Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Nonparametric conditional density specification testing and quantile estimation; with application to S&P500 returns

by

Patrick Marsh

Granger Centre Discussion Paper No. 19/02

University of Nottingham
UK | CHINA | MALAYSIA

Nonparametric Conditional Density Specification Testing and Quantile Estimation; with Application to S&P500 Returns.

Patrick Marsh
School of Economics and
Granger Centre for Time Series Econometrics
University of Nottingham
NG7 2RD
email: patrick.marsh@nottingham.ac.uk

February 11, 2019

Abstract:

This paper develops a two stage procedure to test for correct dynamic conditional specification. It exploits nonparametric likelihood for an exponential series density estimator applied to the in-sample Probability Integral Transforms obtained from a fitted conditional model. The test is shown to be asymptotically pivotal, without modification. Numerical experiments illustrate both this and also that it can have significantly more power than equivalent tests based on the empirical distribution function, when applied to a number of simple time series specifications. In the event of rejection, the second stage nonparametric estimator can both consistently estimate quantiles of the data, under empirically relevant conditions, as well as correct the predictive log-scores of mis-specified models. Both test and estimator are applied to monthly S&P500 returns data. The estimator leads to narrower predictive confidence bands which also enjoy better coverage and contributes positively to the predictive log-score of Gaussian fitted models. Additional application involves risk evaluation, such as Value at Risk calculations or estimation of the probability of a negative return. The contribution of the nonparametric estimator is particularly clear during the financial crisis of 2007/8 and highlights the usefulness of a specification procedure which offers the possibility of partially correcting rejected specifications.

Keywords: Conditional specification, series density estimator, nonparametric likelihood ratio, predictive quantiles for returns, log-score.

1 Introduction

Testing for the correctness of a particular distributional specification is a fundamental step in determining the adequacy of fitted models for Economic and Financial variables. Conditional specification tests generally involve goodness-of-fit type tests applied to the in-sample Probability Integral Transforms (PIT), see Diebold, Gunther and Tay (1998). Corradi and Swanson (2006b, 2012) provides a thorough review of such procedures as well as related tests of predictive densities. The latter being based on the out of sample PITs, obtained either recursively or via a rolling estimation window.

This paper provides a test of conditional specification based upon a consistent nonparametric density estimator, applied to the sequence of in-sample PITs. It is a direct generalization of the procedures developed in Marsh (2007, 2018) for independent and identically distributed (IID) data. These apply Portnoy's (1988) test in the context of Barron and Sheu's (1988) density estimator for the standard goodness-of-fit problems with independent sampling. The procedure is two-stage. In the first a standard dynamic conditional econometric model is specified and estimated via any quasi-likelihood approach that yields appropriately consistent nuisance parameter estimates. Allied to a probability specification a second stage consistently estimates the density function of the PITs. A likelihood ratio test applied to this density yields the specification test.

The approach is related to the smooth moment tests of Ledwina (1994), Kallenberg and Ledwina (1997) and Bontemps and Meddahi (2012) and is analogous to the generalization of Claeskens and Hjort (2004) to the evaluation of predictive densities by Lin and Wu (2017). It differs, significantly, in that the test statistic is a likelihood ratio based on a density estimator obtained from the moments, rather than on the moments themselves.

Typically tests for conditional specification (and predictive densities) have been based upon the empirical distribution function (EDF) of the PIT. Specifically, a correctly specified distribution will generate independent and uniform PITs, which may be tested via (adaptation of) standard EDF based tests, such as the Kolmogorov-Smirnov (KS) or Cramer-von Mises (CvM). Andrews (1997), Bai (2003) and Corradi and Swanson (2006b) have all proposed variants of such tests. Those tests based on the EDF can suffer from three potential shortcomings.

First, standard applications of KS or CvM tests are not asymptotically pivotal, in general. Simply testing for the simplest independent and identically distributed Gaussian formulation requires four sets of asymptotic critical values depending on what combination of mean and/or variance needs to be estimated, see Stephens (1976). In the predictive evaluation context this lack of pivotal-ness is termed 'estimation bias', Rossi and Sekhposyan (2015). Bai's (2003) marginalization approach is not generally applicable, and the bootstrap of Corradi and Swanson (2006b) requires strict stationarity, ruling out recursive estimation schemes, for instance. Second, tests based on the EDF tend to have low power compared to parametric procedures. This is exacerbated in the context of tests based on the out-of-sample PITs since a significant fraction of the sample must be dedicated to estimation. Thirdly, in the event of rejection, such tests do not offer any insight into how either the specification or predictive ability of the model can be improved. Indeed, such tests are only applied, if they are applied, after a battery of standard diagnostic tests. If the latter are not rejected, but goodness-of-fit is, then the applied researcher is left with no obvious avenue down which to proceed.

