Finance and Forecasting*

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

Abstract

We explore convenient analytic properties of distributions constructed as mixtures of scaled and shifted t -distributions. A feature that makes this family particularly desirable for econometric applications is that it possesses closed-form expressions for its anti-derivatives (e.g., the cumulative density function). We illustrate the usefulness of these distributions in two applications. In the first application, we use a scaled and shifted t -distribution to produce density forecasts of U.S. inflation and show that these forecasts are more accurate, out-of-sample, than density forecasts obtained using normal or standard t -distributions. In the second application, we replicate the option-pricing exercise of Abadir and Rockinger (2003) using a mixture of scaled and shifted t -distributions and obtain comparably good results, while gaining analytical tractability.

JEL Classification: C63; C53, C45.

Keywords: ARMA-GARCH models, Neural Networks, Nonparametric Density Estimation, Forecast Accuracy, Option Pricing, Risk Neutral Density.

*The authors wish to thank Karim Abadir, John Bilson, Manfred Deistler, Carl FitzGerald, Jens Jackwerth, Eugene Kandel, Christian Krattenthaler, Andrew Patton, Michael Rockinger, Michael Wolf, and participants of the European Summer Symposium in Financial Markets in Gerzensee for helpful discussions and valuable suggestions. Two anonymous referees and an associate editor provided helpful suggestions that substantially improved this paper. Haefke acknowledges financial support from EU grant HPMF-CT-2001-01252. Part of this paper was written while Haefke was visiting UCLA. He thanks the department of economics for its hospitality. White's participation was supported by NSF grants SBR 9811562 and SES 0111238.

[†] Department of Economics, University of California, Los Angeles. E-mail: giacomini@econ.ucla.edu

[‡] Deutsche Bank AG, Credit Risk Management. E-mail: andreas.gottschling@db.com

[§] [Institute for Advanced Studies, Vienna](#) and [Instituto de Análisis Económico, CSIC](#) , E-mail: haefke@ihs.ac.at

[¶] Department of Economics, University of California, San Diego. E-mail: hwhite@ucsd.edu

1 Introduction

Integrals of particular functions play a central role in economics, econometrics, and finance. For 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. As another 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. 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 calls attention to the very convenient properties of Student's t -distribution in yielding closed form¹ expressions for the cdf and its integrals for particular subsets of parameter values. Even when the expressions for these integrals are not in closed form, they are still analytically quite convenient. In special cases, the inverse cdf (quantile function) also has a closed form expression, which is especially convenient for analyzing Value at Risk. Although a closed form for the t -distribution cdf has been long known (see e.g. Moran (1968, pp 326 – 328)) this convenient property and the convenient expression for the integral of the cdf that we provide have not previously been recognized for their usefulness in economic and financial applications.

Significant flexibility is achieved by considering mixtures of scaled and shifted t -distributions, which inherit all the convenient properties of the t -distribution with respect to cdf's and their integrals. Moreover, these mixtures have the structure of a single hidden layer artificial neural network (ANN). This ANN structure ensures that with sufficiently many terms in the mixture and under suitable regularity conditions, these mixtures are capable of approximating any function and its derivatives to any desired degree of accuracy. This further suggests that mixtures of t -densities may be useful as a substitute for quadrature methods in numerical integration.

Two separate empirical applications illustrate the advantages of using t -distributions

¹ The definition of “closed form” is not universally agreed upon. Here, by “closed form” we mean an expression containing only a finite number of symbols, and including only the operators $+$, $-$, $*$, $/$, and a small list of trigonometric functions, inverse trigonometric functions, factorial and gamma functions, and so forth.

and their mixtures to model economic and financial time series. In the first application, we use a shifted, rescaled t -distribution to produce density forecasts of U.S. inflation and establish whether the new distribution can provide any improvements in the out-of-sample performance of density forecasts relative to the use of the normal or the standard t -distributions. We evaluate the forecasts using the framework suggested by Diebold, Gunther and Tay (1998), which makes use of the antiderivative of the density, thus highlighting the usefulness of the t -distribution in such applications. In the second application, we use a mixture of scaled and shifted t -distributions to estimate risk-neutral densities associated with financial options. In particular, we consider the same option-pricing application as that in Abadir and Rockinger (2003), and obtain comparably good results, while gaining in analytical tractability.

The outline of the paper is as follows. Section 2 sets forth the convenient properties of the t -distribution that are our focus here. Section 3 provides a brief discussion of artificial neural networks and extends the integrability results using mixtures and artificial neural networks. This has the further benefit of bringing conditional densities into our framework. Section 4 presents our application to U.S. inflation forecasts. Section 5 presents the application to option prices. Section 6 concludes. The appendices contain mathematical details and the proofs.

2 A Flexible Family of Density Functions

Student's (1908) t -distribution with ν degrees of freedom has the familiar density

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

where $\Gamma(\cdot)$ is the standard gamma function. It is a standard result that for all $0 < m < \nu - 1$

$$\int_{-\infty}^{\infty} |x|^m t_\nu(x) dx < \infty$$

and that the integer moments are given by:

$$\int_{-\infty}^{\infty} x^m t_\nu(x) dx = \begin{cases} 0 & m \text{ odd} \\ \nu^{\frac{m}{2}} \frac{\Gamma(\frac{\nu-m}{2})\Gamma(\frac{m+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi}} & m \text{ even.} \end{cases}$$

Now we consider the antiderivatives² of t_ν . 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 $D(D^{-1}f) = f$.

² In writing the antiderivatives, the “constant of integration” is here always taken to be zero. In particular applications, other values may be appropriate.

In giving the antiderivative of t_v , 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

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)}{\Gamma(c+k)} \frac{z^k}{k!}. \quad (2)$$

The series converges absolutely for $|z| < 1$, as a ratio test verifies. In our applications, we are interested in the hypergeometric function for any real x . For this, we make use of the transformation $z = \frac{x^2}{x^2 + v}$ which yields $|z| < 1$ (see e.g. equation (89) in Abadir (1999)). 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. Analogous to Amos (1964, Eqn. 15) we obtain the cdf of t_v in closed form making use of the hypergeometric function ${}_2F_1$:

Proposition 1 *Let $t_{v,\sigma}$ be a scale-generalized version of (1) such that*

$$t_{v,\sigma}(x) = \frac{1}{\kappa_{v,\sigma}} \left(1 + \frac{x^2}{\sigma^2 v} \right)^{-\frac{v+1}{2}}, \quad (3)$$

where $\kappa_{v,\sigma}$ denotes the normalization factor:

$$\kappa_{v,\sigma} = \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{\sigma^2 v} \right)^{-\frac{v+1}{2}} dx = \sigma \sqrt{v\pi} \frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})}. \quad (4)$$

Then for all $x \in R$, $\sigma > 0$, and $1 < v < \infty$:

$$D^{-1}t_v(x) = \frac{1}{2} + \frac{x}{\kappa_{v,\sigma} \sqrt{(1 + \frac{x^2}{\sigma^2 v})}} \cdot {}_2F_1\left(\frac{1}{2}, 1 - \frac{v}{2}; \frac{3}{2}; \frac{x^2}{x^2 + \sigma^2 v}\right). \quad (5) \quad \square$$

For nonnegative integers n such that

$$n = \frac{v}{2} - 1 \quad (6)$$

the infinite sum in ${}_2F_1$ terminates after n terms.

