# Spectral backtests unbounded and folded

**Michael B. Gordy and Alexander J. McNeil**

**2024-060**

# Spectral backtests unbounded and folded[*]

Michael B. Gordy

Federal Reserve Board, Washington DC


Alexander J. McNeil

School for Business and Society, University of York


15 July 2024

### Abstract

In the spectral backtesting framework of Gordy and McNeil (2020) a probability measure on the unit interval is used to weight the quantiles of greatest interest in the validation of forecast models using probability-integral transform (PIT) data. We extend this framework to allow general Lebesgue-Stieltjes kernel measures with unbounded distribution functions, which brings powerful new tests based on truncated location-scale families into the spectral class. Moreover, by considering uniform distribution preserving transformations of PIT values the test framework is generalized to allow tests that are focused on both tails of the forecast distribution.

*JEL* Codes: C52; G21; G32
*Keywords*: Backtesting; Volatility; Risk management

# 1 Introduction

Gordy and McNeil (2020) study a class of backtests for forecast distributions in which the test statistic depends on a spectral transformation of a quantile exceedance indicator function. The spectral transformation weights quantile exceedance events using a kernel measure which is chosen by the validator to reflect the validator's priorities for model performance. The present paper extends the original treatment in two directions. First, whereas Gordy and McNeil (2020) restrict the kernel measure to the class of probability measures, in this paper we allow the kernel measure to be unbounded, subject to an integrability condition. We show that unbounded kernels deliver tests materially more powerful than tests based on the bounded kernels studied by Gordy and McNeil (2020). Second, we introduce a pre-processing of the data by a folding transformation that leaves the size of the backtest unaltered but increases its power against misspecifications of forecast volatility that are extremely common in practice.

Our extensions to the spectral backtesting framework are germane to any validation exercise in which performance throughout one or both tails of the forecast distribution is of special interest. Our investigation is motivated by recent developments in the capital regulation of the trading operations of large banks. Under the current Basel III rules (Basel Committee on Bank Supervision, 2019), minimum capital requirements for a bank's trading book are determined by the bank's self-reported daily Expected Shortfall (ES) at the $\alpha_1 = 97.5\%$ confidence level. The adoption of ES departs from earlier Basel regimes tied to Value-at-Risk (VaR) at the $\alpha^* = 99\%$ confidence level. Left unchanged in Basel III is the role of the regulator in validation of the bank's model through backtesting. For this purpose, banks in the United States report to regulators for each trading day the probability associated with the realized profit-and-loss (P&L) in the prior day's forecast distribution, i.e., the probability integral transform (PIT) associated with realized P&L. Observing the PIT values is equivalent to observing VaR exceedances *at every level* $\alpha \in [0, 1]$. Besides Gordy and McNeil (2020), bank-reported PIT data have been studied by Lynch et al. (2023)

and Iercosan et al. (2023).

Under a VaR-based regime, the regulator would have particular interest in testing model performance over some range of confidence levels in the neighborhood of $\alpha^*$. Accordingly, Gordy and McNeil (2020) illustrate their methods with kernels placing mass in a window $[\alpha_1, \alpha_2]$ with $0 < \alpha_1 < \alpha^* < \alpha_2 < 1$, e.g., $[0.985, 0.995]$. In such a setting, only bounded measures produce finite test statistics, and since our test statistic is invariant to the measure of the window, without loss of generality we can restrict attention to probability measures. Because ES is an integral of VaR above a threshold level, it is natural under the new regime to consider continuous kernels that weight on every $\alpha$ above some threshold, e.g., $\alpha_1 = 0.975$ in the Basel context. In this setting, even some unbounded measures can be guaranteed to yield valid test statistics. Further, because bank models tend to break down under extreme market events and unbounded measures weight most heavily on such tail events, we expect unbounded measures to deliver more powerful tests. We confirm this intuition in simulation exercises and show as well that this power does not come at the expense of size distortions.

The topic of backtesting expected shortfall has led to a lively debate about whether or not ES is amenable to backtesting (Gneiting, 2011; Acerbi and Székely, 2014; Fissler et al., 2016; Acerbi and Székely, 2023). A growing literature, including Patton et al. (2019) and Barendse et al. (2023), employs elicitability theory to develop joint backtests of VaR and ES. For regulatory use, this methodology would generally require that banks submit time series of both VaR and ES estimates, although a recent paper of Bayer and Dimitriadis (2022) suggests a workaround to obtain a test of ES estimates only, at the possible expense of some model misspecification.

Issues related to backtesting estimates of risk measures such as ES are sidestepped in our framework because we test the *forecast distributions* from which risk measures are estimated, rather than the estimates themselves. It may be noted that a number of recent papers propose PIT-based approaches to backtesting expected shortfall and, in particular, exploit the *cumulative violation process* of Du and Escanciano (2017), which can be viewed as a

particular choice of spectral transformation. These include Du et al. (2023), who propose an improved conditional ES backtest, Hoga and Demetrescu (2023), who propose a real-time monitoring procedure for ES forecasts and Hué et al. (2024) who use orthogonal polynomials to jointly test moment conditions for the cumulative violation process and the process of durations between VaR exceedances.

Even if the regulator is interested exclusively in the upper tail of the PIT distribution, it is often the case that models that are misspecified in the upper tail may be similarly misspecified in the lower tail. For example, in a risk management setting, a failure to capture stochastic volatility in the distribution of financial returns leads to underestimation of extreme gains as well as extreme losses. Berkowitz et al. (2011) and O'Brien and Szerszen (2017) provide evidence of neglected stochastic volatility in the banking context by showing that simple GARCH models fitted to bank P&L often outperform bank internal models. Expressed in terms of the observed PIT values, such a misspecification would produce too few middling PIT and too many low and high PIT. Thus, even if the regulator is concerned only with large losses, a kernel that assigns no weight to the lower tail of the PIT distribution fails to capture data that may be relevant to detecting misspecification in the upper tail. We show how the regulator can pre-process the PIT values, by an operation we describe as *folding*, so that tail values from left and right in the original PIT distribution are mapped to the upper tail of the pre-processed distribution without altering the distribution of the test statistic under the null hypothesis.

A simple example of a suitable pre-processor would apply the v-shaped mapping $T(u) = |1 - 2u|$ to the PIT values. Under this mapping, the event that a pre-processed PIT value is in the upper tail, $T(\text{PIT}) \in [v, 1]$, is equivalent to the event that the PIT lies in a union of intervals in both tails, $\{\text{PIT} \in [0, (1-v)/2] \cup [(1+v)/2, 1]\}$. It is straightforward to see that if the PIT are in fact uniformly distributed (as under the null hypothesis of the backtest) then the transformed PIT are uniformly distributed as well. The linear symmetric mapping $T(u) = |1 - 2u|$ is only a single example of a very large class of uniform distribution preserving

(u.d.p.) transformations. A common finding in the empirical literature is that the distribution of market returns is asymmetric such that the tail of large losses is heavier than the tail of large gains, a phenomenon that led to the development of asymmetric GARCH-type models incorporating both leverage effects and skewed innovation distributions, including AGARCH (Engle, 1990), EGARCH (Nelson, 1991) and GJR-GARCH (Glosten et al., 1993). We show that asymmetric members of the u.d.p. class can be chosen to highlight model skewness as well as kurtosis.

In Section 2, we extend the backtesting framework of Gordy and McNeil (2020) to allow for unbounded kernels and u.d.p. folding transformations. A key result demonstrates that folding is not redundant, i.e., pre-processing delivers backtests that cannot otherwise be obtained. In Section 3 we introduce two novel families of unbounded kernels. Monte Carlo simulations demonstrate that these kernels deliver backtests that are well-sized and highly sensitive to unmodelled kurtosis. A parsimonious but flexible family of v-shaped pre-processors is introduced in Section 4. Monte Carlo simulations show how pre-processing further highlights unmodelled kurtosis. Pre-processors can be effective as well in the presence of unmodelled skewness. However, in the absence of material excess kurtosis, a poorly chosen pre-processor can mask rather than enhance the signature of model misspecificaton. Section 5 offers guidance on implementation in practical settings.

## 2 Extended spectral backtesting

### 2.1 Backtesting set-up

We assume that a forecaster models portfolio losses $(L_t)$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{N}_0}, \mathbb{P})$ where $\mathcal{F}_t$ represents the information available to the forecaster at time $t$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and $\mathbb{N}$ denotes the non-zero natural numbers.[1] For any time $t \in \mathbb{N}$, the loss $L_t$ is an $\mathcal{F}_t$-measurable random variable with conditional distribution function (df) given by

---

[1] $L_t$ is the negative value of P&L, so large losses are associated with the right tail of the distribution.

$F_t(x) = \mathbb{P}\left(L_t \leqslant x \mid \mathcal{F}_{t-1}\right)$. In most applications this distribution is not time-invariant, due to serial dependencies in $(L_t)$ and changes in the composition of the portfolio over time.

At time $t$ the forecaster builds a model $\widehat{F}_t$ of $F_t$ based on the information $\mathcal{F}_{t-1}$. PIT-values are the random variables $(P_t)$ obtained by setting $P_t = \widehat{F}_t(L_t)$. If the models $\widehat{F}_t$ form a sequence of ideal probabilistic forecasts in the sense of Gneiting et al. (2007), i.e., coinciding with the conditional laws $F_t$ of $L_t$ for every $t$, then the result of Rosenblatt (1952) implies that the process $(P_t)$ is a sequence of iid standard uniform variables. PIT-values contain information about exceedances of quantile estimates at any level $u$: if $\widehat{\mathrm{VaR}}_{u,t} = \widehat{F}_t^{\leftarrow}(u)$ denotes the estimate of the $u$-quantile of $F_t$ calculated using the generalized inverse of $\widehat{F}_t$ at probability level $u$, then $P_t \geqslant u \iff L_t \geqslant \widehat{\mathrm{VaR}}_{u,t}$.

We adopt the position of an external model validator, such as a regulator, who uses the PIT-values $(P_t)$ to take a decision on the quality of the forecasting methodology. For the purposes of this paper, we assume that the validator has access only to these PIT values although this restriction could be relaxed considerably. What is essential is that the validator does not observe the entire distribution $\widehat{F}_t$ which reflects the reality of most regulatory regimes. Further, for brevity, we consider only tests of unconditional coverage. Application of unbounded measures and folding pre-processors would apply without complication to the tests of conditional coverage described in Gordy and McNeil (2020).

## 2.2 Spectral backtests

The model validator employs a spectral transformation of the PIT values of the form

$$W_t = \int_I \mathbb{1}_{\{T(P_t) \geqslant u\}} \mathrm{d}\nu(u) \tag{1}$$

where (i) $\nu$ is a Lebesgue-Stieltjes measure referred to as the *kernel measure* and (ii) $T : I \to I$ is a *uniform distribution preserving (u.d.p.) transformation*; if $U \sim U(0,1)$ is a standard uniform random variable and $T$ a u.d.p. transformation, then $T(U) \sim U(0,1)$. Throughout

the paper, $I$ denotes the unit interval $[0, 1]$.

In Gordy and McNeil (2020) the measure $\nu$ was restricted to be a probability measure and the transformation $T$ was simply the identity transformation $T(v) = v$. This set-up was appropriate for a focus on the right tail of the forecast distribution. By looking at PIT exceedances of levels $u$ and using the probability measure $\nu$ to select and weight levels of interest $u$ at the upper end of the unit interval, test statistics were derived that that were sensitive to forecast model specification at a range of quantiles in the right tail.

With any Lebesgue-Stieltjes measure $\nu$ on domain $I$, there is an associated increasing right-continuous function $G_\nu$, referred to as a distribution function (df), such that $\nu([0, u]) = G_\nu(u)$. It is easily seen that (1) is equivalent to the closed-form expression

$$W_t = \nu([0, T(P_t)]) = G_\nu(T(P_t)) \tag{2}$$

which shows that $W_t$ is increasing in $T(P_t)$. Note that we employ df in a generalized sense, since $G_\nu$ is a *probability* df only if $\lim_{u \to 1} G_\nu(u) = 1$. To streamline the presentation, we will henceforth impose the following mild regularity condition on $\nu$.

**Assumption 1.** $G_\nu$ has at most a finite set of discontinuities and is otherwise absolutely continuous.

The univariate transformation extends naturally to the multivariate case in which a set of distinct kernel measures $\nu_1, \ldots, \nu_m$ is applied to PIT-values to obtain the vector-valued variables $\boldsymbol{W}_1 \ldots, \boldsymbol{W}_n$ where

$$\boldsymbol{W}_t = (W_{t,1}, \ldots, W_{t,m})', \quad W_{t,j} = \nu_j([0, T(P_t)]) = G_j(T(P_t)), \; j = 1, \ldots, m. \tag{3}$$

We refer to any backtest based on $\boldsymbol{W}_1 \ldots, \boldsymbol{W}_n$ as a *spectral backtest*. The null hypothesis addressed by an unconditional spectral backtest is

$$H_0: \quad \boldsymbol{W}_t \sim F_W^0 \tag{4}$$

7

where $F_W^0$ denotes the df of $\boldsymbol{W}_t$ when $P_t$ is uniform. In Gordy and McNeil (2020) two types of tests were considered: spectral Z-tests based on central limit theorem arguments and spectral likelihood-ratio tests (LR-tests). The results showed a number of advantages of the former over the latter, including better control of size for similar or superior power, ease of implementation and speed of execution. In this paper we focus on Z-tests and provide the necessary extension of the theory to the Lebesgue-Stieltjes case.[2]

When $\dim \boldsymbol{W}_t = m$ a spectral Z-test is based on the fact that under the multivariate CLT $\sqrt{n}\left(\overline{\boldsymbol{W}}_n - \boldsymbol{\mu}_W\right) \xrightarrow[n\to\infty]{d} N_m(\boldsymbol{0}, \Sigma_W)$ where $\overline{\boldsymbol{W}}_n = n^{-1}\sum_{t=1}^n \boldsymbol{W}_t$ and $\boldsymbol{\mu}_W$ and $\Sigma_W$ are the mean vector and covariance matrix of the null distribution $F_W^0$. Hence a test can be based on assuming for large enough $n$ that

$$T_n = n\left(\overline{\boldsymbol{W}}_n - \boldsymbol{\mu}_W\right)' \Sigma_W^{-1} \left(\overline{\boldsymbol{W}}_n - \boldsymbol{\mu}_W\right) \sim \chi_m^2, \tag{5}$$

where we refer to $T_n$ as an $m$-spectral Z-test statistic. When $m = 1$ the chi-squared test is equivalent to a two-sided test based on

$$Z_n = \frac{\sqrt{n}(\overline{W}_n - \mu_W)}{\sigma_W} \xrightarrow[n\to\infty]{d} N(0,1) \tag{6}$$

where $\mu_W = \mathbb{E}(W_t)$ and $\sigma_W^2 = \mathrm{var}(W_t)$ are the moments in the null model $F_W^0$ for $W_t$.