To address those three shortcomings, the proposed test is first shown to be asymptotically pivotal, without modification, and does not require strict stationarity. Numerical experiments involving some simple distributional specifications clearly indicate the pivotal nature of the procedures. Secondly, it has both theoretical and numerical power advantages. It is applied to in-sample PITs, and thus all the sample is available to test, it does not have to be split to first cover parameter estimation. Numerical power superiority is demonstrated over even the unfeasible variants of the in-sample EDF tests - i.e. when size corrected critical values are used. Experiments demonstrate such against empirically relevant alternatives such as mis-specification of unconditional skewness or kurtosis or of the dynamic structure of the conditional mean or variance.

Thirdly, the test is based upon the likelihood ratio of a nonparametric density estimator in the second stage. In the event of rejection this estimator can itself be used to correct probability or interval predictions. That is, we can consistently correct the quantile function of the in-sample PITs generated from a mis-specified conditional distribution. These quantiles then can be mapped back to the original sample space to correct the quantile function of the original fitted conditional distribution. Additional numerical experiments illustrate the accuracy of these corrected quantiles. Applying a proper score function, see Gneiting and Raftery (2007), to the corrected predictive density, the log-score is decomposable into the sum of two components. The first is the log-score from the initial fitted model. The second is the log-score of the non-parametric density estimator. Numerical experiments show the contribution of the latter is small, but positive. This analysis also highlights a desirable interpretation for the test itself. It may be viewed as a test of conditional predictive ability, in the spirit of Giacomini and White (2006) comparing the original fitted model with that of the two-stage procedure.

The tests and quantile estimator are applied to monthly data (adjusted for splits and dividends) on the S&P500 index, from December 1997 on. A number of simple time series models (from IID to AR-GARCH specifications) are estimated across the full sample as well as various sub-samples either side of the onset of the Global Financial Crisis of 2007/8. Each model's specification is checked via the proposed test. Then, exploiting the consistent quantile estimator, predictive confidence bands can be generated via both recursive and rolling estimation schemes. The quantiles from the density estimator are shown to be both narrower and have better empirical coverage than Gaussian quantiles, and those based on a rolling scheme more accurate than those from the recursive. Value at Risk calculations, as well as predictions for the probability of a negative return, based on both estimated and fitted Gaussian quantiles can be directly compared, with significant divergence observed beginning early in 2008.

The plan for the rest of the paper is as follows. In the next section a simpler, unfeasible, goodness-of-fit procedure is presented, in preparation for generalization to the conditional, dynamic framework with nuisance parameters. Section 3, proves that the density estimator introduced by Barron and Sheu (1989) remains consistent in this context, and that the likelihood ratio test is asymptotically standard normal and consistent against fixed alternatives. The test has power against the same rate of local alternatives as the original test of Marsh (2007). Section 4 simulates the finite sample size and power, and compares the latter to that of size-corrected KS and CvM tests. A corollary to Theorem 1 demonstrates that the quantiles of the data can be consistently estimated via this nonparametric procedure, further numerical experiments demonstrate this All of the properties of these procedures are demonstrated in an extended application to monthly S&P500 in Section 5. Section 6 concludes while all proofs, the tables for the numerical experiments as well as the graphs and further tables used in the application are in the appendix.

2 Preliminaries

2.1 Perfect Specification

Suppose that $\{Y_i\}_{i=1}^n$ is a sequence of random variables having conditional density functions $f_i = f_i\left(Y_i|\mathcal{F}_i;\beta\right)$, where β is a $k \times 1$ vector of parameters and $\mathcal{F}_i$ represents the information set available at point i in the sample, typically $\mathcal{F}_i$ will consist of both past values of Y as well as past and current values of any predictors. When β is known, testing the specification of f_i trivially collapses to the distribution free goodness-of-fit (GoF) problem. Marsh (2007) introduced a test for such, via an exponential series density estimator. Here we first simplify that procedure prior to subsequent generalization to conditional specifications with unknown parameters.