The second antiderivative, $D^{-2}t_v$, is also of interest. For example, the price of a European call option with strike K and risk neutral cdf $F(\cdot)$ can be expressed in terms of an integral of the cumulative distribution function (cdf) of risk neutralized asset returns: $C(K) = \int_K^{\infty} [1 - F(S)] dS$. The second antiderivative is given by our next result.

Theorem 1 *Let $t_{v,\sigma}$ be as in 3 and $\kappa_{v,\sigma}$ as in Equation 4. Then for all $x \in R$, $\sigma > 0$, and $1 < v < \infty$:*

$$D^{-2}t_{v,\sigma}(x) = \frac{x}{2} + \frac{v \sqrt{(1 + \frac{x^2}{\sigma^2 v})}}{(v-1)\kappa_{v,\sigma}} {}_2F_1\left(\frac{-1}{2}, 1 - \frac{v}{2}; \frac{1}{2}; \frac{x^2}{x^2 + \sigma^2 v}\right).$$

These expressions also terminate after n terms for all $v = 2(n+1)$. $\square$

3 Mixture Distributions and Artificial Neural Networks

Further flexibility can be achieved by considering mixtures of t -distributions, that is, by taking a convex combination of densities of scaled and shifted standard t -distributions. Just as with mixtures of normals, these mixtures can deliver skewed distributions, distributions with tail properties unachievable by a single t -distribution 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. In addition to analytic tractability, another potential advantage of using the t -distribution instead of the normal to form a mixture is that, because of its greater flexibility, one may require fewer terms (mixing densities) in the convex combination to achieve a given accuracy of approximation to the true density.

We establish our result for mixtures of t -distributions 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:

$$\psi(\mathbf{x}, \beta, \gamma) = \sum_{j=1}^q \beta_j \cdot g(\tilde{\mathbf{x}}^T \gamma_j), \quad (7)$$

where $\tilde{\mathbf{x}} = (1, x_1, x_2, \dots, x_r)$, $\gamma = (\gamma_1^T, \gamma_2^T, \dots, \gamma_q^T)^T$, $\gamma_j \in R^{r+1}$, $\beta = (\beta_1^T, \dots, \beta_q^T)^T$ and $g : R \rightarrow R$ is the hidden unit “activation” function. See Kuan and White (1994) for additional background.

In our discussion of desirable approximation properties, the notion of ℓ -finiteness will be useful:

Definition 1 Let ℓ be a non-negative integer. A function g is ℓ -finite if g is continuously differentiable of order ℓ and has Lebesgue integrable ℓ^{th} derivative. $\square$

Mixtures of the form (7) are able to approximate large classes of functions (and their derivatives) arbitrarily well, for ℓ -finite activation functions g (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 (White, 1996, e.g. Theorem 19.1). The form of (7) delivers not only flexibility, but it also provides the basis 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: $x_2, \dots, x_r$. 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 $g = t_v$. Our next result shows that these mixtures can deliver arbitrarily accurate approximations to a large class of densities under suitable conditions.

Theorem 2 *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*

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

for some integer $m \geq 0$ (for further background see Gallant and White (1992)). For integer $\ell < v - 1$, t_v 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 (8) using a single hidden layer feedforward network of the form

$$\psi_v(x, \theta) = \sum_{j=1}^q \beta_j \cdot t_v(\tilde{x}^T \gamma_j), \quad (9)$$

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

Observe that t_v is always 0-finite by construction.

Corollary 1 *Let $T_v = D^{-e_i} t_v$ denote the antiderivative of t_v with respect to the i -th variable, and let $l \leq u$ be real numbers. Then the integral of the neural net (9) has the form*

$$\int_l^u \psi_v(x, \theta) dx_i = \Psi_v(x_{(i)}(u); \theta) - \Psi_v(x_{(i)}(l); \theta),$$

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

$$\Psi_v(x_{(i)}(a); \theta) = \sum_{j=1}^q \beta_j \cdot T_v(a_{ij}(x_{(i)}(a), \gamma_{ij})),$$

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, $\Psi_{\mathbf{v}}(x_{(i)}(a); \theta)$ has a closed form expression (i.e. terminates after n terms) for all $\mathbf{v}$ of the form $\mathbf{v} = 2n + 2$, $n = 0, 1, 2, \dots$ $\square$

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

The networks $\Psi_{\mathbf{v}}$ of Corollary 1 have desirable approximation properties:

Theorem 3 Let f and $t_{\mathbf{v}}$ be as in Theorem 2, and let $T_{\mathbf{v}}$ be as in Corollary 1. Then for integer $\ell < \mathbf{v}$, $T_{\mathbf{v}}$ 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 $\Psi_{\mathbf{v}}(\cdot)$ given in Corollary 1. $\square$

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

We also have analogs of Corollary 1 and Theorem 3 for the integral of $\Psi_{\mathbf{v}}$.

Corollary 2 Let $\Xi_{i,\mathbf{v}} = D^{-2e_i} t_{\mathbf{v}}$ denote the second antiderivative of $t_{\mathbf{v}}$ with respect to the i -th variable. Let $l \leq u$ be real numbers. Then the integral

$$\int_l^u \Psi_{\mathbf{v}}(x_{(i)}(a); \theta) da$$

has the form

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

In addition, in $\Lambda_{i,\mathbf{v}}$ the series terminates after $n + 1$ terms for all $\mathbf{v}$ of the form $\mathbf{v} = 2n + 2$, $n = 0, 1, 2, \dots$ $\square$

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

Corollary 3 *Let f and t_ν be as in Theorem 1, and let $\Xi_{i,\nu}$ be as in Corollary 2. Then for integer $\ell < \nu + 1$, $\Xi_{i,1\nu}$ 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 $\Lambda_{i,\nu}$ given in Corollary 2.* $\square$

When f is the antiderivative of a cdf, $\Lambda_{i,\nu}$ can approximate it. Its derivatives (the cdf and pdf) can be approximated by the derivatives of $\Lambda_{i,\nu}$ due to the denseness in Sobolev norm and the fact that the associated activation function is always 2-finite. This property is useful in option pricing contexts, for example, as risk neutral densities can be well approximated by fitting networks involving our Ξ 's to the option price and then differentiating twice.

4 An Application to Inflation Density Forecasting

In this section, we investigate the potential usefulness of the scaled and shifted t -distributions in producing density forecasts of U.S. inflation. Our goal is to establish whether we can achieve any improvements in the out-of-sample performance of density forecasts relative to the use of more common but restrictive distributional assumptions for the conditional density of inflation, such as the normal or the standard t -distributions.

The evaluation of the forecasts is based on the framework suggested by Diebold, Gunther and Tay (1998) which utilizes the c.d.f. of the variable of interest. The c.d.f. for the t -distribution is computed easily, which makes this evaluation method particularly suitable for our application.