By definition, the u.d.p. transformation $T(P)$ does not alter the moments of $P$ under the null hypothesis. Thus, as in Gordy and McNeil (2020), the first moment $\mu_W$ of the transformed PIT-values $W_t$ is easily obtained as

$$\mu_W = \int_I (1-u)\mathrm{d}\nu(u) \tag{7}$$

The variance $\sigma_W^2$ of $W_t$ and the cross-moments in the covariance matrix $\Sigma_W$ of $\boldsymbol{W}_t$ are obtained using a simple product rule for spectrally transformed PIT values.

---

[2]The theory of spectral LR-tests presented in Gordy and McNeil (2020) carries through in the more general case without the need for any significant modification.

**Theorem 2.1.** *The set of spectrally transformed PIT values defined by $W_{t,j} = G_j(T(P_t))$ is closed under multiplication. The product $W_t^* = W_{t,1}W_{t,2}$ is given by $W_t^* = G^*(T(P_t))$ where $\nu^*$ is a Lebesgue-Stieltjes measure and the associated function $G^*$ satisfies*

$$G^*(u) = \int_{[0,u]} \left(G_2(s) - \frac{1}{2}\nu_2(\{s\})\right) \mathrm{d}\nu_1(s) + \int_{[0,u]} \left(G_1(s) - \frac{1}{2}\nu_1(\{s\})\right) \mathrm{d}\nu_2(s).$$

It follows that $\sigma_W^2 = \mu_{W^*} - \mu_W^2$, where $\mu_{W^*}$ is found by applying (7) under the measure $\nu^*$ obtained when $\nu_1 = \nu_2 = \nu$. This yields

$$\mu_{W^*} = \int_I (1 - u)\left(2G_\nu(u) - \nu(\{u\})\right) \mathrm{d}\nu(u). \tag{8}$$

The central limit theorem underpinning the Z-test requires finite second moments. For the univariate case, the following proposition provides a sufficient condition on the tail behavior of $G_\nu$.

**Proposition 2.2.** *If $G_\nu(u) = \mathcal{O}((1-u)^{-0.5+\epsilon})$ as $u \to 1$ for some $\epsilon > 0$, then $\sigma_W^2$ is finite.*

In the multivariate setting, the asymptotic distribution in (5) holds if the condition in Proposition 2.2 is satisfied for each $\nu_j$, $j = 1, \ldots, m$.

## 2.3 Uniform distribution preserving transformations

We are interested in u.d.p. transformations $T$ that can extend our testing framework to uncover deficiencies in forecast models that are not revealed by the identity transformation (general theory for u.d.p. transformations can be found in Porubský et al., 1988, among others). Since the choice of kernel measure in our framework is quite flexible, one might ask whether the insertion of any given u.d.p. transformation $T$ in (1) delivers a new Z-test that could not be obtained by changing the kernel measure.

To this end we introduce the concept of *redundancy* in the test framework. Let $\mu$ and $\sigma$ be the moments associated with kernel $\nu$. We say that a u.d.p, transformation $T$ is *redundant*

for kernel $\nu$ if there exists another kernel $\tilde{\nu}$ with moments $\tilde{\mu}$ and $\tilde{\sigma}$ that always delivers the same magnitude $|Z_n|$ for the test-statistic in (6). That is, let $\{P_1, \ldots, P_n\}$ be an arbitrary sample of PIT values, and let $\overline{W}_n = (1/n) \sum_{i=1}^n G_\nu(T(P_i))$ and $\widetilde{W}_n = (1/n) \sum_{i=1}^n G_{\tilde{\nu}}(P_i)$. Then $T$ is redundant if

$$\frac{|\widetilde{W}_n - \tilde{\mu}|}{\tilde{\sigma}} = \frac{|\overline{W}_n - \mu|}{\sigma}$$

almost surely.

As a simple example, consider the u.d.p. transformation $T(v) = 1 - v$.

**Lemma 2.3.** *The u.d.p. transformation $T(v) = 1 - v$ is redundant for any bounded measure $\nu$ and not redundant for any unbounded measure $\nu$.*

Further examples of non-redundant transformations are obtained by considering u.d.p. transformations that are *folding*. By this we mean transformations $T$ for which almost all values $u \in I$ are associated with multiple values in the preimage of $\{u\}$ under $T$. To make this precise we introduce some additional notation and give a definition. For a generic function $f : D \to Y$ and for a set $D_1 \subseteq D$ we write $f[D_1]$ to mean the image of $D_1$ under $f$; similarly, for a set $Y_1 \subseteq Y$, $f^{-1}[Y_1]$ is the preimage of $Y_1$ under $f$.

**Definition 2.4.** For a u.d.p. transformation $T : I \to I$ let $I_T \subseteq I$ be the set defined by $I_T = \{u | \operatorname{card}(T^{-1}[\{u\}]) \geq 2\}$. $T$ is *folding* if $I_T$ has Lebesgue measure one.

For example, the u.d.p. transformation $T(v) = |1 - 2v|$ has $I_T = I \setminus \{0.5\}$ and is clearly folding. The folding class includes v-shaped, m-shaped, w-shaped and more general saw-shaped functions. Our general result for the folding class is

**Proposition 2.5.** *Let $\nu$ be a measure for which $G_\nu$ is strictly increasing on a sub-interval of $I$. If $T$ is a folding u.d.p. transformation then it is not redundant for $\nu$.*

The intuition is that $G_{\tilde{\nu}}(P)$ must be weakly increasing in $P$, which implies that lower and upper tail observations contribute in opposite signs (thereby offsetting one another) in the sample test statistic. By contrast, when we pre-process the PIT values with a folding

u.d.p. transformation, $G_\nu(T(P))$ cannot be monotonic in $P$. PIT values from non-contiguous regions of the unit interval will map to the same value of $G_\nu(T(P))$. Which PIT observations contribute in the same sign and which offset each other depends on the shape of the pre-processor.

# 3    Two families of Lebesgue-Stieltjes kernels

We consider some possibilities for novel kernels which are not necessarily probability measures. For notational simplicity we present the theory for the case where $T$ is the identity transformation.

For the remainder of the paper, the tests we consider are based on dfs $G_\nu$ with densities $g_\nu$ satisfying $g_\nu(u) > 0$ for $\alpha_1 < u < \alpha_2$ and $g_\nu(u) = 0$ for $u < \alpha_1$ and $u > \alpha_2$. In certain cases we allow for mass at the boundaries, i.e., $\nu(\{\alpha_i\}) \geqslant 0$. We refer to the interval $[\alpha_1, \alpha_2]$ as the *kernel window*.

**Remark 3.1.** For unbounded measures, we have $G_\nu(u) \to \infty$ as $u \to \alpha_2$. In such cases, only $\alpha_2 = 1$ is admissible. Were we to choose $\alpha_2 < 1$, we would have $\Pr(\alpha_2 < P_t \leqslant 1) = (1 - \alpha_2) > 0$ under the null hypothesis, so the first moment $\mu_W$ would be infinite.

## 3.1    Simple kernels of power form

Gordy and McNeil (2020) observe that the beta-type density $(u - \alpha_1)^{a-1}(\alpha_2 - u)^{b-1}$ provides a flexible yet parsimonious and tractable form for the density of $G_\nu$. Since that paper restricted $\nu$ to the set of probability measures, it was necessary to restrict $a > 0, b > 0$ and to regularize the kernel by the beta function $B(a, b)$. Here we relax the restriction on $b$ and discard the regularization.

For $u$ in the unit interval, let $B(u; a, b)$ denote the (unregularized) incomplete beta function

$$B(u; a, b) = \int_0^u x^{a-1}(1 - x)^{b-1}\, dx. \tag{9}$$

We define the beta kernel $\nu$ via the df

$$G_\nu(u) = B\left(\frac{(\alpha_1 \vee u \wedge \alpha_2) - \alpha_1}{\alpha_2 - \alpha_1}; a, b\right).$$

This kernel is purely continuous, i.e., $\nu(\{\alpha_1\}) = \nu(\{\alpha_2\}) = 0$. When $b > 0$, $B(u; a, b)$ is bounded from above by $B(a, b)$ so $W_t = G_\nu(P_t)$ certainly has finite moments. However, when $b \leqslant 0$, $B(u; a, b)$ is unbounded as $u \to 1$ and, by Remark 3.1, we have to set $\alpha_2 = 1$. Moreover, the existence of moments has to be checked with the help of Proposition 2.2 and the following result.

**Proposition 3.2.** *As $u \to 1$,*

$$B(u; a, b) = \begin{cases} \mathcal{O}((1-u)^b) & b < 0, \\ \mathcal{O}(-\ln(1-u)) & b = 0. \end{cases}$$

In combination with Proposition 2.2, it follows that $W_t = G_\nu(P_t)$ has finite first and second moments if and only if $b > -1/2$. In the case $b = 0$ we note that

$$-\ln(1-u) = \mathcal{O}((1-u)^k), \qquad u \to 1, \tag{10}$$

for any $k < 0$, a fact that is used a number of times in the following sections. The $b = 0$ case is particularly important for practical application. For small $|b|$, standard algorithms for $B(u; a, b)$ may be numerically unstable for $u$ near 1. However, for $b = 0$, González-Santander (2021, Theorem 1) provides a finite series expansion in elementary functions.

We perform Monte Carlo analyses to explore how the size and power of spectral backtests with beta-type kernels depend on the beta parameters $(a, b)$. We consider four different choices for the df $F$ of the true model of $L_t$: the standard normal, and the scaled $t_{10}$, scaled $t_5$ and scaled $t_3$. The Student t distributions are scaled to have variance one so differences stem from different tail shapes rather than different variances. We take the forecaster's

12

model $\widehat{F}$ to be the standard normal, i.e., we transform the sampled $L_t$ to PIT-values as $P_t = \Phi(L_t)$. Therefore, when the samples of $L_t$ are drawn from the standard normal, the PIT-values are uniformly distributed and are used to evaluate the size of the tests. The PIT samples arising from the Student t distributions show the kind of departures from uniformity that are observed when the forecaster's model is too thin-tailed.

We fix a kernel window of $[\alpha_1, 1]$ for $\alpha_1 = 0.975$. Our sample size is fixed to $n = 500$ corresponding to two-year samples of trading day returns. Our tables report the percentage of rejections of the null hypothesis at the 5% confidence level based on $2^{16} = 65{,}536$ replications. All reported $p$-values are based on two-sided tests.

| Parameters | (1,1) | (2,1) | (1,$^1\!/_4$) | (1,$^1\!/_8$) | (1,0) | (2,0) | (5,0) |
|---|---|---|---|---|---|---|---|
| Normal | 4.7% | 4.6% | 4.6% | 4.5% | 4.4% | 4.3% | 4.9% |
| Scaled $t_{10}$ | 13.7% | 19.4% | 24.1% | 28.6% | 34.2% | 40.8% | 45.1% |
| Scaled $t_5$ | 21.2% | 34.0% | 45.7% | 55.0% | 64.6% | 72.2% | 76.4% |
| Scaled $t_3$ | 13.1% | 28.7% | 46.5% | 61.3% | 75.0% | 82.2% | 86.5% |

Table 1: Size and power of tests based on beta monokernels.
Kernel window is $[0.975, 1]$. 2^16 trials with 500 observations per trial.

Results for univariate beta kernels are reported in Table 1. For all sets of beta parameters, tests are well-sized. For each alternative true model $F$, we find that power increases as $b$ declines and $a$ increases. The magnitude of the effect is extremely large. For example, against the scaled $t_3$ alternative, the rejection rate increases from 13.1% for the uniform (beta$(1,1)$) kernel to 75.0% for the beta$(1,0)$ to 86.5% for the beta$(5,0)$. This pattern is confirmed across a finer grid of beta parameters for the case of the Student $t_5$ in Figure 1.

To understand this pattern, observe that for any two PIT values $\alpha_1 < p_1 < p_2 < 1$, the ratio of the beta kernels $g_\nu(p_2)/g_\nu(p_1)$ decreases in $b$ and increases in $a$. The higher this ratio, the greater the weight in the test on PIT in the neighborhood of $p_2$ relative to PIT in the neighborhood of $p_1$. As shown in Figure 2, within the kernel window of $[0.975, 1]$, the distributions of PIT under the scaled Student t alternatives differ most from the distribution under the null (green solid line) as we move deeper into the tail. Thus, we generally expect
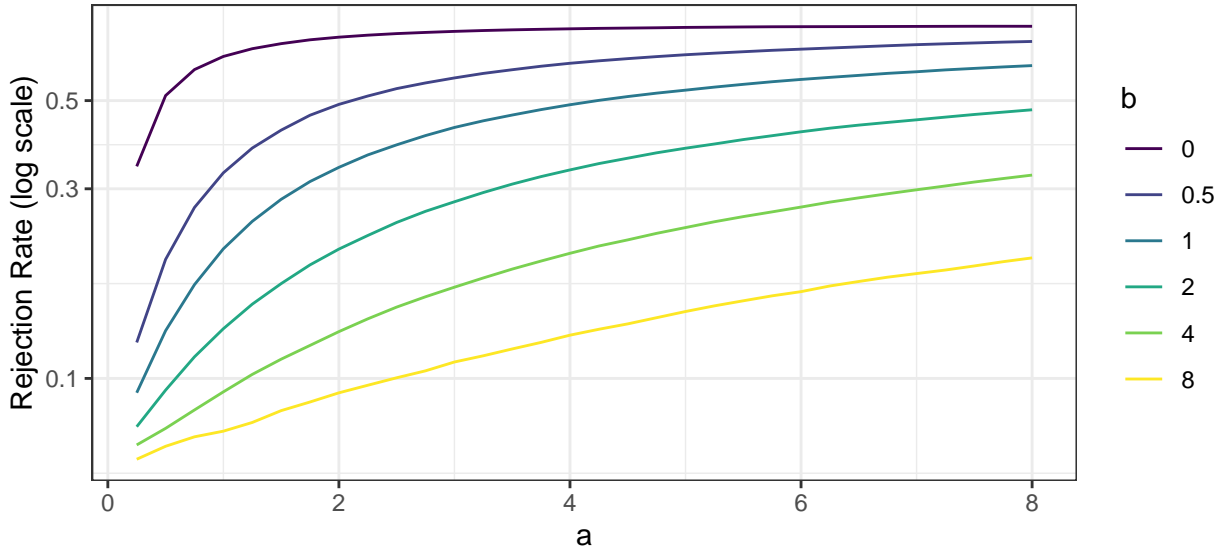
Figure 1: Power of tests based on beta monokernels against scaled $t_5$ alternative. Kernel window is $[0.975, 1]$. 2^16 trials with 500 observations per trial.

tests that weight more heavily on the right-hand tail to deliver higher power.