According to Lemma 1 of Bai (2003) evaluating the conditional cumulative distribution functions of $\{Y_i\}_{i=1}^n$ at those outcomes generates a sequence of IID Uniform random variables. That is we can test whether f_i is *perfectly specified* (i.e. the density family, parameter and conditioning set are all known) via;

$$H_0 : F_i = F_i \left(Y_i|\mathcal{F}_i;\beta\right) = \int_{-\infty}^{Y_i} f_i \left(y|\mathcal{F}_i;\beta\right) dy \sim IID\, U\left[0,1\right]. \tag{1}$$

To proceed, construct a sequence $\{X_i\}_1^n$, where $X_i = h\left(F_i\right)$ for some user-chosen monotone function $h\left(.\right)$. Under H_0 the X_i are IID copies of a variable X, having known distribution and density,

$$U_0(x) = \Pr\left[X < x\right] \quad \& \quad u_0\left(x\right) = \frac{dU_0\left(x\right)}{dx} \quad for \quad x \in (0,1).$$

The density $u\left(x\right)$ is first approximated via the exponential family,

$$p_x(\theta) = \exp\left\{\sum_{k=1}^{m} \theta_k\phi_k(x) - \psi_m\left(\theta\right)\right\}, \quad \psi_m\left(\theta\right) = \ln\int_0^1 \exp\left\{\sum_{k=1}^{m}\theta_k\phi_k(x)\right\} dx, \tag{2}$$

where the $\phi_k(x)$ are linearly independent functions spanning $\mathbb{R}^m$ and $\psi_m\left(\theta\right)$ is the cumulant function, such that $p_x\left(\theta\right)$ integrates to one over x.

Let the density $u\left(x\right)$ on $(0,1)$, satisfy $\log[u(x)] \in W_2^r$, the Sobolev space of functions on $(0,1)$ for which $d^{r-1}u\left(x\right)/dx^{r-1}$ is absolutely continuous and $d^r u\left(x\right)/dx^r$ is square-integrable. According to Crain (1974) and Barron and Sheu (1991) there exists a unique $\theta_{(m)} = (\theta_1, .., \theta_m)'$ satisfying

$$\int_0^1 \phi_k(x)p_x\left(\theta_{(m)}\right) dx = \int_0^1 \phi_k(x)u\left(x\right) dx \quad \text{for } k = 1, 2, ..., m, \tag{3}$$

and, as $m \to \infty$, $p_x\left(\theta_{(m)}\right)$ and $u\left(x\right)$ converge in relative entropy, with

$$E_U\left[\ln\left(\frac{u\left(x\right)}{p_x\left(\theta_{(m)}\right)}\right)\right] = \int_0^1 \ln\left(\frac{u\left(x\right)}{p_x\left(\theta_{(m)}\right)}\right) u\left(x\right) dx = O\left(m^{-2r}\right).$$

That $\theta_{(m)}$ is unique implies that H_0 can be tested instead via a simple hypothesis on $\theta_{(m)}$,

$$H_0 : F_i \sim IID\, U\left[0,1\right] \leftrightarrow H_0 : \theta_{(m)} = \theta_{(m)}^0, \tag{4}$$

where $\theta_{(m)}^0$ solves (3) with $u\left(x\right) = u_0\left(x\right) = dU_0\left(x\right)/dx$. A nonparametric likelihood ratio test in the exponential family (2) is,

$$\lambda_m = 2\sum_{i=1}^{n} \ln\left[\frac{p_{X_i}\left(\bar{\theta}_{(m)}\right)}{p_{X_i}\left(\theta_{(m)}^0\right)}\right],$$

where $\bar{\theta}_{(m)}$ is the unique maximum likelihood estimator (MLE) for $\theta_{(m)}$ satisfying

$$\int_0^1 \phi_k(x)\, p_x\left(\bar{\theta}_{(m)}\right) dx = \frac{\sum_{i=1}^n \phi_k(X_i)}{n} \quad \text{for } k = 1, 2, ..., m. \tag{5}$$