4.1 Generating time-varying density forecasts

We consider competing one-month-ahead density forecasts of U.S. inflation obtained from conditional parametric models. We use monthly, seasonally unadusted U.S. Consumer Price Index (CPI) data from 1959:1 to 2006:12 available through the St. Louis Fed website. We calculate inflation as the log-12th difference of CPI over the sample period, multiplied by a factor of 100.³

We conduct a specification search for the appropriate model of the conditional mean and variance of inflation on the subsample 1959:2-1985:12, which constitutes the in-sample portion. We consider models within the classes ARMA(p,q), ARMA(p,q)-GARCH(1,1), ARMA(p,q)-ARCH(1) and ARMA(p,q)-EGARCH(1,1) with

³The Augmented Dickey-Fuller unit root test, using 6 lags of the change in the dependent variable, rejects the unit root hypothesis at the 5% level.

$0 \leq p \leq 6$ and $0 \leq q \leq 2$ and select the specification which minimizes the Schwartz BIC information criterion, which is an ARMA(1,1)-ARCH(1) model.

Letting Y_t denote inflation at time t and I_{t-1} the information set available at time $t - 1$, the competing ARMA(1,1)-ARCH(1) forecasting models are thus the following:

Model 1 : $Y_t = c + \phi Y_{t-1} + u_t + \theta u_{t-1}$,

$$u_t = \sqrt{h_t} v_t \quad (10)$$

$$h_t = k + \alpha_1 u_{t-1}^2 \quad (11)$$

and

$$v_t \mid I_{t-1} \sim N(0, 1). \quad (12)$$

Model 2 : Same as Model 1 except

$$v_t \mid I_{t-1} \sim t_{v_2}. \quad (13)$$

Model 3 : Same as Model 2 except

$$v_t \mid I_{t-1} \sim t_{v_3}(\mu, \sigma), \quad (14)$$

where $t_v(\mu, \sigma)$ represents a scaled and shifted version of the standard t -distribution, obtained by substituting $x - \mu$ to x in equation (3). Note that the effect of model 3 relative to model 2, besides introducing the scale factor σ , is to introduce a term $(\frac{\mu}{\sigma}) \sqrt{h_t}$ into the conditional mean of the model.

4.2 Estimation and out-of-sample evaluation

We generate a sample of forecast densities of inflation from these models described in the previous section using a recursive sampling scheme, as, e.g., in Clements and Smith (2000), to allow for the possibility of time-varying densities. We first divide the available sample of monthly U.S. inflation data into two parts, 1959:2-1985:12 and 1986:1-2006:12, with the first part used for estimation and the second part left for out-of-sample evaluation. We estimate the parameters of each model by maximum likelihood over the first sample, and we then use the estimated model to generate a one-step-ahead density forecast. We then augment the estimation sample by adding the following observation, re-estimate the model's parameters, and produce the second density forecast. Continuing in this fashion until all observations from the second part of the sample are utilized results in a sequence of $T = 240$ 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 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 2 and 3 will have parameters v_2 and v_3 that vary with time. The time plots of recursively estimated parameters $\hat{v}_2$ and $\hat{v}_3$ are contained in Figure 1.

The top panel of Figure 1 reveals that the estimate of v_2 decreases over time, suggesting that inflation has fatter tails towards the end of the sample. The lower panel of Figure 1 shows a decrease in $\hat{v}_3$, indicating the same phenomenon.

In line with Clements and Smith (2000) and Diebold et al. (1998) we do not perform diagnostic tests on the estimated models, and we ignore parameter estimation uncertainty. This approach 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 forecasts.

To evaluate the sample of density forecasts, we utilize 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 standard techniques for evaluating point forecasts, which typically assume quadratic loss.

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

$$z_t = \int_{-\infty}^{y_t} p_t(u) du, \quad t = 1, \dots, T \quad (15)$$

where y_t is the realization of inflation at time t and $p_t(y_t)$ the estimated density forecast. Diebold et al. (1998) show that, if the sequence of density forecasts is correctly specified, the corresponding sequence of probability integral transforms 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. $U(0, 1)$ for 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 makes it difficult to sort out the causes of a possible rejection. We therefore consider a number of tests of the i.i.d. $U(0, 1)$ hypothesis, ranging from formal tests to more informal, graphical tests. To test the joint hypothesis of $U(0, 1)$ and identical distribution, we implement the Kolmogorov-Smirnov (KS) test; to test the i.i.d. hypothesis alone, we consider the BDS test (Brock, Dechert, Scheinkman and LeBaron, 1996), the CK test of time reversibility (Chen and Kuan, 2002) and the Breusch-Godfrey LM (LM) test for serial correla-

tion up to 10 lags in the series⁴ $(z_t - \bar{z})^i, i = 1, \dots, 4$.

To assess uniformity, we consider the histogram plot of the z 's, and evaluate its distance from the theoretical p.d.f. of a $U(0, 1)$. We accompany the estimates of the p.d.f. by 95% confidence intervals. 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.

To further evaluate and compare the forecast performance of the three distributions we consider the sequence of time-varying 75% and 99% forecast confidence intervals implied by the three density forecasts of inflation, along with the realizations of inflation over the out-of-sample period. We also formally evaluate the performance of the three interval forecast series by conducting a test of correct coverage (Kupiec, 1995), which establishes whether the realizations of inflation fall within the confidence interval a proportion of times that equals the interval's nominal coverage (i.e., 75% and 99%).⁵

4.3 Results

For each model of inflation, we test the null hypothesis of i.i.d. $U(0, 1)$ for the sequence of probability integral transforms $\{z_t\}_{t=1}^{240}$ of the realizations of inflation with respect to the density forecasts generated by each of the three models described in section 4.1. The results for the KS test are reported in Table 1; results for the BDS test, the CK test, and the LM test are reported in Table 2. Finally, Table 3 reports the results of the test for correct coverage of the 75% and 99% interval forecasts of inflation. The empirical p.d.f. and c.d.f. are shown in Figure 2. The interval forecasts are shown in Figure 3.

4.3.1 Model 1: Normal disturbances

Table 1 reveals that the KS test leads to rejection of the hypothesis of i.i.d. $U(0, 1)$ for the series of z 's, at the 5% confidence level. Rejection of uniform distribution is also confirmed by the histogram of the z 's (Figure 2), which is characterized by some bins that fall outside the 95% confidence interval.

⁴The LM test of serial correlation in the series $(z_t - \bar{z})^i, i = 1, \dots, 4$ is designed to detect misspecifications in the conditional mean, variance, skewness and kurtosis.

⁵ Let $a = (1/T) \sum_{t=1}^T 1(Y_{t+1} \in CI)$ denote the empirical coverage and let α be the nominal coverage. Then the relevant null hypothesis is $H_0 : a = \alpha$ and the likelihood ratio test statistic is $LR = 2[\log(a^{Ta}(1-a)^{T-Ta}) - \log(\alpha^{Ta}(1-\alpha)^{T-Ta})]$, which has an asymptotic χ_1^2 distribution.

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

Model	Test Statistic	p-value
M1:ARMA(1,1)-ARCH(1)-nor	0.0896*	0.0398
M2:ARMA(1,1)-ARCH(1)- t_{v_2}	0.0923*	0.0313
M3:ARMA(1,1)-ARCH(1)- $t_{v_3}(\mu, \sigma)$	0.0796	0.0906

Values of the test statistic and p-values for the Kolmogorov-Smirnov test of the hypothesis of $i.i.d.U(0, 1)$ of the probability integral transforms from each model. A ‘**’ indicates rejection of the null hypothesis at the 5% confidence level.

Table 2: p-values of BDS test, CK test and LM test.