Results for bivariate beta kernels are reported in Table 2. Gordy and McNeil (2020) demonstrated that bikernel tests are generally more powerful than monokernel tests when the component kernels of the bivariate test emphasize opposite ends of the kernel support. Put another way, the lower the correlation between $W_{t,1}$ and $W_{t,2}$, the greater the additional information gain in introducing the second kernel. Gordy and McNeil (2020) illustrate the point by way of a bikernel, given mnenomic ZPP, with component parameter pairs (25,1) and (1,25), which they showed outperformed bivariate kernels of lesser curvature. Consistent with their results, we find that ZPP offers higher power than the bivariate kernel with two linear densities (parameter pairs (2,1) and (1,2)) against all three Student t alternatives. However, by allowing for unbounded beta kernels (i.e., $b = 0$), we can obtain even better performance without resorting to extreme values of parameter $a$. For example, the bivariate kernel with component parameters (2,0) and (1,3) outperforms ZPP on all three Student t alternatives.
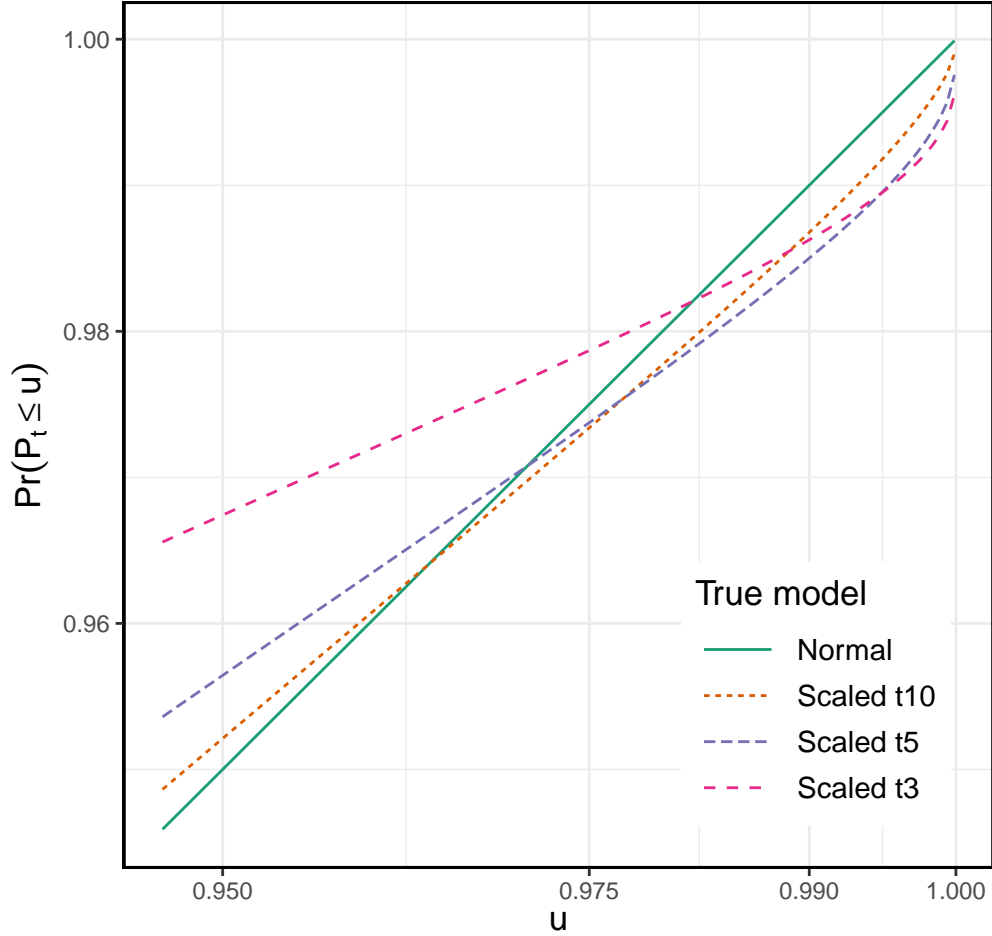
14

Figure 2: Distribution functions for reported PIT-values.
Dfs for the reported PIT-values when the forecaster assumes standard normal losses ($\widehat{F} = \Phi$) but the true loss model $F$ is standard normal, scaled $t_{10}$, scaled $t_5$, or scaled $t_3$.

| Parameters: | (2,1) | (25,1) | (2,0) | ($^5/_2$,0) | ($^9/_2$,0) |
| | (1,2) | (1,25) | (1,3) | ($^1/_2$,3) | ($^1/_2$,6) |
| --- | --- | --- | --- | --- | --- |
| Normal | 4.8% | 5.5% | 5.4% | 5.4% | 5.5% |
| Scaled $t_{10}$ | 22.5% | 38.0% | 41.2% | 41.3% | 42.5% |
| Scaled $t_5$ | 47.3% | 69.7% | 74.3% | 74.5% | 75.3% |
| Scaled $t_3$ | 64.1% | 84.7% | 88.2% | 88.5% | 89.0% |

Table 2: Size and power of tests based on beta bikernels.
Kernel window is [0.975, 1]. 2^16 trials with 500 observations per trial. Gordy and McNeil (2020) assign mnemonic ZPP to the parameter pair (25,1), (1,25).

## 3.2 Kernels derived from truncated location-scale families

In this section we look at bispectral tests that arise as score tests in truncated location-scale families (TLSFs). A test of this kind based on the normal distribution was proposed in a PhD thesis by Lok (2017) and can be viewed as a Z-test analog of the test of Berkowitz (2001), which is a likelihood-ratio test of the uniformity of PIT values in an interval against a family of non-uniform alternatives. We show below that a wide class of TLSFs yield viable bispectral tests.

Let $\mathcal{R}(\mu, \sigma)$ denote a family of continuous probability distributions with location parameter $\mu$ and scale parameter $\sigma$ (which need not be equal to the mean and standard deviation) and let $R$ denote the df of the $\mathcal{R}(0,1)$ distribution and $\rho$ its density. Examples include the cases where $R$ is standard normal (Lok, 2017), standard logistic or Gumbel distribution. If we assume that $R^{-1}(P_t) \sim \mathcal{R}(\mu, \sigma)$ and write $\boldsymbol{\theta} = (\mu, \sigma)'$, the df and density of $P_t$ are respectively[3]

$$F_P(p \mid \boldsymbol{\theta}) = R\left(\frac{R^{-1}(p) - \mu}{\sigma}\right), \quad f_P(p \mid \boldsymbol{\theta}) = \frac{\rho\left(\frac{R^{-1}(p) - \mu}{\sigma}\right)}{\rho(R^{-1}(p))\sigma}, \quad p \in [0,1], \tag{11}$$

and the uniform distribution corresponds to $\boldsymbol{\theta} = \boldsymbol{\theta}_0 = (0,1)'$.

The truncated location-scale family corresponding to the window $[\alpha_1, \alpha_2] \subseteq [0,1]$ is the mixed probability distribution of the truncated random variable $P_t^* = \alpha_1 \vee P_t \wedge \alpha_2$. This is described by the density $f_P(p \mid \boldsymbol{\theta})$ on $[\alpha_1, \alpha_2)$, an atom of size $F_P(\alpha_1, \boldsymbol{\theta})$ at $\alpha_1$ if $\alpha_1 > 0$, and an atom of size $\overline{F}_P(\alpha_2, \boldsymbol{\theta})$ at $\alpha_2$ if $\alpha_2 < 1$; if $\alpha_2 = 1$ then $P_t^* = \alpha_1 \vee P_t$ and the density $f_P(p \mid \boldsymbol{\theta})$ applies to the closed interval $[\alpha_1, 1]$.

If $\mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t)$ denotes the likelihood contribution of the PIT observation $P_t$ in this

---

[3]The assumption $R^{-1}(P_t) \sim \mathcal{R}(\mu, \sigma)$ imposes no restriction on the underlying distribution $F_t$ for the loss $X_t$ or for the modeler's belief $\widehat{F}_t$. An auxiliary variable $\tilde{X}_t = R^{-1}(P_t)$ can be conceived as a quasi-loss in the sense that $X_t$ and $\tilde{X}_t$ will be comonotonic but otherwise unconnected in their marginal distributions.

truncated model, then the score vector is given by

$$S_t(\theta) = \left( \frac{\partial}{\partial \mu} \ln \mathcal{L}_{P^*}(\theta \mid P_t), \frac{\partial}{\partial \sigma} \ln \mathcal{L}_{P^*}(\theta \mid P_t) \right)'. \tag{12}$$

Let $\overline{S}_n(\theta_0) = \frac{1}{n} \sum_{t=1}^{n} S_t(\theta_0)$ be the mean of the observed score vectors under the null. Standard likelihood theory implies that $\sqrt{n} \overline{S}_n(\theta_0) \xrightarrow[n\to\infty]{d} N_2(0, \Upsilon(\theta_0))$ under the null, where $\Upsilon(\theta)$ denotes the covariance matrix of $S_t(\theta)$, i.e., the Fisher information matrix. For large $n$ the score test statistic satisfies

$$n \overline{S}_n(\theta_0)' \Upsilon(\theta_0)^{-1} \overline{S}_n(\theta_0) \sim \chi_2^2. \tag{13}$$

Computation of the score vector and the information matrix is detailed in Appendix D.

We now give conditions under which the score test (13) can be viewed as a bispectral Z-test with kernel measures $\nu_1$ and $\nu_2$ given by sums of discrete and continuous parts. The following assumption will be imposed in our key result.

**Assumption 2.** The df $R$ underlying the TLSF score test is absolutely continuous with log-concave density $\rho$ and support $\mathbb{R}$.

**Theorem 3.3.** *Let $0 < \alpha_1 < \alpha_2 \leqslant 1$ and assume that Assumption 2 holds. Let $\lambda_\rho$ denote the function $\lambda_\rho(x) = -\rho'(x)/\rho(x)$. Then the equation*

$$x \left( \frac{\rho(x)}{R(x)} + \lambda_\rho(x) \right) - 1 = 0, \tag{14}$$

*has a unique root $\underline{x} > 0$ and, provided that $\alpha_1 \geqslant \underline{\alpha} \equiv R(\underline{x})$, the score vector $S_t(\theta_0)$ satisfies $S_t(\theta_0) = W_t - \mu_W$ where $W_{t,i} = G_i(P_t)$,*

$$G_i(u) = \gamma_{i,1} \mathbb{1}_{\{u \geqslant \alpha_1\}} + \gamma_{i,2} \mathbb{1}_{\{u \geqslant \alpha_2, \alpha_2 < 1\}} + \int_0^u g_i(x) \mathrm{d}x, \tag{15}$$

*the densities $g_i$ are given by the derivatives*

$$g_1(u) = \frac{\mathrm{d}}{\mathrm{d}u} \left( \lambda_\rho(R^{-1}(u)) \right) \mathbb{1}_{\{\alpha_1 \leqslant u \leqslant \alpha_2\}}, \qquad g_2(u) = \frac{\mathrm{d}}{\mathrm{d}u} \left( R^{-1}(u)\lambda_\rho(R^{-1}(u)) \right) \mathbb{1}_{\{\alpha_1 \leqslant u \leqslant \alpha_2\}}, \quad (16)$$

*the constants $\gamma_{i,j}$ are given by the non-negative values*

$$\gamma_{1,1} = \frac{\rho\left(R^{-1}(\alpha_1)\right)}{\alpha_1} + \lambda_\rho\left(R^{-1}(\alpha_1)\right), \qquad \gamma_{2,1} = R^{-1}(\alpha_1)\gamma_{1,1} - 1$$

$$\gamma_{1,2} = \frac{\rho\left(R^{-1}(\alpha_2)\right)}{1 - \alpha_2} - \lambda_\rho\left(R^{-1}(\alpha_2)\right), \qquad \gamma_{2,2} = R^{-1}(\alpha_2)\gamma_{1,2} + 1$$

*and $\boldsymbol{\mu}_W = \alpha_1^{-1}\rho(R^{-1}(\alpha_1)) \left(1, \ R^{-1}(\alpha_1)\right)'$.*

**Remark 3.4.** If the density $\rho$ is not log-concave, we would typically have upper as well as lower bounds on the window over which the score test gives a well-defined bispectral test. For elaboration and illustration, see Appendix B.

We now consider the two cases $\alpha_2 < 0$ and $\alpha_2 = 1$ separately. The former case is straightforward since the $W_{t,i}$ variables are bounded, guaranteeing that the elements of $\Upsilon(\boldsymbol{\theta}_0)$ are finite. In this case it would be possible to normalize the measures $\nu_i$ to be probability measures by dividing by $G_i(\alpha_2)$ although there is no practical advantage in doing so.

If $\alpha_2 = 1$, it follows from Theorem 3.3 and formula 15 that for $u \in [\alpha_1, 1]$ the dfs $G_i$ have the forms

$$G_1(u) = \rho(R^{-1}(\alpha_1))/\alpha_1 + \lambda_\rho(R^{-1}(u))$$
$$G_2(u) = R^{-1}(\alpha_1)\rho(R^{-1}(\alpha_1))/\alpha_1 + R^{-1}(u)\lambda_\rho(R^{-1}(u)). \tag{17}$$

Since $\lambda_\rho$ is an increasing function and $R^{-1}(u) \to \infty$ as $u \to 1$, we can infer that $G_1(u) \leqslant G_2(u)$ and that $G_2(u) \to \infty$ as $u \to 1$. In this case $G_2$ is unbounded and we cannot normalize the measure $\nu_2$ to be a probability measure. We need to verify that the condition of Proposition 2.2 is satisfied for $G_2$ to be sure that the elements of $\Upsilon(\boldsymbol{\theta}_0)$ are finite.

**Remark 3.5.** While $G_2$ is always unbounded, $G_1$ can be bounded if $\lim_{x \to \infty} \lambda_\rho(x)$ is finite. For example, this occurs when $\rho$ is the logistic density or the Gumbel density. These cases are analysed in Examples 3.8 and Examples 3.9 below; in both cases $\lim_{x \to \infty} \lambda_\rho(x) = 1$.

**Remark 3.6.** If a TLSF distribution $R$ is skewed, then there exists a complementary TLSF distribution of opposite skew with df $R_c(x) = 1 - R(-x)$. If $\rho$ is log-concave, then so is $\rho_c$. To simplify implementation, we can exploit the relationships $\rho_c(x) = \rho(-x)$, $\lambda_c(x) = -\lambda_\rho(-x)$, and $R_c^{-1}(p) = -R^{-1}(1-p)$.

The main limitation on the application of bispectral tests based on TLSF score test is the requirement that $\alpha_1 \geqslant \underline{\alpha}$ where $\underline{\alpha} = R(\underline{x})$ and $\underline{x} > 0$ solves (14). The fact that $\alpha_1 > R(0)$ shows that a portion of the interval $[0, 1]$ must be eliminated from consideration. We illustrate the TLSF score test and the constraint on $\alpha_1$ with several examples of log-concave densities. Expressions for $\Upsilon(\boldsymbol{\theta}_0)$ and other computational details are found in the Online Supplement (Appendix 3).