Suppose that $m, n \to \infty$ with $m^3/n \to 0$, then according to Theorem 1 of Barron and Sheu (1991), $p_x\left(\bar{\theta}_{(m)}\right)$ converges in relative entropy to $u(x)$

$$E_U\left[\ln\left(\frac{u(x)}{p_x\left(\bar{\theta}_{(m)}\right)}\right)\right] = \int_0^1 \ln\left(\frac{u(x)}{p_x\left(\bar{\theta}_{(m)}\right)}\right) u(x)\, dx = O_p\left(\frac{m}{n} + m^{-2r}\right),$$

while Theorem 1 of Marsh (2007) proves,

$$\Lambda_m = \frac{\lambda_m - m}{\sqrt{2m}} \to_d N(0,1). \tag{6}$$

Additionally, Λ_m diverges under any fixed (IID) alternative (i.e. the test is consistent) and it has power against local alternatives parametrized by $\theta_{(m)}^1 - \theta_{(m)} = c\sqrt{\frac{\sqrt{m}}{n}}$ with $c'c = 1$ and $\theta_{(m)}^1$ satisfies (3) but with $u(x) = u_1(x)$, the density of X under the alternative. Note that if we allow m to grow arbitrarily slowly then the local alternative rate approaches that of EDF based tests, $O\left(n^{1/2}\right)$.

Application of this procedure can become overly complicated, even in the simplest of GoF problems as in Marsh (2007), through the choices of both the monotone function $h(F)$ and on the basis $\phi_k(x)$. Although different choices may be worth pursuing on computational or numerical grounds, to simplify as much as is possible, here we impose $h(F) = F$ and choose the trigonometric basis, $\phi_k(x) = \{\cos[2k\pi x], \sin[2k\pi x]\}_{k=1}^{m/2}$.

2.2 Correct Specification

Assume now that β in (1) is unknown and must be estimated as a preliminary step prior to application of the likelihood ratio test described above. That is we test that the conditional density f_i is *correctly specified*, i.e. we know everything about f_i except β. Let $\hat{\beta}_n$ denote a (quasi) maximum likelihood estimator of β obtained from the sample $\{Y_i\}_{i=1}^n$ using the specified likelihood $L = \prod_{i=1}^n f(Y_i|\mathcal{F}_i; \beta)$.

Typically the alternative will be the (unspecified) negation of H_0. Under such an alternative, suppose that the observations are instead outcomes of random variables having (unspecified) density and distribution functions $g_i = g_i(y|\mathcal{G}_i)$ and $G_i(y|\mathcal{G}_i)$, for some information set $\mathcal{G}_i$, such that either $f_i \neq g_i$ and/or $\mathcal{F}_i \nsubseteq \mathcal{G}_i$ for some i. Denote this alternative via;

$$H_1 : G_i = \int_{-\infty}^{Y_i} g_i(y|\mathcal{G}_i)\, dy \sim IID\,U\,[0,1]\,.$$

We require the following assumptions on both $F_i(y|\mathcal{F}_i;\beta)$, $G_i(y|\mathcal{G}_i)$ and the respective densities $f(y|\mathcal{F}_i;\beta)$ and $g_i(y|\mathcal{G}_i)$, to ensure the existence of $\hat{\beta}_n$ and under which the asymptotic distribution of the proposed test will be derived.

Assumption 1 For all $i \in \mathbb{Z}^+$:

(i) The densities $f_i(y|\mathcal{F}_i;\beta)$ are measurable in y for every $\beta \in \mathbf{B}$, a compact subset of $p-$dimensional Euclidean space, and are continuous in β for every y.

(ii) The $G_i(y|\mathcal{G}_i)$ are absolutely continuous distribution functions and such that $\sup_i E_{\mathcal{F}_i}[\log[g_i(y|\mathcal{G}_i)]$ exists and $\sup_i |\log f_i(y_i|\mathcal{F}_i,\beta)| < v(y)$ for all β where $v(.)$ is integrable with respect to $G(.)$.

(iii) Let

$$I_i(\beta) = E_{\mathcal{F}_i}\left[\ln\left[\frac{g_i(y|\mathcal{G}_i)}{f_i(y_i|\mathcal{F}_i,\beta)}\right]\right] = \int_y \ln\left[\frac{g_i(y|\mathcal{G}_i)}{f_i(y_i|\mathcal{F}_i,\beta)}\right] g_i(y|\mathcal{F}_i)\, dy,$$

such that $\bar{I}(\beta) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} I_i(\beta)$ has a unique minimum at $\beta_* \in \mathbf{B}$.