BDS test of $H_0 : \{z_t\} \sim i.i.d.$				
Model	p-value			
M1:ARMA(1,1)-ARCH(1)-nor	0.167			
M2:ARMA(1,1)-ARCH(1)- t_{v_2}	0.233			
M3:ARMA(1,1)-ARCH(1)- $t_{v_3}(\mu, \sigma)$	0.233			
CK test of time reversibility				
Model	p-value			
M1:ARMA(1,1)-ARCH(1)-nor	0.3575			
M2:ARMA(1,1)-ARCH(1)- t_{v_2}	0.3948			
M3:ARMA(1,1)-ARCH(1)- $t_{v_3}(\mu, \sigma)$	0.3252			
LM test of no serial correlation in $(z_t - \bar{z})^k, k = 1, \dots, 4$				
Model	Series			
	$(z_t - \bar{z})$	$(z_t - \bar{z})^2$	$(z_t - \bar{z})^3$	$(z_t - \bar{z})^4$
M1:ARMA(1,1)-ARCH(1)-nor	0.0003*	0.0027*	0.0017*	0.0002*
M2:ARMA(1,1)-ARCH(1)- t_{v_2}	0.0004*	0.0038*	0.0025*	0.0003*
M3:ARMA(1,1)-ARCH(1)- $t_{v_3}(\mu, \sigma)$	0.0004*	0.0058*	0.0011*	0.0003*

p-values for: BDS test of independence implemented using the Matlab routine `bds.m` of Ludwig Kanzler (1998); CK test of time reversibility up to 10 lags (setting the user-defined constant beta equal to 0.5); Breusch-Godfrey LM test of serial correlation up to 10 lags in the series $(z_t - \bar{z})^k, k = 1, \dots, 4$, with test statistic computed as the number of observations times the (uncentered) R^2 from a regression of the series on 10 of its lags. A ‘**’ indicates rejection of the null hypothesis at the 5% confidence level.

Table 3: Test of correct coverage of interval forecasts.

Model	p-value	
	75% nominal coverage	99% nominal coverage
M1:ARMA(1,1)-ARCH(1)-nor	0.001*	0.708
M2:ARMA(1,1)-ARCH(1)- t_{ν_2}	0.000*	0.304
M3:ARMA(1,1)-ARCH(1)- $t_{\nu_3}(\mu, \sigma)$	0.047*	0.304

p-values for test of correct coverage (Kupiec, 1995) of 75% and 99% interval forecasts derived from the three models. A “*” indicates rejection of the null hypothesis of correct coverage at the 5% confidence level. See footnote 5 for details.

Regarding the hypothesis of independence, Table 2 reveals that both the BDS test and the CK test fail to reject the hypothesis of independence. However, the LM test detects serial correlation in all powers of $(z_t - \bar{z})$, overturning the conclusion of the BDS and the CK test (possibly due to low power of the BDS and the CK tests).

From Table 3, we see that Model 1 fails the test of correct coverage for the 75% interval forecast, while it does not fail the test for the 99% interval forecast, at 5% confidence level.

Overall, Model 1 fails on all counts, poorly capturing the dynamics of inflation and assuming a functional form which appears to be misspecified.

4.3.2 Model 2: Student’s t disturbances

As in the case of Model 1, Table 1 shows that the KS 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 analysis of the histogram also suggests that the assumption of standard t disturbances does not achieve significant improvements relative to the assumption of normality. The appearance of the p.d.f. plot for Model 2 is very similar to that for Model 1, with some of the bins in the histogram falling outside the 95% confidence interval.

The results of the test for independence also mirror those for Model 1: the BDS and the CK test fail to reject the hypothesis of independence, while the LM test finds serial correlation in the first four powers of $(z_t - \bar{z})$.

From the perspective of interval forecast performance, the 75% interval forecasts from Model 2 fail the test of correct coverage, whereas the 99% interval forecast has correct coverage, as can be seen in Table 3.

Overall, Model 2 seems not to improve on the performance of Model 1, leading to the conclusion that the standard t -distributional assumption appears overall to

be inadequate.

4.3.3 Model 3: Scaled and shifted t disturbances

Unlike the case of the previous two models, the KS test fails to reject the null hypothesis of i.i.d. $U(0, 1)$ for the sequence of probability integral transforms derived from Model 3, at the 5% confidence level. The apparent superiority of Model 3 over the normal and the standard t is further confirmed by an analysis of the histogram plot in Figure 2, which displays all but one bin falling within the 95% confidence bounds. Thus the probability integral transforms of the density forecasts generated by Model 3 pass most tests of the $U(0, 1)$ hypothesis.

Further, the results in Table 3 suggest that the 99% interval forecasts from Model 3 display correct coverage, and for the 75% interval the hypothesis of correct coverage is rejected only marginally, with p-value equal to 0.047 (in contrast to p-values equal to 0.001 and 0.000 for Models 1 and 2).

However, the results for the tests of independence are analogous to those for the previous two models: although the BDS and CK tests fail to reject the hypothesis of independence, the LM test lead to rejection of the null hypothesis.

In conclusion, density forecasts obtained under the assumption of scaled and shifted t disturbances appear to provide the best approximation for the true density of inflation over the sample considered in the paper. This suggests that the scaled and shifted t -distribution constitutes an improvement over the more common assumptions of normality or standard t -distribution, which generate forecasts that fail all evaluation tests. Nevertheless, all three models apparently fail to adequately capture the dynamics of inflation, as suggested by the rejection of the hypothesis of independence for the probability integral transforms implied by the model. A possible explanation for this failure could be the fact that we kept the specification of the conditional mean and variance fixed throughout the out-of-sample period, while in practice the dynamics of inflation may have changed substantially over time, making the ARMA(1,1)-ARCH(1) model a poor approximation for the data-generating process.⁶

5 An Application to Option Pricing

In this section we illustrate the flexibility of mixtures of scaled and shifted densities and the usefulness of having closed form expressions for the cdf and higher

⁶Given the extent of our specification search, it would not be feasible to conduct a new search for each of the 240 out-of-sample forecast periods.

antiderivatives of these densities.

Under standard assumptions the price of a European call option at time t is given by

$$C_t(K) = e^{-r(T-t)} \int_0^\infty \max(0, S_T - K) f(S_T) dS_T \quad (16)$$

where K denotes the strike price, t denotes current date, T the expiration date, r a risk free discount rate, S_T the price of the underlying asset at expiration, and $f(\cdot)$ the unique risk neutral density of the underlying asset price at expiration. The first term can be interpreted as a discount factor and the integral is just the expected pay-off under the risk neutral probability. For background see e.g. Lamberton and Lapeyre (1996). It is well known in the literature that this risk neutral density can be estimated by estimating the option price function C_t and then differentiating twice (Breen and Litzenberger, 1978):

$$f(S_T) = \left. \frac{d^2}{dK^2} e^{r(T-t)} C_t(K) \right|_{K=S_T} \quad (17)$$

Specifications for $f(\cdot)$ in the literature range from parametric (Melick and Thomas, 1997) and density functionals (Abadir and Rockinger, 2003) to fully nonparametric estimators (Aït-Sahalia and Lo, 1998).