**Example 3.7** (Test based on normal distribution). When $\mathcal{R}(\mu, \sigma) = N(\mu, \sigma^2)$ and $R = \Phi$, $P_t$ is said to follow a probitnormal distribution on the unit interval. This distribution also appears as the nesting model in the well-known LR-test of Berkowitz (2001). The normal distribution has a log-concave density with $\lambda_\rho(x) = x$ and (14) has unique root $\underline{x} \approx 0.84$. The conditions for a bispectral test are satisfied if $\alpha_1 \geqslant \underline{\alpha} = \Phi(\underline{x}) \approx 0.80$, which is unlikely to bind in application to the range of tail probability levels of practical interest. From (17) we have that, when $\alpha_2 = 1$,

$$G_1(u) \sim \Phi^{-1}(u) \quad \text{and} \quad G_2(u) \sim \Phi^{-1}(u)^2, \qquad \text{as } u \to 1.$$

Because $G_2(u) \sim -2 \ln(1-u)$ as $u \to 1$ (Abramowitz and Stegun, eds, 1965, eq. 26.2.22), it follows from (10) and Proposition 2.2 that second moments are finite.

**Example 3.8** (Test based on logistic distribution). The standard logistic distribution has $R(x)$ given by the logistic function $S(x) = 1/(1 + \exp(-x))$. The density is log-concave

with $\lambda_\rho(x) = S(x) - S(-x)$ and (14) has unique root $\underline{x} \approx 1.28$. A bispectral test may be constructed for $\alpha_1 \geqslant \underline{\alpha} \approx 0.78$. Interestingly, the first kernel density function satisfies $g_1(u) = 2$, implying constant weighting in the kernel window, while, for $\alpha_2 = 1$, (17) implies that $G_2(u) \sim -\ln(1-u)$ as $u \to 1$ and we argue as in the normal case that second moments are finite.

**Example 3.9** (Two tests based on Gumbel distribution). Unlike the normal and logistic distributions, the Gumbel distribution is skewed, so weights left and right tails asymmetrically. Both the standard Gumbel (positive skew) and complementary Gumbel (negative skew) have log-concave density. For the standard Gumbel we have $\lambda_\rho(x) = 1 - \exp(-x)$ and (14) has unique root $\underline{x} = 1$. A bispectral test may be constructed for $\alpha_1 \geqslant \underline{\alpha} = R(1) = e^{-1/e} \approx 0.69$. The first kernel density function is the decreasing function $g_1(u) = u^{-1}\mathbb{1}_{\{\alpha_1 \leqslant u \leqslant \alpha_2\}}$ while it may be easily deduced from (17) that the df of the second kernel satisfies $G_2(u) \sim -\ln(-\ln u)$ as $u \to 1$ when $\alpha_2 = 1$. Since $-\ln u \sim 1 - u$ as $u \to 1$ we once again have that $G_2(u) \sim -\ln(1-u)$ as $u \to 1$, guaranteeing finiteness of second moments.

For the complementary Gumbel a bispectral test can be constructed for $\alpha_1 \geqslant \underline{\alpha}_c \approx 0.87$. The first kernel density function is the increasing function $g_1(u) = (1-u)^{-1}\mathbb{1}_{\{\alpha_1 \leqslant u \leqslant \alpha_2\}}$ while it may be deduced from (17) that $G_2(u) \sim \ln\left(\frac{1}{1-u}\right)\ln\left(\ln\left(\frac{1}{1-u}\right)\right)$.

**Example 3.10** (Family of tests based on the logistic-beta distribution). The test based on the logistic distribution of Example 3.8 is a special case in a larger family of tests based on the logistic-beta distribution for which the df is a composition of the beta df $\mathcal{I}(z; a, b)$ and the logistic function, i.e.,

$$R(x) = \mathcal{I}(S(x); a, b)$$

where $a > 0$ and $b > 0$. The standard logistic is the case where $a = b = 1$.

The density is log-concave. Cumulants are known in closed-form, from which is it easily verified that the skew of the distribution has the same sign as $a - b$. By the reflection symmetry property of the beta df, the complementary distribution of the logistic-beta $(a, b)$
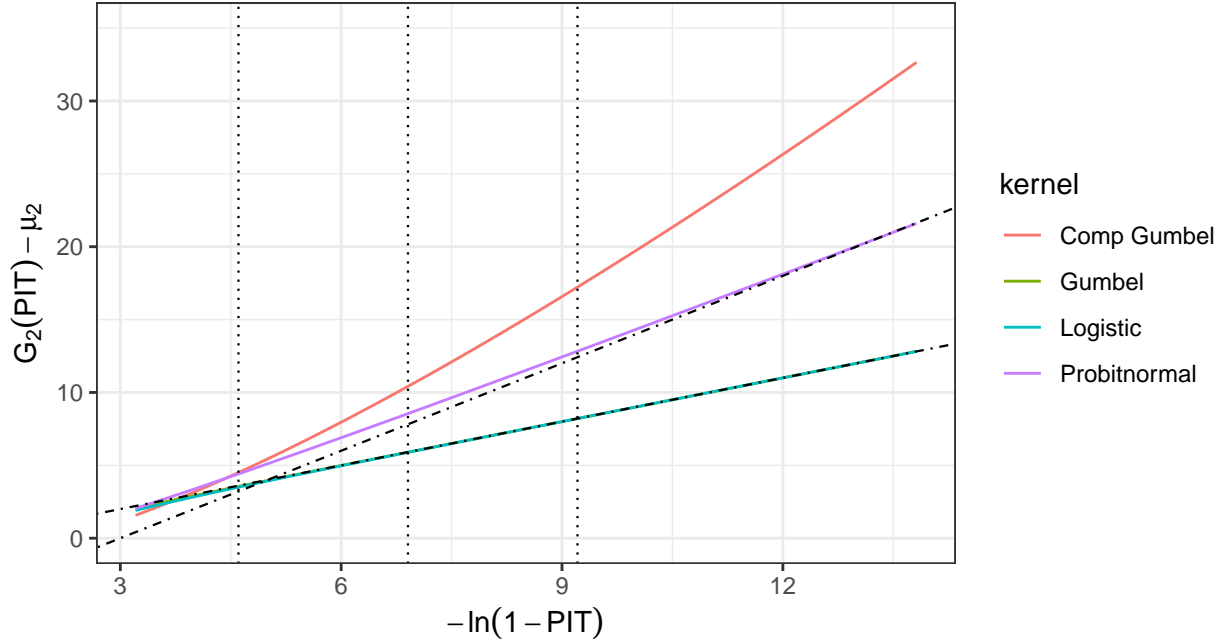
Figure 3: Tail behavior of the $G_2$ kernel.
The Gumbel and logistic lines are visually indistinguishable. The asymptote lines are $y = -1 + x$ and $y = -6 + 2x$.

is itself a logistic-beta distribution with parameters $(b, a)$. By varying parameters $(a, b)$ one obtains a rich family of distributions which vary materially in their shapes and higher moments. Nonetheless, all members of the family share the same tail behavior for the second kernel:

**Proposition 3.11.** *For all parameters $a > 0$, $b > 0$ of the logistic-beta family, $G_2(u) \sim -\ln(1 - u)$ as $u \to 1$.*

Figure 3 summarizes the tail behaviors of the examples described above. We do not include the logistic-beta because it is virtually indistinguishable from the logistic and Gumbel cases.

We assess size and power of TLSF tests following the methodology of Section 3.1. We include two logistic-beta distributions, one left-skewed and one right-skewed, in addition to the logistic. As a benchmark, we also include the Berkowitz (2001) LR-test. Results are reported in Table 3. Comparing the first column to the last, we see that the probitnormal TLSF test is

notably more powerful than its LR analog, the Berkowitz test. The complimentary Gumbel is the most powerful of the TLSF tests. It is comparable in size and power to the bivariate beta kernel tests with parameters $((2,0),(1,3))$ as reported in Table 2. The most striking feature of the table is that the Gumbel, logistic, and both logistic-beta TLSF tests deliver virtually identical size and power against all three alternatives. This is a consequence of the identical tail behavior of their respective $G_2(p)$ kernels as $p \to 1$. The probitnormal, which outperforms the Gumbel and logistic somewhat, has a steeper asymptote in Figure 3, while the complimentary Gumbel is steepest of all and convex.

| F | Probitnormal | Compli Gumbel | Gumbel | Logistic-Beta Logistic | $(3/2,1/2)$ | $(1/3,2/3)$ | Berkowitz |
|---|---|---|---|---|---|---|---|
| Normal | 5.1% | 5.5% | 5.1% | 5.1% | 5.0% | 5.1% | 5.1% |
| Scaled $t_{10}$ | 38.9% | 41.1% | 36.8% | 36.8% | 36.8% | 36.8% | 28.9% |
| Scaled $t_5$ | 72.9% | 74.5% | 71.1% | 71.2% | 71.2% | 71.1% | 65.5% |
| Scaled $t_3$ | 88.0% | 88.8% | 87.0% | 87.0% | 87.0% | 87.0% | 86.5% |

Table 3: Size and power of tests based on TLSF bikernels.
Kernel window is [0.975, 1]. 2^16 trials with 500 observations per trial. "Berkowitz" denotes the Berkowitz (2001) LR-test.

# 4 A family of v-shaped u.d.p. transformations

While the spectral tests we have developed will work with any u.d.p. transformation, we will confine our practical examples to transformations $T(v)$ that are *v-transforms*. V-transforms constitute a flexible parametric class of u.d.p. transformations that are well-suited to modeling volatile financial time series. Such transforms map values near a central *fulcrum* point to near zero and values near the boundaries to near one, so are useful in situations where we wish to emphasize PIT values in either tail. The symmetric linear v-transform $T(v) = |1-2v|$ is the simplest case and an obvious choice when we have no reason to suspect asymmetry in the true model.

Following McNeil (2021), we define

**Definition 4.1.** A v-transform is a mapping $T : I \to I$ with the following properties:

1. $T(0) = T(1) = 1$ and there exists a fulcrum point $0 < \delta < 1$ such that $T(\delta) = 0$;

2. $T$ is continuous on $I$, strictly decreasing on $[0, \delta]$ and strictly increasing on $[\delta, 1]$;

3. For every point $u \in I$ there is a point $l(u) \in [0, \delta]$ satisfying $T(l(u)) = u$ and $T(l(u) + u) = u$.

It is straightforward to verify that such a transformation preserves uniformity: if $u \in I$ and $P_t \sim U(0, 1)$ then $\mathbb{P}(T(P_t) \leqslant u) = \mathbb{P}(l(u) \leqslant P_t \leqslant l(u) + u) = u$ so that $T(P_t) \sim U(0, 1)$.

V-transforms may be characterized as mappings $T : I \to I$ taking the form

$$T(v) = \begin{cases} (1 - v) - (1 - \delta)\Psi\left(\frac{v}{\delta}\right) & v \leqslant \delta, \\ v - \delta\Psi^{-1}\left(\frac{1-v}{1-\delta}\right) & v > \delta, \end{cases} \tag{18}$$

where $0 < \delta < 1$ and $\Psi$ is a continuous and strictly increasing distribution function on $I$ (McNeil, 2021, Theorem 1). Figure 4 shows a number of examples of linear and nonlinear v-transforms constructed using (18) and the function $\Psi(v) = v^\kappa$, which we refer to as a *generator* of a v-transform. The symmetric linear v-transform corresponds to $\delta = 1/2$ and $\kappa = 1$. For small $\epsilon$ ($\epsilon < \min\{\delta, 1 - \delta\}$) in the linear case, it is straightforward to show that $T(\epsilon) - T(1 - \epsilon)$ is increasing in $\delta$. When $\kappa > 1$, the right arm of the v-transform is convex and the left arm concave. For small $\epsilon$, we can show that $T(\epsilon) - T(1 - \epsilon)$ is increasing in $\kappa$. That is, moving the fulcrum to the right or increasing convexity in $\Psi(v)$ increases the emphasis on PIT observations in the left tail relative to those in the right tail.

We will sometimes refer to v-transforms satisfying Definition 4.1 as *proper* v-transforms and describe the identity transformation $T(v) = v$ and the u.d.p. transformation $T(v) = 1 - v$ as *degenerate* v-transforms; the latter are degenerate because they satisfy $\lim_{\delta \to 0} T(v; \delta) = v$ for $v \in (0, 1]$ and $\lim_{\delta \to 1} T(v; \delta) = 1 - v$ for $v \in [0, 1)$, for any family of proper v-transforms $T(\cdot; \delta)$ indexed by $\delta$. It is straightforward to see that all proper v-transforms are *folding* in the sense of Definition 2.4 since, in all cases, $I_T = I \setminus \{\delta\}$.
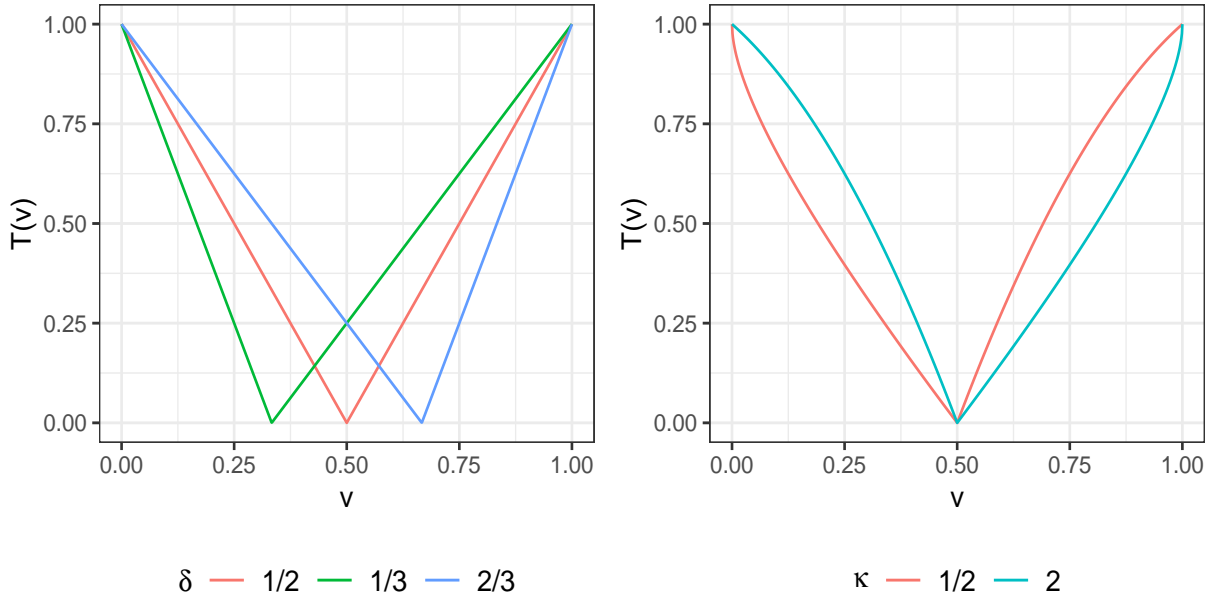
23

Figure 4: Examples of v-transforms with generator $\Psi(v) = v^\kappa$.
Linear v-transforms are shown in the left panel, for which we fix $\kappa = 1$ and vary the fulcrum $\delta$. Nonlinear v-transforms are shown in the right panel, for which we fix $\delta = 1/2$ and vary the exponent $\kappa$.