The goal in this section is to derive a closed form expression for the call option price in equation (16) where we assume the risk neutral density to be one of our proposed mixture densities. Once we have a closed form expression, we can then readily estimate the free parameters using suitable nonlinear econometric methods. We follow Abadir and Rockinger (2003) in extracting risk neutral densities for specific day/maturity combinations which yields a model with one endogenous and one exogenous variable.

Integrating (16) by parts we find that the option price is the integral of the survival function from K to ∞ . Solving this integral we find that the call option price can be obtained from

$$C_t(K) = e^{-r(T-t)} D^{-1} F(x) \Big|_{x=K}. \quad (18)$$

Exploiting the linearity of our mixtures we can thus write the call option price as a convex combination of second antiderivatives of the scaled and shifted t -densities:

$$c(K; \mathbf{v}, \mu, \sigma, \beta) = \sum_{j=1}^q \beta_j D^{-2} t_{\mathbf{v}_j, \mu_j, \sigma_j}(x) \Big|_{x=K}, \quad (19)$$

where an expression for the second antiderivative of the scaled and shifted t , $t_{\mathbf{v}_j, \mu_j, \sigma_j}$, is given in Theorem 1. This option price can now be approximated using nonlinear

Table 4: Estimation Results for one unrestricted scaled and shifted t kernel.

Maturity	46	58	98	176
Date	930503	931020	930611	930923
μ	447.522	469.363	456.649	468.274
σ	1.128	1.043	1.187	1.116
ν	2.15	2.13	2.09	2.05
Observations	15	20	17	24
R^2	0.997	0.998	0.996	0.996
$\bar{R}^2$	0.997	0.998	0.996	0.995
BIC	-5.272	12.297	5.058	39.867

least squares estimation by solving

$$\min_{\nu, \mu, \sigma, \beta} \sum_{j=1}^q \left(e^{r(T-t)} C_t(K) - c(K; \nu, \mu, \sigma, \beta) \right)^2.$$

Because the approximation is, under regularity conditions, consistent in Sobolev norm (see e.g. Gallant and White (1992)), we can approximate the derivatives of $C_t(K)$ with those of $\hat{C}_t(K) = e^{-r(T-t)} c(K; \hat{\nu}, \hat{\mu}, \hat{\sigma}, \hat{\beta})$.

To illustrate the usefulness of our method, we use the data described in Aït-Sahalia and Lo (1998) on European options for 1993. The density functionals of Abadir and Rockinger (2003) are based on confluent hypergeometric functions and constitute a natural and ambitious benchmark to which to compare our densities. For this reason we pick exactly the same dates and maturities as they do.

Estimation results are reported in Tables 4–6. Our mixtures of two scaled and shifted t 's obtain a fit comparable to the density functionals of Abadir and Rockinger (2003) while estimating the same number of parameters (seven) as do their density functionals. Our results are strongly superior to those based on a lognormal density.

Aït-Sahalia and Lo (1998) have argued that estimated risk neutral densities exhibit strong kurtosis. Based on our results of Sections 2 and 3, we therefore expect the mixture approach to perform well in this application. When they exist, the mean and the central second to fourth moments of the mixture distribution are given by:

$$\begin{aligned} \mu_0 &\equiv \int_{-\infty}^{\infty} x \psi(x) dx = \sum_{j=1}^q \beta_j \mu_j \\ \int_{-\infty}^{\infty} (x - \mu_0)^2 \psi(x) dx &= \sum_{j=1}^q \beta_j \left\{ \frac{\sigma_j^2 \nu_j}{\nu_j - 2} + (\mu_j - \mu_0)^2 \right\} \end{aligned}$$

Table 5: Estimation Results for optimal mixture of scaled and shifted t kernels.

Maturity	46	58	98	176
Date	930503	931020	930611	930923
β_1	0.47116	0.87635	0.48122	0.47826
β_2	0.52884	0.12365	0.51878	0.52174
μ_1	455.71195	463.88917	439.02152	467.40523
μ_2	436.74045	-180.07235	465.97645	427.68524
σ_1	1.05161	0.71091	1.13348	1.17359
σ_2	1.09967	0.84172	1.08543	0.61103
ν_1	2.27139	1.11081	2.06186	2.00751
ν_2	2.16015	2.40935	21.94908	1.04186
Observations	15	20	17	24
R^2	0.99968	0.99982	0.99959	0.99989
$\bar{R}^2$	0.99943	0.99973	0.99934	0.99985
BIC	-25.45502	-20.35048	-22.00398	-36.21021

Table 6: Estimation Results for Alternative models.

Maturity	46	58	98
Date	930503	931020	930611
Abadir-Rockinger Density Functionals			
R^2	0.999922	0.999911	0.999661
$\bar{R}^2$	0.999864	0.999857	0.999504
Hermite			
R^2	0.997214	0.984918	0.993403
$\bar{R}^2$	0.996750	0.982764	0.992627
Jumps			
R^2	0.997926	0.991013	0.995244
$\bar{R}^2$	0.997580	0.989729	0.994685
Mixtures			
R^2	0.998267	0.990682	0.996039
$\bar{R}^2$	0.997573	0.987577	0.994983
Lognormal Density			
$R^2 = \bar{R}^2$	0.951508	0.928570	0.980671

These statistics are taken from Abadir and Rockinger (2003).

$$\begin{aligned}\int_{-\infty}^{\infty} (x - \mu_0)^3 \psi(x) dx &= \sum_{j=1}^q \beta_j (\mu_j - \mu_0) \left\{ \frac{3\sigma_j^2 \nu_j}{\nu_j - 2} + (\mu_j - \mu_0)^2 \right\} \\ \int_{-\infty}^{\infty} (x - \mu_0)^4 \psi(x) dx &= \sum_{j=1}^q \beta_j \left\{ \frac{3\sigma_j^4 \nu_j^2}{8 - 6\nu_j + \nu_j^2} + \frac{6\sigma_j^2 \nu_j (\mu_j - \mu_0)}{\nu_j - 2} + (\mu_j - \mu_0)^4 \right\}\end{aligned}$$

Our estimation results indicate that not even the second moment exists (cf. [Table 5](#)[ν_j]). In terms of quality of fit the mixture of scaled and shifted t -distributions is comparable to the density functionals of Abadir and Rockinger (2003) where the $\bar{R}^2$ is higher in the fourth decimal. Our mixture of scaled and shifted t densities clearly outperforms all other methods reported in Abadir and Rockinger (2003) as can be seen from [Table 6](#).

6 Conclusion

We explore convenient analytic properties of mixtures of scaled and shifted t -distributions that make these well suited for applications to the analysis of economic and financial time series. Two particularly appealing features of this family are its flexibility and the fact that it possesses analytically convenient expressions for its antiderivatives. We illustrate the usefulness of such features in applications to inflation density forecasting and to option pricing. In the first application, we show that density forecasts of inflation obtained using a scaled and shifted t -distribution are more accurate than forecasts that use the normal or the standard t -distribution. This application makes use of the techniques for density forecast evaluation proposed by Diebold et. al (1998), which rely on computation of the cdf, thus highlighting the desirability of having convenient expressions for the cdf in such cases. The second application replicates results of Abadir and Rockinger (2003), who proposed a flexible family of distributions and illustrated its usefulness in an application to option-pricing. We show that the use of our mixture of distributions allows us to obtain comparably good results, while affording analytical tractability and ease of implementation. These results suggest that models based on mixtures of scaled and shifted t -distributions have a useful role to play in econometrics, given their convenience, generality, and flexibility.

Appendix A Proofs

PROOF of Proposition 1

To establish our result, we take $\nu > 0$. Consider the general case

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

with

$$\begin{aligned}\lambda &= \frac{1}{\sigma^2 \nu} \\ b &= \frac{\nu + 1}{2}.\end{aligned}$$

By symmetry and definition of the normalization factor we have

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

so that for $x < 0$ we can write

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

and for $x > 0$ we can write

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

To evaluate the integral, substitute $(1 - u) = (1 + \lambda t^2)^{-1}$. Then

$$\begin{aligned}t &= \lambda^{-\frac{1}{2}} u^{\frac{1}{2}} (1 - u)^{-\frac{1}{2}} \\ dt &= \frac{1}{2} \lambda^{-\frac{1}{2}} u^{-\frac{1}{2}} (1 - u)^{-\frac{3}{2}} du.\end{aligned}$$

After this substitution the integral becomes:

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

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

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

We can now write $F(x)$ as

$$F(x) = \frac{\kappa_{\nu, \sigma}}{2} + \text{sign}(x) \frac{1}{\sqrt{\lambda}} \left(\frac{\lambda x^2}{1 + \lambda x^2} \right)^{1/2} {}_2F_1 \left(\frac{1}{2}, \frac{3}{2} - b; \frac{3}{2}; \frac{\lambda x^2}{1 + \lambda x^2} \right) .$$

Substituting out b and λ we obtain

$$F(x) = \frac{\kappa_{v,\sigma}}{2} + \frac{x}{\sqrt{(1 + \frac{x^2}{\sigma^2 v})}} {}_2F_1\left(\frac{1}{2}, 1 - \frac{v}{2}; \frac{3}{2}; \frac{x^2}{\sigma^2 v + x^2}\right). \quad \blacksquare$$

Normalizing by $\kappa_{v,\sigma}$ now gives the desired result.

PROOF of Theorem 1

Multiplying with $\kappa_{v,\sigma}$ and applying Euler's transformation (Snow (1952), equation II(2)) yields:

$$\kappa_{v,\sigma} D^{-1} t_{v,\sigma}(x) = \frac{\kappa_{v,\sigma}}{2} + x {}_2F_1\left(\frac{1}{2}, \frac{v+1}{2}; \frac{3}{2}; -\frac{x^2}{\sigma^2 v}\right).$$

Direct integration gives

$$\kappa_{\lambda,\xi} D^{-2} t_v(x) = \frac{x \kappa_{\lambda,\xi}}{2} - \frac{1}{2\lambda(\frac{v-1}{2} - 1)} {}_2F_1\left(\frac{-1}{2}, \frac{v-1}{2} - 1; \frac{1}{2}; -\lambda x^2\right),$$

and reapplying Euler's transformation gives

$$D^{-2} t_v(x) = \frac{x}{2} + \frac{v \sqrt{(1 + \frac{x^2}{\sigma^2 v})}}{(v-1)\kappa_{v,\sigma}} {}_2F_1\left(\frac{-1}{2}, 1 - \frac{v}{2}; \frac{1}{2}; \frac{x^2}{x^2 + \sigma^2 v}\right). \quad \blacksquare$$

PROOF of Theorem 2

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

$$\psi_\lambda(x, \theta) = \sum_{j=1}^q \beta_j t_v(\tilde{x}^T \gamma_j) \quad (20)$$

is ℓ -finite. Due to the finitely additive nature of [Equation 20](#) the result is not vacuous if t_v is ℓ -finite for some ℓ . From the continuity of t_v and $\kappa_\lambda < \infty$ we have that t_v is ℓ -finite for $\ell = 0$. We proceed to verify that t_v is also ℓ -finite for $\ell < v$. Omitting the normalizing factor κ_λ for clarity, we have the following:

1. Some algebra gives $D^\ell t_v(x) = p_\ell(x) \left(1 + \frac{x^2}{v}\right)^{\frac{v-1}{2} - \ell}$ where $p_\ell(x)$ is a polynomial in x of degree ℓ . This is a continuous function for all v and ℓ .

2. $\int_{-\infty}^{\infty} |D^\ell t_v(x)| dx < \infty$ follows from

$$\begin{aligned} \int_{-\infty}^{\infty} |D^\ell t_v(x)| dx &= \int_{-\infty}^{\infty} \left| p_\ell(x) \left(1 + \frac{x^2}{v} \right)^{\frac{v-1}{2}-\ell} \right| dx \\ &\leq \int_{-\infty}^{\infty} |p_\ell(x)| \left(1 + \frac{x^2}{v} \right)^{\frac{v-1}{2}-\ell} dx \\ &\leq \sum_{j=0}^{\ell} \int_{-\infty}^{\infty} |p_{j\ell}(x)| \left(1 + \frac{x^2}{v} \right)^{\frac{v-1}{2}-\ell} dx, \end{aligned}$$

where $p_{j\ell}$ is a monomial of order $j \leq \ell$. As

$$\int_{-\infty}^{\infty} |x|^\ell t_v(x) dx < \infty$$

for $\ell < v-1$, the result follows. ■

PROOF of Corollary 1

By definition

$$\int_l^u \psi_\lambda(x, \theta) dx_i = \sum_{j=1}^q \beta_j \int_l^u t_{v,\sigma}(\tilde{x}^T \gamma_j) dx_i.$$

Let us define

$$\begin{aligned} x &:= \tilde{x}^T \gamma \\ a_{ij}(a, x_{(i)}, \gamma_{ij}) &= a\gamma_{ij} + \sum_{k=1, k \neq i}^{r+1} \tilde{x}_k \gamma_{kj} \\ u_{ij} &:= a_{ij}(u, x_{(i)}, \gamma_{ij}) \\ l_{ij} &:= a_{ij}(l, x_{(i)}, \gamma_{ij}) \\ T_{v,\sigma}(a_{ij}(a, x_{(i)}, \gamma_{ij})) &:= D^{-e_i} t_{v,\sigma}(x), \end{aligned}$$

which allows us to write

$$\begin{aligned} \int_l^u t_{v,\sigma}(\tilde{x}^T \gamma) dx_i &= \frac{1}{\kappa_{v,\sigma}} \int_l^u \left(\frac{(\tilde{x}^T \gamma)^2}{\sigma^2 v} + 1 \right)^{-\frac{v+1}{2}} dx_i \\ &= \frac{1}{\beta_i \kappa_{v,\sigma}} \int_{l_i}^{u_i} \left(\frac{x^2}{\sigma^2 v} + 1 \right)^{-\frac{v+1}{2}} dx. \end{aligned}$$

Defining

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

we may consequently write the desired integral as

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

PROOF of Theorem 3

Theorems 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 (7) are sufficient for this purpose if the activation function g is ℓ -finite. This is shown in Theorem 2 for $t_{v,\sigma}$. Since the ℓ -finiteness of any non-negative function implies the $(\ell + 1)$ -finiteness of its antiderivative, the result follows for $T_{v,\sigma}$. ■

PROOF of Corollary 2

This follows directly from Corollary 1 by substituting the functions from Theorem 1. ■