Fix a kernel $\nu_1$ on the unit interval, and let $\nu_\eta$ be the same kernel scaled to the window of width $\eta$ at the upper end of the unit interval, i.e., the window $[1 - \eta, 1]$. Similar to the scaling of the beta kernel in Section 3.1, we can write

$$G_\eta(u) = G_1 \left( \frac{(1 - \eta \vee u) - (1 - \eta)}{\eta} \right).$$

If the PIT data $\{P_t\}$ are pre-processed by the v-transform $T(v) = |1 - 2v|$, then transformed values $\{W_t\}$ satisfy $W_t > 0$ if and only if $P_t \in [0, \eta/2] \cup [1 - \eta/2, 1]$. In particular, PIT observations in $[1 - \eta, 1 - \eta/2]$ no longer receive weight. Thus, a possibly unintended consequence of the pre-transformation is that it concentrates weight on more extreme portions of the tails. To remedy this situation, we can double the kernel width, i.e., we replace $G_\eta(u)$ by $G_{2\eta}(u)$ so that $W_t > 0$ if and only if $P_t \in [0, \eta] \cup [1 - \eta, 1]$.

We illustrate with a simple exercise. We fix the kernel to the beta bikernel with parameters pairs $(1, 0)$ and $(1, 2)$. As in Section 3, we consider four different choices for the df $F$ of

the true model of $L_t$: the standard normal, and the scaled $t_{10}$, scaled $t_5$ and scaled $t_3$, and we assume that the forecaster's model is $\widehat{F} = \Phi$. In Table 4, the column labeled "identity" reports rejection rates in the absence of pre-processing (equivalently, by application of the degenerate v-transform $T(v) = v$) whereas the final column reports rejection rates when pre-processing by the proper v-transform $T(v) = |1 - 2v|$. The narrow window has kernel width $\eta = 0.025$ whereas the wide window has kernel width $2\eta = 0.05$.

| F | identity | \|1-2u\| |
|---|---|---|
| **Kernel window: Narrow** | | |
| Normal | 5.3% | 5.5% |
| Scaled $t_{10}$ | 40.8% | 60.9% |
| Scaled $t_5$ | 74.1% | 92.1% |
| Scaled $t_3$ | 88.1% | 97.9% |
| **Kernel window: Wide** | | |
| Normal | 5.0% | 5.1% |
| Scaled $t_{10}$ | 38.6% | 58.8% |
| Scaled $t_5$ | 75.4% | 92.2% |
| Scaled $t_3$ | 93.9% | 98.7% |

Table 4: Size and power of v-transformed tests under excess kurtosis. Beta bikernel with parameters ((1,0),(1,2)). 2^16 trials with 500 observations per trial.

Comparing the upper left quadrant to the lower left and the upper right quadrant to the lower right, we see that power can increase or decrease when the kernel width is doubled. As discussed in Section 3.1, tests are most powerful when the kernel window coincides with the range over which the true distribution differs most strongly from the forecaster's model. Doubling the kernel width can strengthen or dilute that coincidence. Comparing the upper left quadrant to the lower right, we see that the combination of pre-processing with adjustment of the kernel window substantially increases the power of the backtest. Because in the comparison both sets of tests weight identically on the upper 0.025 percent of the distribution of PIT values (under the null), the improved performance is coming from the addition of the lower 0.025 of the distribution of PIT values in the tests of the lower right quadrant.

When the true model exhibits skewness as well as excess kurtosis, in some situations one can obtain a more powerful backtest by pre-processing with an asymmetric v-transform. However, when the model is close to mesokurtic, the pre-processor must be well-chosen. To illustrate, consider the family of Fernández and Steel (1998) (hereafter "FS") skew-t distributions. We use a standardized version of this family with density

$$f_{FS}(x; \gamma, \zeta) = \frac{2\gamma}{\sigma(1 + \gamma^2)} \begin{cases} f_t\left(\frac{\gamma(x-\mu)}{\sigma}; \zeta\right) & x \leqslant \mu \\ f_t\left(\frac{x-\mu}{\gamma\sigma}; \zeta\right) & x > \mu \end{cases} \tag{19}$$

where $f_t(\cdot; \zeta)$ denotes the density of the standard t-distribution with $\zeta > 2$ degrees of freedom, $\gamma$ is a skewness parameter, and location $\mu = \mu(\zeta, \gamma)$ and scale $\sigma = \sigma(\zeta, \gamma)$ are chosen so that the FS distribution has mean 0 and variance 1; see the Online Supplement (Appendix 1) on this standardization. In the case of $\zeta = \infty$, we have the FS-Normal($\gamma$) distribution. In the case of $\gamma = 1$, we have the scaled $t_\zeta$ distribution as used earlier in Table 4 and in Section 3.

Figure 5 illustrates the PIT signature for unmodelled skewness. In the case of the FS-Normal distribution (left panel), for $\gamma = 3/2$ (blue line) the PIT density features a local maximum in $(0, 0.5)$ and a local minimum in $(0.5, 1)$. A symmetric pattern arises when we invert $\gamma$ ($\gamma = 2/3$, red line). Numerically, we can show that skewness of the PIT distribution takes the same sign as $\gamma - 1$ (so right skew for $\gamma = 3/2$, left skew for $\gamma = 2/3$) and increases in magnitude with $|\gamma - 1|$. Kurtosis of the PIT distribution is invariant to $\gamma$ (kurtosis of $9/5$, equal to that of the uniform distribution).

When we increase kurtosis in the true model ($\zeta = 5$, right panel), the departure from uniformity in the PIT density is magnified but the visual pattern is qualitatively similar. Numerically, we find that skewness is decreasing in magnitude as $\zeta \to \infty$ and as $\gamma \to 1$. Kurtosis is decreasing in $\zeta$ and increasing in $|\gamma - 1|$.

The exercise of Table 5 is similar to that of Table 4, except that we vary the skewness of the true model rather than its kurtosis. The rejection rates in the top row, labeled "Normal"
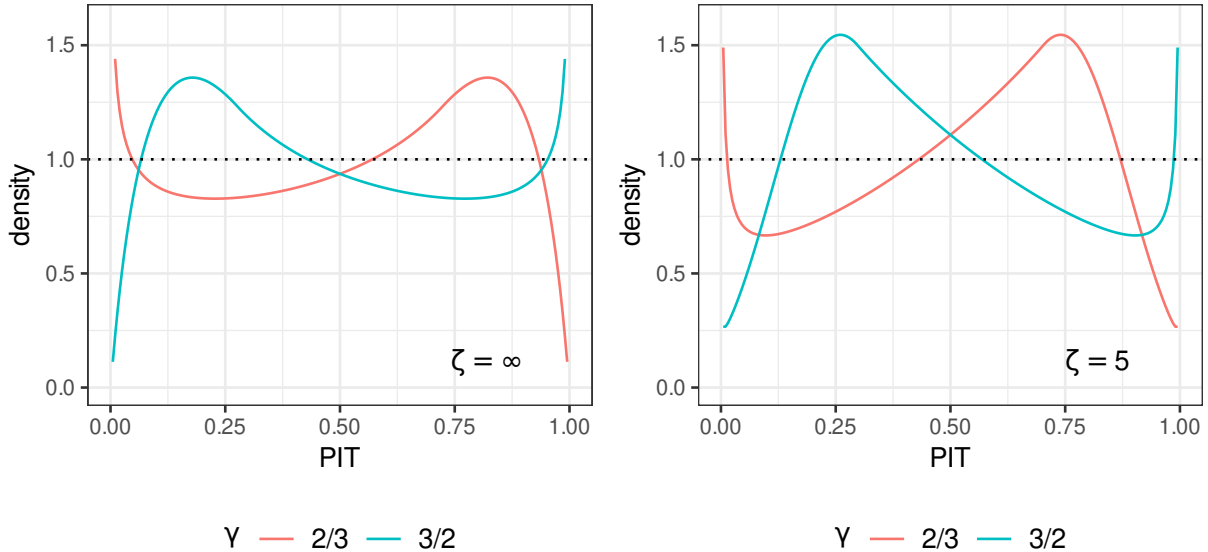
Figure 5: Density for PIT under model misspecification.
The true model is FS-t with $\zeta$ degrees of freedom and skewness parameter $\gamma$. The dotted line is the uniform density.

(equivalent to FS-Normal(1)), capture the size of the test. Parameter $\gamma$ increases as we move down the rows of the table, which increases right skew. Comparing the first two columns shows that pre-processing with $T(v) = |1 - 2v|$ decreases power, regardless of whether we double the width of the kernel window. Under the FS-Normal($\gamma$) model, left-tail PIT values are underrepresented and right-tail PIT values overrepresented. Folding the PIT distribution averages over these effects, thereby obscuring the departure from uniformity.

The remaining columns of Table 5 vary parameters $(\delta, \kappa)$ of the pre-processor. For the right-skewed true models, we observe that an asymmetric pre-processor can materially increase the power of the backtest relative to a symmetric pre-processor but all pre-processors reduce power relative to the case of no pre-processor for the reasons just discussed. Rejection rates are higher for $\delta = 1/3$ than for $\delta = 2/3$ and higher for $\kappa = 1/2$ than for $\kappa = 2$. The intuition is most straightforward for the linear case: as we reduce $\delta$, the pre-processor $T(v; \delta)$ converges towards the identity pre-processor (equivalent to $\delta = 0$).

Finally, in Table 6, we report rejection rates for a true model with considerable excess kurtosis as well as skewness. The results inherit some features from each of the previous

27

| F | identity | \|1-2u\| | ($^1/_3$, 1) | ($^2/_3$, 1) | ($^1/_2$, $^1/_2$) | ($^1/_2$, 2) |
|---|---|---|---|---|---|---|
| Kernel window: Narrow | | | | | | |
| Normal | 5.3% | 5.5% | 5.4% | 5.4% | 5.3% | 5.2% |
| FS-Normal(51/50) | 7.0% | 5.6% | 5.9% | 5.1% | 6.9% | 4.2% |
| FS-Normal(26/25) | 8.8% | 5.7% | 6.5% | 4.8% | 8.7% | 3.4% |
| FS-Normal(6/5) | 31.9% | 10.2% | 15.3% | 6.4% | 30.5% | 3.7% |
| FS-Normal(4/3) | 52.2% | 17.9% | 27.1% | 11.5% | 50.3% | 11.9% |
| Kernel window: Wide | | | | | | |
| Normal | 5.0% | 5.1% | 5.1% | 5.2% | 4.9% | 5.1% |
| FS-Normal(51/50) | 6.4% | 5.3% | 5.6% | 4.9% | 6.4% | 4.3% |
| FS-Normal(26/25) | 8.2% | 5.4% | 6.0% | 4.8% | 7.8% | 3.6% |
| FS-Normal(6/5) | 30.9% | 9.0% | 13.7% | 5.8% | 28.7% | 5.4% |
| FS-Normal(4/3) | 51.1% | 16.4% | 25.1% | 11.2% | 48.5% | 16.3% |

Table 5: Size and power of v-transformed tests under skewness with minimal excess kurtosis. Column headers of form $(\delta, \kappa)$ refer to the v-transform with fulcrum $\delta$ and generator $\Psi(u) = u^\kappa$. The *identity* pre-processor is equivalent to $(\delta = 0, \kappa = 1)$ and $|1 - 2u|$ equivalent to $(\delta = ^1/_2, \kappa = 1)$. All tests utilize the beta bikernel with parameters $((1,0), (1,2))$. $2^{16}$ trials with 500 observations per trial.

two tables. As in Table 4, the symmetric linear pre-processor materially increases the power of the backtest when $\gamma$ is close to 1. As skewness increases, the considerations at work in Table 5 begin to dominate. As in Table 5, we can further increase the power in some cases by nudging the fulcrum of the pre-processor to the left. For larger values of $\gamma$, a nonlinear pre-processor ($\kappa = ^1/_2$) is even more effective.

In Section 3.1 and in Gordy and McNeil (2020), we highlight the intuition that a kernel is most powerful against a given alternative when it places greater mass on regions of the unit interval over which the forecaster's PIT distribution departs most from uniformity. The intuition extends to the choice of pre-processor. In Figure 6, we plot the df of the linear $T(P; \delta)$ for $P \sim$ FS-t$(5, 26/25)$ and various values of $\delta$. Consistent with the findings in Table 6, we see that the df of the PITs crosses the uniform df within the wide kernel window of $[0.95, 1]$. The dfs of the pre-processed PIT values are pointwise more distant from the uniform df except at the left edge of the window. Departure from uniformity is slightly decreasing across the three values of $\delta$, which accords with the slight differences in power of

| F | identity | \|1-2u\| | ($^1/_3$, 1) | ($^2/_3$, 1) | ($^1/_2$, $^1/_2$) | ($^1/_2$, 2) |
|---|---|---|---|---|---|---|
| Kernel window: Narrow | | | | | | |
| Normal | 5.3% | 5.5% | 5.4% | 5.4% | 5.3% | 5.2% |
| Scaled $t_5$ | 74.1% | 92.1% | 91.7% | 91.7% | 81.8% | 81.8% |
| FS-t(5,51/50) | 77.2% | 92.2% | 92.0% | 91.7% | 83.7% | 80.0% |
| FS-t(5,26/25) | 79.7% | 92.2% | 92.2% | 91.6% | 85.3% | 78.1% |
| FS-t(5,6/5) | 92.1% | 93.1% | 94.1% | 91.4% | 93.5% | 63.9% |
| FS-t(5,4/3) | 96.1% | 94.3% | 95.5% | 92.4% | 96.5% | 58.0% |
| Kernel window: Wide | | | | | | |
| Normal | 5.0% | 5.1% | 5.1% | 5.2% | 4.9% | 5.1% |
| Scaled $t_5$ | 75.4% | 92.2% | 91.8% | 91.9% | 81.4% | 81.3% |
| FS-t(5,51/50) | 78.2% | 92.3% | 92.0% | 91.8% | 83.2% | 79.7% |
| FS-t(5,26/25) | 80.6% | 92.3% | 92.3% | 91.7% | 84.9% | 77.8% |
| FS-t(5,6/5) | 92.1% | 93.4% | 94.3% | 91.9% | 93.5% | 64.7% |
| FS-t(5,4/3) | 96.1% | 94.9% | 95.9% | 93.4% | 96.5% | 61.6% |