PROOF of Corollary 3

This result follows from Theorem 2 and Theorem 3 by applying the recursive ℓ -finiteness argument given in the proof of Theorem 3 one more time. ■

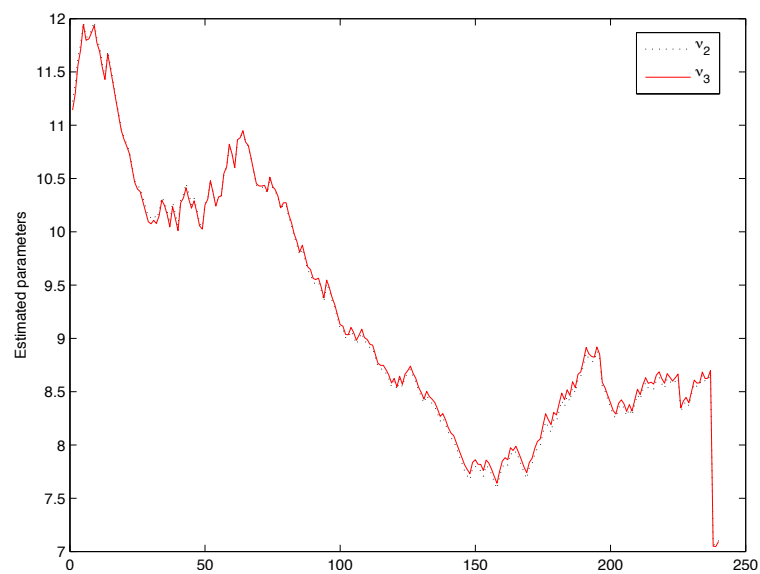
Appendix B Data Description

Data for the inflation application is the monthly consumer price index, all items, not seasonally adjusted. It is freely available from <http://research.stlouisfed.org/fred2/series/CPIAUCSL/downloaddata>. We computed the inflation rate as the 12th log difference of the CPI series.

We obtained the data for the option pricing application from Michael Rockinger, to whom we are most grateful. The data are carefully described in Aït-Sahalia and Lo (1998).

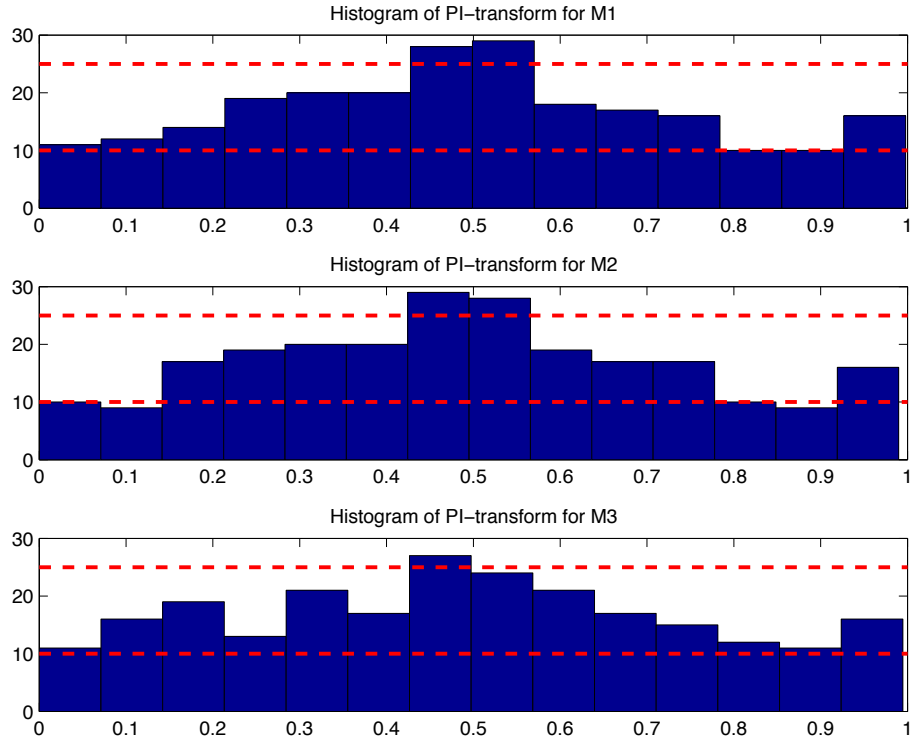
Appendix C Figures

Figure 1: Recursive estimates of $\hat{v}_2$ and $\hat{v}_3$ for the out-of-sample period.



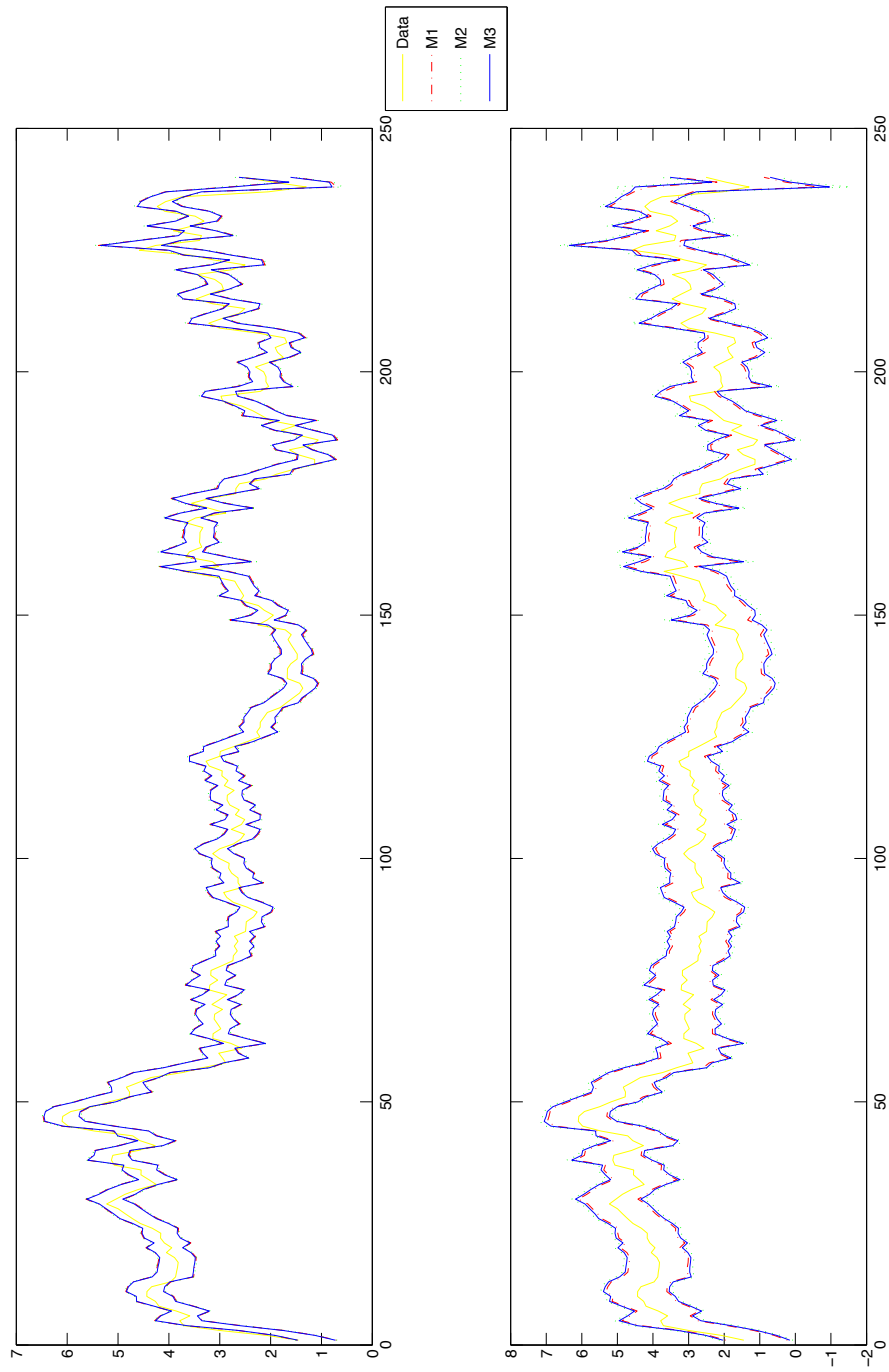
Recursive estimates of $\hat{v}_2$ and $\hat{v}_3$ for the scaled and shifted t-density forecast for the out-of-sample period 1/1986 – 12/2006.

Figure 2: Histograms of probability integral transforms z_t



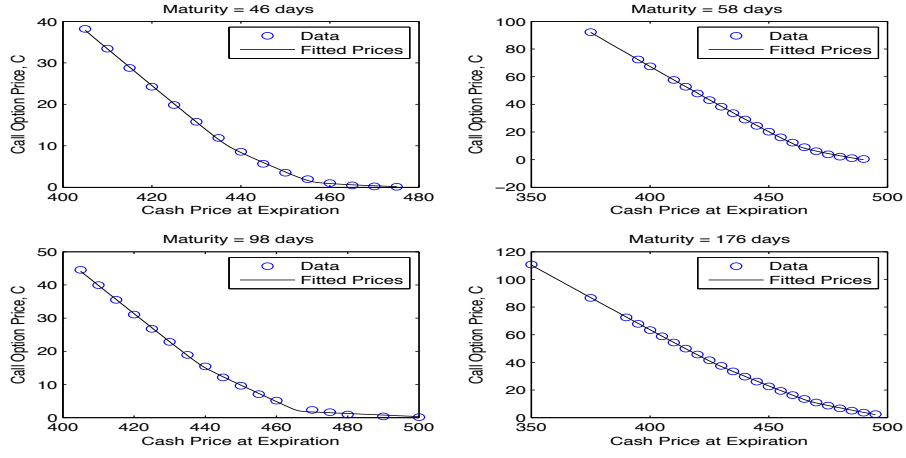
From top to bottom, the three models are $M1 : ARMA(1,1) - ARCH(1) - \text{normal}$; $M2 : ARMA(1,1) - ARCH(1) - t_{\nu_2}$ and $M3 : ARMA(1,1) - ARCH(1) - t_{\nu_3}(\mu, \sigma)$. The dashed lines indicate 95% confidence intervals computed under the hypothesis that $z_t \sim i.i.d.U(0,1)$. See text for details.

Figure 3: Out of Sample Inflation Forecasts



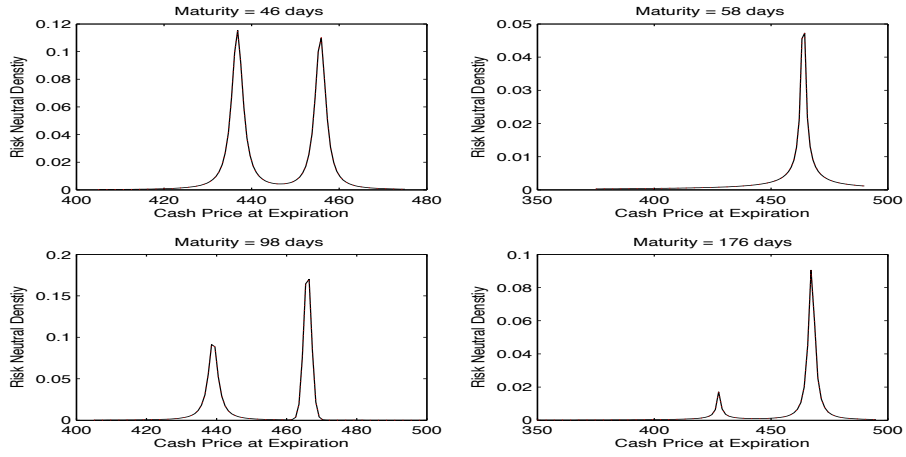
This figure plots the realizations of inflation over the out-of-sample period together with the corresponding 75% (top panel) and 99% (bottom panel) interval forecasts implied by the three models described in [Section 4.1](#).

Figure 4: Actual and fitted option prices, scaled and shifted t .



Call prices for the mixture density estimator for various times-to-maturity as a function of the option's underlying asset price.

Figure 5: Estimated Risk Neutral Densities, scaled and shifted t .



These estimates of the risk neutral densities are derived from the mixture of densities estimator. The second derivative of the call pricing function is evaluated at S_T and we plot the function $f(S_T)$ for day/maturity combinations chosen by Abadir and Rockinger (2003).

References

- Abadir, K.**, "An Introduction to Hypergeometric Functions for Economists," *Econometric Reviews*, 1999, 18, 287–330.
- and **M. Rockinger**, "Density Functionals, With an Option-Pricing Application," *Econometric Theory*, 2003, 19, 778–811.
- Aït-Sahalia, Y. and A.W. Lo**, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 1998, LIII, 499–547.
- Amos, D.E.**, "Representations of the central and non-central t distributions," *Biometrika*, 1964, 51, 451–458.
- Bailey, W.N.**, *Generalized Hypergeometric Series*, New York: Strechert-Hafner Service Agency, 1962.
- Breeden, D.T. and R. Litzenberger**, "Prices of State Contingent Claims Implicit in Option Prices," *Journal of Business*, 1978, 51, 621–657.
- Brock, W., W. Dechert, J. Scheinkman, and B. LeBaron**, "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews*, 1996, 15, 197–235.
- Chen, Y.T. and C.M. Kuan**, "Time Irreversibility and EGARCH effects in US stock index returns," *Journal of Applied Econometrics*, 2002, 17, 565–578.
- 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., T.A. Gunther, and A.S. Tay**, "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 1998, 39, 863–883.
- 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.

- Kuan, C.M. and H. White**, "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 1994, 13, 1–91.
- Kupiec, P.**, "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 1995, 2, 173–184.
- Lamberton, D. and B. Lapeyre**, *Introduction to Stochastic Calculus Applied to Finance*, London: Chapman & Hall, 1996.
- Melick, W.R. and C.P. Thomas**, "Recovering an asset's implied PDF from option prices: An application to crude oil during the Gulf crisis," *Journal of Financial and Quantitative Analysis*, 1997, 32, 91–116.
- Moran, P.A.P.**, *An Introduction to Probability Theory*, Oxford: Clarendon Press, 1968.
- Slater, L.J.**, *Generalized Hypergeometric Functions*, Cambridge: 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.
- Student**, "The Probable Error of a Mean," *Biometrika*, 1908, 6, 1–25.
- White, H.**, "Parametric Statistical Estimation with Artificial 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.