Table 6: Size and power of v-transformed tests under skewness and excess kurtosis.
Column headers of form $(\delta, \kappa)$ refer to the v-transform with fulcrum $\delta$ and generator $\Psi(u) = u^\kappa$. The *identity* pre-processor is equivalent to $(\delta = 0, \kappa = 1)$ and $|1 - 2u|$ equivalent to $(\delta = ^1/_2, \kappa = 1)$. The Scaled $t_5$ alternative is equivalent to FS-t(5, 1). All tests utilize the beta bikernel with parameters $((1,0), (1,2))$. 2^16 trials with 500 observations per trial.

the corresponding tests.[4]

---
[4]The line for $\delta = ^1/_2$, not shown in the plot, lies between those of $\delta = ^1/_3$ and $\delta = ^2/_3$, and is difficult to distinguishable from the former.
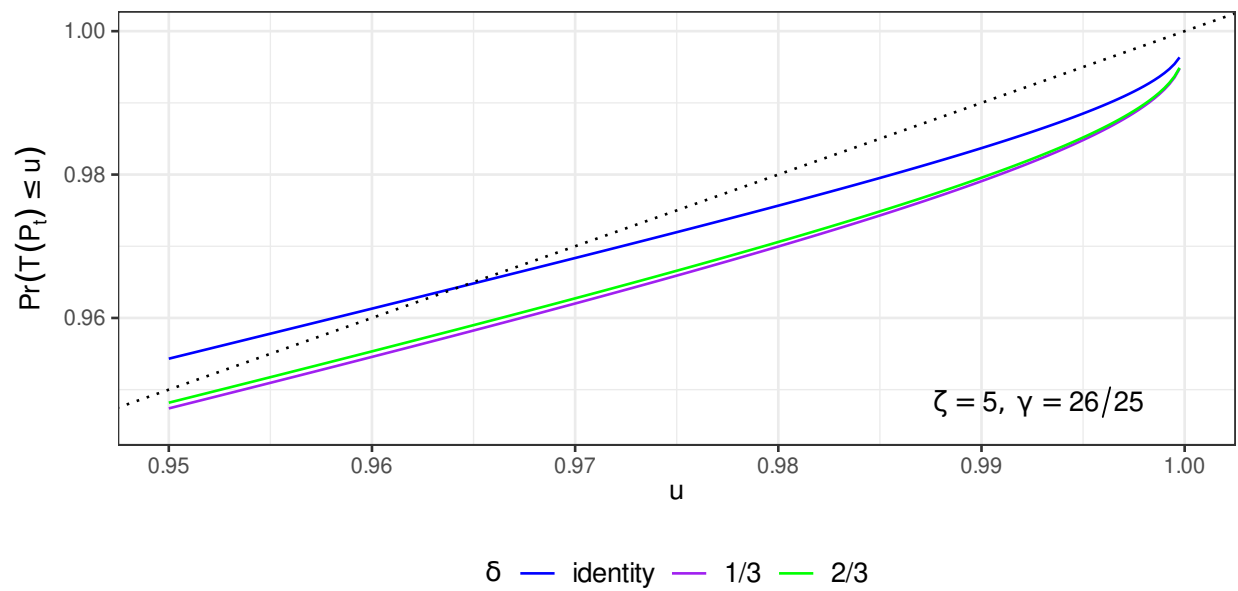
Figure 6: Tails of distribution functions for v-transformed PIT-values.
Dfs for v-transformed PIT-values when the forecaster assumes standard normal losses ($\widehat{F} = \Phi$) but the true loss model $F$ is FS-t(5,26/25) and the validator applies a linear pre-processor.

# 5  Conclusion

We conclude with some practical guidance on the choice of kernel and pre-processor. As in Gordy and McNeil (2020), we emphasize that such choices express implicitly the preferences of the validator (i.e., the locus of departure from PIT uniformity that the validator would deem most worrisome), so it is not our place to be prescriptive. To the extent that the validator may be uncertain or agnostic over her preferences, other considerations could come into play. In particular, if the validator has some prior familiarity with the forecaster's methodology, the validator might wish to craft the test to highlight suspected flaws. Power and size are natural considerations as well, and the validator might also prefer that the test statistic have good numerical properties, e.g., simple to program, quick to calculate, and robust to small changes in parameters and data values.

The unbounded beta and TLSF families are similar in the power that can be achieved against alternatives featuring excess kurtosis, and substantially more powerful than tests based on bounded kernels. Our evidence indicates that all cases of these families yield well-sized tests. Where they differ is in their computational and aesthetic properties. Tests based on the beta kernel are easy to program and fast but depend in general on the availability of routines for the hypergeometric functions. Under special cases, in particular when parameters take integer and half-integer values, the hypergeometric functions can be sidestepped, resulting in simple expressions. With regard to the unbounded beta kernel, the case of $b = 0$ is easily programmed and numerically stable, but for non-zero values of $b$ near zero, numerical instability can arise.

The TLSF kernels arise naturally as score tests, so can be seen as moment-based analogues to well-known LR tests such as that of Berkowitz (2001). Relative to the beta family, tests based on the TLSF families are more intricate to program and somewhat slower to execute, though still much faster than their LR-test analogs and generally more powerful on samples of typical length. These tests appear to be numerically stable. The TLSF family of tests are valid on the upper tail of the PIT distribution but not over the entire unit interval. In some

applications, the restriction on the choice of kernel window could be a limitation.

When the validator suspects unmodeled kurtosis in the forecast model, as might arise when a model adapts inadequately to changes in market volatility, pre-processing the PIT data with a folding transformation is highly effective in highlighting an excess of tail observations. The simple symmetric linear v-transform $T(v) = |1 - 2v|$ performs well. When the validator additionally suspects unmodeled skewness, an asymmetric v-transform may offer somewhat greater power than the symmetric linear v-transform but at some cost to robustness. All of the pre-processors considered in our analysis can be implemented in a few lines of trivial code. Furthermore, because the test statistic is a sample mean of a simple composition of the pre-processor and the kernel df, the algorithm can be implemented in modular fashion. That is, starting with a sample of PIT values $\{P_i\}$, we first apply the pre-processor as $\tilde{P}_i = T(P_i)$, and then feed the sample $\{\tilde{P}_i\}$ through the spectral backtest. There is no alteration to the code for the spectral backtest.

# A  Proofs

## A.1  Proof of Theorem 2.1

For notation simplicity assume $T$ is the identity transformation. Since $G_1$ and $G_2$ are increasing, right-continuous functions, it follows that the function $G^*(u) = G_1(u)G_2(u)$ must also be increasing and right-continuous and thus it can be used to define a Lebesgue-Stieltjes measure $\nu^*$ by setting $\nu^*(\{0\}) = G^*(0) = 0$ and $\nu^*((a,b]) = G^*(b) - G^*(a)$ for any $0 \leqslant a < b \leqslant 1$. It follows that $W_t^* = G^*(P_t) = \nu^*([0, P_t])$.

The formula for $G^*$ is obtained by applying the integration-by-parts formula for the Lebesgue-Stieltjes integral (Hewitt, 1960, Theorem A); see also Refuz and Yor (2004, Ch. 0).

## A.2 Proof of Proposition 2.2

Since $G_\nu(u) = \mathcal{O}((1-u)^{-1/2+\epsilon})$ as $u \to 1$ for some small $\epsilon$, there exists a value $u_0$ and a positive constant $C$ such that $G_\nu(u) \leqslant C(1-u)^{-1/2+\epsilon}$ for $u \geqslant u_0$. Let $\bar{u}$ be the larger of $u_0$ and the last point at which $G_\nu$ is not differentiable (there are only finitely so many points by Assumption 1). We can decompose (8) as

$$\mathbb{E}(W_t^2) = \int_{[0,\bar{u}]} (1-u)\,(2G_\nu(u) - \nu(\{u\}))\,\mathrm{d}\nu(u) + \int_{(\bar{u},1]} (1-u)\,(2G_\nu(u) - \nu(\{u\}))\,\mathrm{d}\nu(u) \quad (A.1)$$

The integrand in the first term is bounded above by $2G_\nu(\bar{u})$ and so the integral is finite. We only need to prove the finiteness of the second term which can be written as

$$\int_{\bar{u}}^1 (1-u)2G_\nu(u)g_\nu(u)\mathrm{d}u = \int_{\bar{u}}^1 (1-u)\frac{\mathrm{d}}{\mathrm{d}u}\left(G_\nu(u)^2\right)\mathrm{d}u = \left[G_\nu(u)^2(1-u)\right]_{\bar{u}}^1 + \int_{\bar{u}}^1 G_\nu(u)^2\mathrm{d}u$$

using integration by parts.

Since $0 \leqslant G_\nu(u)^2(1-u) \leqslant C^2(1-u)^{2\epsilon}$ for $u \geqslant \bar{u}$ and $(1-u)^{2\epsilon} \to 0$ as $u \to 1$, it follows that $[G_\nu(u)^2(1-u)]_{\bar{u}}^1 = -G_\nu(\bar{u})^2(1-\bar{u})$. Moreover, the second term is finite because

$$\int_{\bar{u}}^1 G_\nu(u)^2\mathrm{d}u \leqslant C^2 \int_{\bar{u}}^1 (1-u)^{-1+2\epsilon}\mathrm{d}u = \frac{C^2}{2\epsilon}(1-\bar{u})^{2\epsilon}\,.$$

## A.3 Proof of Lemma 2.3

Assume that $\nu$ is a bounded measure with moments $\mu$ and $\sigma$. Without loss of generality $\nu$ may be taken to be a probability measure with df $G_\nu$ satisfying $G_\nu(1) = 1$. In that case let $\tilde{\nu}$ to be the probability measure defined by the df $G_{\tilde{\nu}}(u) = 1 - G_\nu(1-u)$. Observe that if $U$ has df $G_\nu$ then $1 - U$ has df $G_{\tilde{\nu}}$ and hence moment formulas for linear functions of random variables give $\tilde{\mu} = 1 - \mu$ and $\tilde{\sigma} = \sigma$. We obtain the identity

$$\frac{|\widetilde{W}_n - \tilde{\mu}|}{\tilde{\sigma}} = \frac{|(1 - \overline{W}_n) - (1 - \mu)|}{\sigma} = \frac{|\overline{W}_n - \mu|}{\sigma},$$

showing redundancy of $T$.

Now assume the measure $\nu$ is unbounded with finite moments $\mu$ and $\sigma$ and consider PIT samples consisting of the single point $\{v\}$. If the transform $T(v) = 1 - v$ is redundant we must have

$$\frac{|G_{\tilde{\nu}}(v) - \tilde{\mu}|}{\tilde{\sigma}} = \frac{|G_{\nu}(1 - v) - \mu|}{\sigma}$$

for all $v \in [0, 1]$, a measure $\tilde{\nu}$ and finite values $\tilde{\mu}$ and $\tilde{\sigma}$. But the rhs tends to infinity as $v \to 0$ while the lhs tends to the finite limit $\tilde{\mu}/\tilde{\sigma}$ which yields a contradiction.

## A.4  Proof of Proposition 2.5

Suppose that the folding u.d.p. transformation $T$ is redundant. For any value $u \in I_T$, let $\underline{v}_u$ and $\bar{v}_u$ be the smallest and largest elements of $T^{-1}[\{u\}]$. Consideration of the PIT samples $\{\underline{v}_u\}$ and $\{\bar{v}_u\}$ implies that the following identities must hold.

$$\frac{|G_{\tilde{\nu}}(\underline{v}_u) - \tilde{\mu}|}{\tilde{\sigma}} = \frac{|G_{\nu}(u) - \mu|}{\sigma} = \frac{|G_{\tilde{\nu}}(\bar{v}_u) - \tilde{\mu}|}{\tilde{\sigma}} \tag{A.2}$$

Monotonicity of $G_{\tilde{\nu}}$ implies that $G_{\tilde{\nu}}(\underline{v}_u) \leqslant G_{\tilde{\nu}}(\bar{v}_u)$. Now consider two cases.

**Case (a): For some $u$ with $G_{\nu}(u) \neq \mu$, $G_{\tilde{\nu}}(\underline{v}_u) < G_{\tilde{\nu}}(\bar{v}_u)$.**  Equation (A.2) can hold only if $\tilde{\mu} - G_{\tilde{\nu}}(\underline{v}_u) = G_{\tilde{\nu}}(\bar{v}_u) - \tilde{\mu}$, implying $\tilde{\mu} = (G_{\tilde{\nu}}(\underline{v}_u) + G_{\tilde{\nu}}(\bar{v}_u))/2$. Now consider the PIT sample of length two, $\{\underline{v}_u, \bar{v}_u\}$. We can easily verify that $\widetilde{W}_2 = \tilde{\mu}$ while $\overline{W}_2 = G_{\nu}(u) \neq \mu$. Thus, the $Z_n$ statistic must be zero for $\tilde{\nu}$ but non-zero in magnitude for the u.d.p.-transformed $\nu$, which contradicts the supposition that $T$ is redundant for $\nu$.

**Case (b): For all $u$ with $G_{\nu}(u) \neq \mu$, $G_{\tilde{\nu}}(\underline{v}_u) = G_{\tilde{\nu}}(\bar{v}_u)$.**  Observe first that (A.2) implies that $G_{\tilde{\nu}}(\underline{v}_u) = G_{\tilde{\nu}}(\bar{v}_u) = \tilde{\mu}$ whenever $G_{\nu}(u) = \mu$, so in case (b) we must have that $G_{\tilde{\nu}}(\underline{v}_u) = G_{\tilde{\nu}}(\bar{v}_u)$ for all $u \in I_T$. Because $G_{\tilde{\nu}}(v)$ is nondecreasing for all $v$, this implies that each value of $u \in I_T$ maps to a interval $[\underline{v}_u, \bar{v}_u] \subseteq I$ over which $G_{\tilde{\nu}}(v)$ is constant. These intervals do not

overlap, so there can be only countably many. However, because $G_\nu(u)$ is strictly increasing over some interval within $I$, we need uncountably many values in $G_{\tilde{\nu}}[I]$ to satisfy (A.2) at each $u \in I_T$, thus leading to a contradiction.

## A.5    Proof of Proposition 3.2

We maintain the assumption that $a > 0$. Wolfram Research (2023, 06.19.06.0049.01) provides these expansions as $u \to 1$:

$$B(u, a, b) \propto \begin{cases} -\ln(1 - u) - \psi(a) - \gamma & b = 0, \\[2mm] \frac{(-1)^{b-1}\Gamma(a)\ln(1-u)}{(-b)!\Gamma(a+b)} - \frac{(1-u)^b}{b} & -b \in \mathbb{N}, \\[2mm] B(a, b) - \frac{(1-u)^b}{b} & \text{otherwise.} \end{cases} \qquad (A.3)$$

where $\psi$ is the digamma function and $\gamma$ is the Euler-Mascheroni constant. In the middle case, note that $(1 - u)^b$ dominates $\ln(1 - u)$ as $u \to 1$ and further that in the edge case of $-(a + b) \in \mathbb{N}$, $\Gamma(a + b)$ is infinite so the first term simply drops out. In the final case, note that $B(a, b)$ is finite even when $b < 0$. The proposition follows directly.

## A.6    Proof of Theorem 3.3

The likelihood function is given by

$$\mathcal{L}_{P^*}(\boldsymbol{\theta} \mid \boldsymbol{P}) = \prod_{t\,:\,P_t < \alpha_1} F_P(\alpha_1 \mid \boldsymbol{\theta}) \prod_{t\,:\,\alpha_1 \leqslant P_t < \alpha_2} f_P(P_t \mid \boldsymbol{\theta}) \prod_{t\,:\,P_t \geqslant \alpha_2} \bar{F}_P(\alpha_2 \mid \boldsymbol{\theta}) \qquad (A.4)$$

where $\bar{F}_P(u)$ denotes the tail probability $1 - F_P(u)$. The likelihood contributions $\mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t)$ are given by the individual terms in (A.4) according to whether $P_t < \alpha_1$, $\alpha_1 \leqslant P_t < \alpha_2$ or $P_t \geqslant \alpha_2$. With the help of the calculations in Appendix D, and using the functions $C_1(x) = \rho(x)/R(x)$ and $\bar{C}_1(x) = -\rho(x)/\bar{R}(x)$ defined there, we can compute the score vector

and evaluate it at $\boldsymbol{\theta}_0 = (0,1)'$ to obtain

$$
\boldsymbol{S}_t(\boldsymbol{\theta}_0) = \begin{cases}
\boldsymbol{\psi}_1(R^{-1}(\alpha_1)) & P_t < \alpha_1, \\[2mm]
\boldsymbol{\psi}_*(R^{-1}(P_t)) & \alpha_1 \leqslant P_t < \alpha_2, \\[2mm]
\boldsymbol{\psi}_2(R^{-1}(\alpha_2)) & P_t \geqslant \alpha_2.
\end{cases} \tag{A.5}
$$

where

$$
\boldsymbol{\psi}_1(x) = \begin{pmatrix} -C_1(x) \\ -xC_1(x) \end{pmatrix}, \quad \boldsymbol{\psi}_*(x) = \begin{pmatrix} \lambda_\rho(x) \\ x\lambda_\rho(x) - 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\psi}_2(x) = \begin{pmatrix} -\bar{C}_1(x) \\ -x\bar{C}_1(x) \end{pmatrix}.
$$

The third case in (A.5), described by the function $\boldsymbol{\psi}_2$, only comes into play when $\alpha_2 < 1$.

If $\rho$ is log-concave this is equivalent to saying that $\lambda_\rho(x) = -\rho'(x)/\rho(x)$ is an increasing function. We first prove that the equation (14) has a unique root $\underline{x}$ and that $\underline{x} > 0$. Let $\Lambda(x) = x(C_1(x) + \lambda_\rho(x))$ and note that $\Lambda$ is a continuous function satisfying $\lim_{x\to 0} \Lambda(x) = 0$ and $\lim_{x\to\infty} \Lambda(x) = \infty$. Hence there exists at least one $x$ satisfying $\Lambda(x) = 1$ and $x > 0$. The derivative $\Lambda'(x)$ can be split into two parts yielding

$$
\begin{aligned}
\Lambda'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(xC_1(x)\right) + \frac{\mathrm{d}}{\mathrm{d}x}\left(x\lambda_\rho(x)\right) \\
&= C_1(x) - x\lambda_\rho(x)C_1(x) - xC_1(x)^2 + \frac{\mathrm{d}}{\mathrm{d}x}\left(x\lambda_\rho(x)\right) \\
&= C_1(x)\left(1 - \Lambda(x)\right) + \frac{\mathrm{d}}{\mathrm{d}x}\left(x\lambda_\rho(x)\right).
\end{aligned}
$$

At any $x$ satisfying $\Lambda(x) = 1$ the first term must be zero and the second term must be strictly positive, since it is the derivative of a strictly increasing function. Since the gradient of $\Lambda$ is positive at any root of the equation (14) we conclude that the latter has a unique root $\underline{x}$ and that $\underline{x} > 0$.

We now turn to the representation of the score test as a bispectral test. Since $\lambda_\rho$ is an increasing function it follows that both components of $\boldsymbol{\psi}_*(x)$ are also increasing functions

and thus non-negative weighting functions $g_i$ can be obtained by differentiating $\boldsymbol{\psi}_*(R^{-1}(u))$ with respect to $u$ on $[\alpha_1, \alpha_2]$.

The discontinuities at $\alpha_1$ and $\alpha_2$ are given by

$$(\gamma_{1,1}, \gamma_{2,1})' = \boldsymbol{\psi}_*(R^{-1}(\alpha_1)) - \boldsymbol{\psi}_1(R^{-1}(\alpha_1)), \qquad (\gamma_{1,2}, \gamma_{2,2})' = \boldsymbol{\psi}_2(R^{-1}(\alpha_2)) - \boldsymbol{\psi}_*(R^{-1}(\alpha_2))$$

where the $\gamma_{i,2}$ constants need only be considered when $\alpha_2 < 1$. Non-negativity of the $\gamma_{i,j}$ requires that the inequalities

$$\lambda_\rho(x) + C_1(x) \geqslant 0 \tag{A.6}$$

$$x\left(\lambda_\rho(x) + C_1(x)\right) - 1 \geqslant 0 \tag{A.7}$$

hold for $x = R^{-1}(\alpha_1)$ and the inequalities

$$-\left(\lambda_\rho(x) + \bar{C}_1(x)\right) \geqslant 0 \tag{A.8}$$

$$-x\left(\lambda_\rho(x) + \bar{C}_1(x)\right) + 1 \geqslant 0 \tag{A.9}$$

hold for $x = R^{-1}(\alpha_2)$ if $\alpha_2 < 1$.

Since $\rho'(x)/\rho(x)$ is a decreasing function we can infer that

$$\frac{\rho'(x)}{\rho(x)}R(x) = \frac{\rho'(x)}{\rho(x)}\int_{-\infty}^{x}\rho(t)\mathrm{d}t \leqslant \int_{-\infty}^{x}\frac{\rho'(t)}{\rho(t)}\rho(t)\mathrm{d}t = \rho(x),$$

implying that (A.6) holds for all $x \in \mathbb{R}$, and

$$\frac{\rho'(x)}{\rho(x)}\bar{R}(x) = \frac{\rho'(x)}{\rho(x)}\int_{x}^{\infty}\rho(t)\mathrm{d}t \geqslant \int_{x}^{\infty}\frac{\rho'(t)}{\rho(t)}\rho(t)\mathrm{d}t = -\rho(x),$$

implying that (A.8) holds for all $x \in \mathbb{R}$. Since we have assumed that $R^{-1}(\alpha_1) \geqslant R^{-1}(\underline{\alpha}) = \underline{x}$, it follows that $\Lambda(R^{-1}(\alpha_1)) \geqslant 1$ and hence that (A.7) holds for $x = R^{-1}(\alpha_1)$. Moreover, since (A.8) holds at $x = R^{-1}(\alpha_2)$ and $R^{-1}(\alpha_2) > R^{-1}(\alpha_1) \geqslant R^{-1}(\underline{\alpha}) > 0$, then (A.9) clearly

also holds for $x = R^{-1}(\alpha_2)$.

Finally, to determine $\boldsymbol{\mu_W} = \boldsymbol{W}_t - \boldsymbol{S}_t(\boldsymbol{\theta}_0)$, we note that, if $P_t < \alpha_1$, then (15) implies that $W_{t,i} = 0$ for $i = 1, 2$ while (A.5) implies that $\boldsymbol{S}_t(\boldsymbol{\theta}_0) = \boldsymbol{\psi}_1(R^{-1}(\alpha_1))$. It follows that we must have $\boldsymbol{\mu_W} = -\boldsymbol{\psi}_1(R^{-1}(\alpha_1))$.

## A.7 Proof of Proposition 3.11

Recall that asymptotic behavior of $G_2(p)$ as $p \to 1$ depends on that of $R^{-1}(p)\lambda_\rho(R^{-1}(p))$. The df is a composition of the beta df $\mathcal{I}(z; a, b)$ and the logistic function, i.e., $R(x) = \mathcal{I}(S(x); a, b)$. The inverse df is therefore $R^{-1}(p) = \text{logit}(\mathcal{I}^{-1}(p; a, b))$ where $\text{logit}(u) = \ln(u/(1 - u))$. The well-known symmetry for the beta distribution implies a symmetry of the same form for the inverse df:

$$1 - \mathcal{I}^{-1}(p; a, b) = \mathcal{I}^{-1}(1 - p; b, a)$$

Consequently, we can write

$$R^{-1}(p) = \ln\left(\mathcal{I}^{-1}(p; a, b)\right) - \ln\left(\mathcal{I}^{-1}(1 - p; b, a)\right). \tag{A.10}$$

From $\lambda_\rho(x) = bS(x) - aS(-x)$, it follows immediately that $\lim_{p \to 1} \lambda_\rho(R^{-1}(p)) = b$. Combining with (A.10), we can infer $R^{-1}(p)\lambda_\rho(R^{-1}(p)) \sim -b \ln\left(\mathcal{I}^{-1}(1 - p; b, a)\right)$ as $p \to 1$.

From (C.3), we have

$$\lim_{u \to 0} aB(a, b)u^{-a}\mathcal{I}(u; a, b) = 1$$

from which we may infer that

$$\ln\left(\mathcal{I}^{-1}(1 - p; b, a)\right) \sim \frac{1}{b}\left(\ln(1 - p) + \ln(bB(a, b))\right) \text{ as } p \to 1.$$

The constant additive term is negligible asymptotically so

$$R^{-1}(p)\lambda_\rho(R^{-1}(p)) \sim -\ln(1 - p), \quad p \to 1,$$

regardless of $(a, b)$.

# B   TLSF tests when the density is non-log-concave

The issues that arise for a non-log-concave density can be illustrated by considering the Student t distribution with $\nu$ degree of freedom and working through the steps of the proof in Section A.6. In this case $\lambda_\rho(x) = (\nu + 1)x/(\nu + x^2)$ is only an increasing function between two turning points at $x = \pm\sqrt{\nu}$ and so there are immediate constraints on the interval in which the densities (16) are positive; in particular, we cannot construct a bispectral test with $\alpha_2 = 1$. Moreover, we have to check all of the conditions (A.6) to (A.9) individually to find an interval $[\alpha_1, \alpha_2]$ on which the test may be applied. For example, when $\nu = 4$, we need to set $\alpha_1 \geqslant \underline{\alpha} \approx 0.773$ and $\alpha_2 \leqslant \overline{\alpha} \approx 0.887$ to obtain a proper bispectral test. The width of the interval increases for larger degrees of freedom but the extra constraints relative to log-concave densities render the t distribution less suitable for constructing bispectral tests.

# C   Moments for the beta kernel

We seek solutions to the moments and cross-moments of the transformed PIT values when the kernel densities take the form

$$g_\nu(u) = (\alpha_2 - \alpha_1)^{1-a-b}(u - \alpha_1)^{a-1}(\alpha_2 - u)^{b-1}\mathbb{1}_{\{\alpha_1 \leqslant u \leqslant \alpha_2\}}$$

for parameters $(a > 0, b > -1/2)$ and $0 \leqslant \alpha_1 < \alpha_2 \leqslant 1$.

Coding is facilitated by computing moments in terms of the moments of standardized beta kernels with $\alpha_1 = 0, \alpha_2 = 1$. Let $\tilde{\nu}$ denote a beta$(a, b)$ kernel with kernel density $g_\nu(u) = u^{a-1}(1 - u)^{b-1}$ on $[0, 1]$ and let $W = G_\nu(U)$ and $\tilde{W} = G_{\tilde{\nu}}(U)$ for $U \sim \text{Uniform}(0, 1)$.

The uncentered moments of $W$ can be obtained as:

$$\mathbb{E}(W^k) = \begin{cases} (\alpha_2 - \alpha_1)\mathbb{E}(\tilde{W}^k) + (1 - \alpha_2)G_{\tilde{\nu}}(1)^k & \text{if } \alpha_2 < 1 \\ \\ (1 - \alpha_1)\mathbb{E}(\tilde{W}^k) & \text{if } \alpha_2 = 1. \end{cases} \tag{C.1}$$

Proposition 3.2 guarantees that $\lim_{y \to 1}(1 - y)G_{\tilde{\nu}}(y)^k = 0$ for $k = 1, 2$, so the expression in the $\alpha_2 = 1$ case is simply the limit of the expression in the $\alpha_2 < 1$ case.

The first moment of $\tilde{W}$ is

$$\mathbb{E}(\tilde{W}) = \int_0^1 (1 - u)g_{\tilde{\nu}}(u)\, du = B(a, 1 + b) \tag{C.2}$$

Since $b > -1/2$, this expression presents no difficulties for our application.

By Wolfram Research (2023, 06.19.26.0005.01, 06.19.26.0006.01), the kernel function $G_{\tilde{\nu}}(u) = B(u; a, b)$ can be expressed in terms of the Gauss hypergeometric function, $_2F_1$, in two equivalent forms:

$$B(u; a, b) = \frac{u^a}{a}{}_2F_1(a, 1 - b; a + 1; u) \tag{C.3}$$

$$= \frac{u^a(1 - u)^b}{a}{}_2F_1(1, a + b; a + 1; u). \tag{C.4}$$

Say we have two beta variables with (possibly) different parameters $(a_i, b_i)$ for $i = 1, 2$. Let $\tilde{G}_i$ denote the transform function for kernel $i$. To get cross-moments and second moments, we need the integral

$$M(a_1, b_1, a_2, b_2) = \int_0^1 (1 - u)\tilde{g}_1(u)\tilde{G}_2(u)du \tag{C.5}$$

$$= \frac{B(a_1 + a_2, 1 + b_1)}{a_2}{}_3F_2(a_2, a_1 + a_2, 1 - b_2; 1 + a_2, 1 + a_1 + a_2 + b_1; 1) \tag{C.6}$$

$$= \frac{B(a_1 + a_2, 1 + b_1 + b_2)}{a_2}{}_3F_2(1, a_1 + a_2, a_2 + b_2; 1 + a_2, 1 + a_1 + a_2 + b_1 + b_2; 1) \tag{C.7}$$

The two forms come from application of Gradshteyn and Ryzhik (2007, 7.512.5) to (C.3)

and (C.4), respectively. When calculated from its series expansion, equation (C.6) will be numerically stable whenever $b_2 \leqslant 0$ whereas (C.7) will be numerically stable whenever $b_2 > 0$.

In our Online Supplement (Appendix 2), we list numerous special cases for which $_3F_2(1)$ has known closed-form solution.

# D  The score function and information matrix

We impose Assumption 2 in this section and recall that the likelihood function is given by

$$\mathcal{L}_{P^*}(\boldsymbol{\theta} \mid \boldsymbol{P}) = \prod_{t\,:\,P_t < \alpha_1} F_P(\alpha_1 \mid \boldsymbol{\theta}) \prod_{t\,:\,\alpha_1 \leqslant P_t < \alpha_2} f_P(P_t \mid \boldsymbol{\theta}) \prod_{t\,:\,P_t \geqslant \alpha_2} \bar{F}_P(\alpha_2 \mid \boldsymbol{\theta}) \qquad \text{(D.1)}$$

where $\bar{F}_P(u)$ denotes the tail probability $1 - F_P(u)$.

We begin with the case of lower truncation, i.e., $P_t < \alpha_1$. For notational convenience, define $C_0(x) = \ln R(x)$. First and second derivatives follow as

$$C_1(x) = \frac{\rho(x)}{R(x)}, \qquad C_2(x) = -\lambda_\rho(x)C_1(x) - C_1(x)^2.$$

Let $\zeta_{\boldsymbol{\theta}}(p) = (R^{-1}(p) - \mu)/\sigma$. The first derivatives of the log-likelihood of the TLSF distribution are

$$\frac{\partial}{\partial \mu} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t < \alpha_1) = -C_1(\zeta_{\boldsymbol{\theta}}(\alpha_1))/\sigma \qquad \text{(D.2)}$$

$$\frac{\partial}{\partial \sigma} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t < \alpha_1) = -\zeta_{\boldsymbol{\theta}}(\alpha_1)C_1(\zeta_{\boldsymbol{\theta}}(\alpha_1))/\sigma \qquad \text{(D.3)}$$

and for the second derivatives we have

$$\frac{\partial^2}{\partial\mu^2} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t < \alpha_1) = (1/\sigma^2)C_2(\zeta_{\boldsymbol{\theta}}(\alpha_1)) \tag{D.4}$$

$$\frac{\partial^2}{\partial\mu\partial\sigma} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t < \alpha_1) = (1/\sigma^2)\left(C_1(\zeta_{\boldsymbol{\theta}}(\alpha_1)) + \zeta_{\boldsymbol{\theta}}(\alpha_1)C_2(\zeta_{\boldsymbol{\theta}}(\alpha_1))\right) \tag{D.5}$$

$$\frac{\partial^2}{\partial\sigma^2} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t < \alpha_1) = (1/\sigma^2)\left(2\zeta_{\boldsymbol{\theta}}(\alpha_1)C_1(\zeta_{\boldsymbol{\theta}}(\alpha_1)) + \zeta_{\boldsymbol{\theta}}(\alpha_1)^2 C_2(\zeta_{\boldsymbol{\theta}}(\alpha_1))\right). \tag{D.6}$$

The case of upper truncation is similar. Define $\bar{C}_0(x) = \ln \bar{R}(x)$. First and second derivatives follow as

$$\bar{C}_1(x) = -\frac{\rho(x)}{\bar{R}(x)}, \qquad \bar{C}_2(x) = -\lambda_\rho(x)\bar{C}_1(x) - \bar{C}_1(x)^2.$$

The first and second derivatives of the log-likelihood of the TLSF distribution for this case take the same form as in (D.2)–(D.6) except with $C_1$ and $C_2$ replaced by $\bar{C}_1$ and $\bar{C}_2$ and with $\alpha_1$ replaced by $\alpha_2$.

For the intermediate continuous case of $P_t = p \in [\alpha_1, \alpha_2)$, we have

$$\frac{\partial}{\partial\mu} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t = p) = (1/\sigma)\lambda_\rho(\zeta_{\boldsymbol{\theta}}(p)) \tag{D.7}$$

$$\frac{\partial}{\partial\sigma} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t = p) = (1/\sigma)\left(\zeta_{\boldsymbol{\theta}}(p)\lambda_\rho(\zeta_{\boldsymbol{\theta}}(p)) - 1\right) \tag{D.8}$$

and for the second derivatives we have

$$\frac{\partial^2}{\partial\mu^2} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t = p) = (-1/\sigma^2)\lambda'_\rho(\zeta_{\boldsymbol{\theta}}(p)) \tag{D.9}$$

$$\frac{\partial^2}{\partial\mu\partial\sigma} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t = p) = (-1/\sigma^2)\left(\lambda_\rho(\zeta_{\boldsymbol{\theta}}(p)) + \zeta_{\boldsymbol{\theta}}(p)\lambda'_\rho(\zeta_{\boldsymbol{\theta}}(p))\right) \tag{D.10}$$

$$\frac{\partial^2}{\partial\sigma^2} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t = p) = (-1/\sigma^2)\left(2\zeta_{\boldsymbol{\theta}}(p)\lambda_\rho(\zeta_{\boldsymbol{\theta}}(p)) + \zeta_{\boldsymbol{\theta}}(p)^2\lambda'_\rho(\zeta_{\boldsymbol{\theta}}(p)) - 1\right). \tag{D.11}$$

Recall that the expected Fisher information matrix is defined as

$$\Upsilon(\boldsymbol{\theta})_{ij} = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} \ln \mathcal{L}_{P^*}(\boldsymbol{\theta} \mid P_t)\right),$$

implying that we need to integrate across the three cases. For the lower and upper truncation cases, we simply weight the respective expressions by $\alpha_1$ and $1 - \alpha_2$ and evaluate at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. For the intermediate case, we require integrals of the following forms:

$$\mathcal{A}_{\rho,k}(\alpha) = \int_0^\alpha R^{-1}(p)^k \lambda_\rho \left(R^{-1}(p)\right) dp = \int_{-\infty}^{R^{-1}(\alpha)} x^k \lambda_\rho(x)\rho(x)dx, \qquad k = 0, 1 \qquad \text{(D.12)}$$

$$\mathcal{B}_{\rho,k}(\alpha) = \int_0^\alpha R^{-1}(p)^k \lambda'_\rho \left(R^{-1}(p)\right) dp = \int_{-\infty}^{R^{-1}(\alpha)} x^k \lambda'_\rho(x)\rho(x)dx \qquad k = 0, 1, 2. \qquad \text{(D.13)}$$

Integrals of form $\mathcal{A}_{\rho,k}$ have general solution for any density $\rho$:

$$\mathcal{A}_{\rho,0}(\alpha) = -\rho\left(R^{-1}(\alpha)\right) \qquad \text{(D.14)}$$

$$\mathcal{A}_{\rho,1}(\alpha) = \alpha - R^{-1}(\alpha)\rho\left(R^{-1}(\alpha)\right) \qquad \text{(D.15)}$$

Integrals of form $\mathcal{B}_{\rho,k}(\alpha)$ depend on the chosen family of TLSF. When working with complementary pairs of skewed distributions, we can show that

$$\mathcal{A}^c_{\rho,k}(\alpha) = (-1)^{k+1}\left(\mathcal{A}_{\rho,k}(1) - \mathcal{A}_{\rho,k}(1 - \alpha)\right) \qquad \text{(D.16)}$$

$$\mathcal{B}^c_{\rho,k}(\alpha) = (-1)^k\left(\mathcal{B}_{\rho,k}(1) - \mathcal{B}_{\rho,k}(1 - \alpha)\right). \qquad \text{(D.17)}$$

We can now express the elements of the information matrix as:

$$\Upsilon(\boldsymbol{\theta}_0)_{1,1} = \mathcal{B}_{\rho,0}(\alpha_2) - \mathcal{B}_{\rho,0}(\alpha_1) - \alpha_1 C_2\left(R^{-1}(\alpha_1)\right) - (1-\alpha_2)\bar{C}_2\left(R^{-1}(\alpha_2)\right), \tag{D.18}$$

$$\begin{aligned}
\Upsilon(\boldsymbol{\theta}_0)_{1,2} = {}& \mathcal{A}_{\rho,0}(\alpha_2) - \mathcal{A}_{\rho,0}(\alpha_1) + \mathcal{B}_{\rho,1}(\alpha_2) - \mathcal{B}_{\rho,1}(\alpha_1) \\
& - \alpha_1\left(C_1\left(R^{-1}(\alpha_1)\right) + R^{-1}(\alpha_1)C_2\left(R^{-1}(\alpha_1)\right)\right) \\
& - (1-\alpha_2)\left(\bar{C}_1\left(R^{-1}(\alpha_2)\right) + R^{-1}(\alpha_2)\bar{C}_2\left(R^{-1}(\alpha_2)\right)\right), \tag{D.19}
\end{aligned}$$

$$\begin{aligned}
\Upsilon(\boldsymbol{\theta}_0)_{2,2} = {}& 2(\mathcal{A}_{\rho,1}(\alpha_2) - \mathcal{A}_{\rho,1}(\alpha_1)) + \mathcal{B}_{\rho,2}(\alpha_2) - \mathcal{B}_{\rho,2}(\alpha_1) - (\alpha_2 - \alpha_1) \\
& - \alpha_1\left(2R^{-1}(\alpha_1)C_1\left(R^{-1}(\alpha_1)\right) + R^{-1}(\alpha_1)^2 C_2\left(R^{-1}(\alpha_1)\right)\right) \\
& - (1-\alpha_2)\left(2R^{-1}(\alpha_2)\bar{C}_1\left(R^{-1}(\alpha_2)\right) + R^{-1}(\alpha_2)^2\bar{C}_2\left(R^{-1}(\alpha_2)\right)\right). \tag{D.20}
\end{aligned}$$

# References

**Abramowitz, M. and I. A. Stegun, eds**, *Handbook of Mathematical Functions*, New York: Dover Publications, 1965.

**Acerbi, C. and B. Székely**, "Back-testing Expected Shortfall," *Risk*, December 2014, *26* (12).

**_ and B. Székely**, "Backtestability and the ridge backtest," *Frontiers of Mathematical Finance*, 2023, *2* (4), 497–521.

**Barendse, S., E. Kole, and D. van Dijk**, "Backtesting Value-at-Risk and Expected Shortfall in the Presence of Estimation Error," *Journal of Financial Econometrics*, 2023, *21* (2), 528–568.

**Basel Committee on Bank Supervision**, "Minimum capital requirements for market risk," Publication No. 457, Bank for International Settlements January (rev. February) 2019.

**Bayer, S. and T. Dimitriadis**, "Regression-Based Expected Shortfall Backtesting," *Journal of Financial Econometrics*, 09 2022, *20* (3), 437–471.

**Berkowitz, J.**, "Testing the accuracy of density forecasts, applications to risk management," *Journal of Business & Economic Statistics*, 2001, *19* (4), 465–474.

_ , **P. Christoffersen, and D. Pelletier**, "Evaluating value-at-risk models with desk-level data," *Management Science*, 2011, *57* (12), 2213–2227.

**Du, Z. and J.C. Escanciano**, "Backtesting expected shortfall: accounting for tail risk," *Management Science*, 2017, *63* (4), 940–958.

_ , **P. Pei, X. Wang, and T. Yang**, "Powerful backtests for historical simulation expected shortfall models," *Journal of Business and Economic Statistics*, 2023, *42* (3), 864–874.

**Engle, R.F.**, "Stock volatility and the crash of '87: Discussion," *The Review of Financial Studies*, 1990, *3* (1), 103–106.

**Fernández, C. and M.F.J. Steel**, "On Bayesian modeling of fat tails and skewness," *Journal of the American Statistical Association*, 1998, *93* (441), 359–371.

**Fissler, T., J.F. Ziegel, and T. Gneiting**, "Expected shortfall is jointly elicitable with value-at-risk: implications for backtesting," *Risk*, January 2016, *28* (1), 58–61.

**Glosten, L. R., R. Jagannathan, and D. E. Runkle**, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The Journal of Finance*, 1993, *48* (5), 1779–1801.

**Gneiting, T.**, "Making and Evaluating Point Forecasts," *Journal of the American Statistical Association*, 2011, *106* (494), 746–762.

_ , **F. Balabdaoui, and A.E. Raftery**, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society, Series B*, 2007, *69* (2), 243–268.

**González-Santander, Juan Luis**, "A Note on Some Reduction Formulas for the Incomplete Beta Function and the Lerch Transcendent," *Mathematics*, 2021, *9* (13), 1486.

**Gordy, Michael B. and Alexander J. McNeil**, "Spectral backtests of forecast distributions with application to risk management," *Journal of Banking and Finance*, 2020, *116*, 105817.

**Gradshteyn, I.S. and I.M. Ryzhik**, *Table of Integrals, Series, and Products*, seventh ed., New York: Academic Press, 2007.

**Hewitt, E.**, "Integration by Parts for Stieltjes Integrals," *The American Mathematical Monthly*, 1960, *67* (5), 419–423.

**Hoga, Y. and M. Demetrescu**, "Monitoring value-at-risk and expected shortfall forecasts," *Management Science*, 2023, *69* (5), 2954–2971.

**Hué, S., C. Hurlin, and Y. Lu**, "Backtesting expected shortfall: accounting for both duration and severity with bivariate orthogonal polynomials," 2024. Available at https://dx.doi.org/10.2139/ssrn.4816132.

**Iercosan, Diana, Alysa Shcherbakova, David McArthur, and Rebecca Alper**, "Beyond Exceedance-Based Backtesting of Value-at-Risk Models: Methods for Backtesting the Entire Forecasting Distribution Using Probability Integral Transform," in David Lynch, Iftekhar Hasan, and Akhtar Siddique, eds., *Validation of Risk Management Models for Financial Institutions: Theory and Practice*, Cambridge University Press, 2023, chapter 4, pp. 57–83.

**Lok, Hsiao Yen**, "Validating market risk models using realized PIT values." PhD dissertation, Heriot-Watt University, Edinburgh, UK 2017.

**Lynch, David, Valerio Potì, Akhtar Siddique, and Francesco Campobasso**, "Evaluation of Value-at-Risk Models: An Empirical Likelihood Approach," in David Lynch,

Iftekhar Hasan, and Akhtar Siddique, eds., *Validation of Risk Management Models for Financial Institutions: Theory and Practice*, Cambridge University Press, 2023, chapter 5, pp. 84–103.

**McNeil, A.J.**, "Modelling volatility with v-transforms and copulas," *Risks*, 2021, *9* (1), 14.

**Nelson, D. B.**, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 1991, *59*, 347–370.

**O'Brien, J. and P.J. Szerszen**, "An evaluation of bank measures for market risk before, during and after the financial crisis," *Journal of Banking and Finance*, July 2017, *80*, 215–234.

**Patton, A.J., J.F. Ziegel, and R. Chen**, "Dynamic semiparametric models for expected shortfall (and Value-at-Risk)," *Journal of Econometrics*, 2019, *211* (2), 388–413.

**Refuz, D. and M. Yor**, *Continuous martingales and Brownian motion*, Springer-Verlag, Berlin, 2004.

**Rosenblatt, M.**, "Remarks on a multivariate transformation," *Annals of Mathematical Statistics*, 1952, *23*, 470–472.

**Š. Porubský, T. Šalát, and O. Strauch**, "Transformations that preserve uniform distribution," *Acta Arithmetica*, 1988, *XLIX*, 459–479.

**Wolfram Research**, "Mathematical Functions Site," https://functions.wolfram.com/, as of 2023-07-23 2023.