(iv) $F_i(Y_i|\mathcal{F}_i,\beta)$ is continuously differentiable with respect to β and $H_i(\beta) = dF_i(Y_i|\mathcal{F}_i,\beta)/d\beta$ *is finite for all β in a closed ball of radius $n^{-1/2}$ around β_*.*

(v) Both $\log[g_i(y|\mathcal{G}_i)]$ and $\log[f_i(y_i|\mathcal{F}_i,\beta)]$ have $r \geq 2$ derivatives in y which are absolutely continuous and square integrable.

Note that, under H_0, together Assumption 1(i), (ii) and (v) and monotonicity of $F_i(y|\mathcal{F}_i;\beta)$ are sufficient for assumption A1 of Bai (2003) to hold. In addition assumption (iii) implies that, for the log-likelihood criterion, the conditions of Theorems 2.2 and 2.3 of Domowitz and White (1982) are met, and therefore $\hat{\beta}_n$ exists and

$$\hat{\beta}_n = \beta_* + O_p(n^{-1/2}).$$

That is, $\hat{\beta}_n$ is a $\sqrt{n}$ consistent Quasi maximum likelihood estimator for the pseudo-true value β_*. Note that under H_0 we have $\beta_* = \beta$, while under H_1 we will have $\beta_* \neq \beta$.

To derive the test, first denote $\hat{X}_i = F\left(Y_i|\mathcal{F}_i, \hat{\beta}_n\right)$, with the mean value expansion

$$\hat{X}_i = F_i\left(Y_i, \beta_*\right) + \left(\hat{\beta}_n - \beta_*\right)' H_i\left(\beta^+\right),$$

where β^+ lies on a line segment joining $\hat{\beta}_n$ and β_*. As a consequence, we can write

$$\hat{X}_i = X_i + e_i, \tag{7}$$

where, although unobserved, under H_0, $X_i = F_i\left(Y_i, \beta\right) \sim IID\,U\left[0,1\right]$, while under H_1, $X_i = F_i\left(Y_i, \beta_*\right) \nsim IID\,U\left[0,1\right]$. Both by construction and as a consequence of Assumption 1 (iv),

$$e_i \in (-1,1) \quad \& \quad e_i = O_p\left(n^{-1/2}\right). \tag{8}$$

In general, in (7) e_i will be both heterogeneous and dependent. However, for what follows it is only necessary that it is both bounded and degenerate.

The modification required to deal with the fact that β must be estimated is as follows. We are still testing on the distribution $U_0(x)$ (here the Uniform distribution) however we do not observe outcomes on X_i, but instead those on $\hat{X}_i$. Trivially, the uniform density satisfies $\log\left[u\left(x\right)\right] \in W_2^{\infty}$.

The maximum likelihood estimator for the parameter in the exponential family (2), say $\hat{\theta}_{(m)}$, based on the likelihood $\hat{L}\left(\theta_{(m)}\right) = \prod_{i=1}^n p_{\hat{X}_i}\left(\theta_{(m)}\right)$ satisfies

$$\int_0^1 \phi_k\left(x\right) p_x\left(\hat{\theta}_{(m)}\right) dx = \frac{\sum_{i=1}^n \phi_k\left(\hat{X}_i\right)}{n} \quad \text{for } k = 1, 2, ..., m, \tag{9}$$

which follows from (5) and using (7).

In the presence of nuisance parameters, testing the specification of $f_i\left(Y_i|\mathcal{F}_i;\beta\right)$, will entail testing $H_0 : X_i \sim IID\,U[0,1]$ (equivalently, $H_0 : \theta_{(m)} = 0_{(m)}$, similar to (4)) but using the likelihood ratio

$$\hat{\lambda}_m = 2\sum_{i=1}^n \ln\left[\frac{p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)}{p_{\hat{X}_i}\left(0_{(m)}\right)}\right] = 2\sum_{i=1}^n \ln p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right), \tag{10}$$

since the X_i are not observed. The following section details the asymptotic properties of $\hat{\lambda}_m$.

3 Asymptotic Properties

3.1 Density estimator under H_0

First it is required that the density estimator still converges in relative entropy to the (in this case Uniform) density of X_i. If not then $H_0 : X_i \sim IIDU(0,1)$ is not equivalent to $H_0 : \theta_{(m)} = 0_{(m)}$.

Key to the required generalization is that, in (7), we do not observe directly a sample from the random variable upon which the hypothesis is being tested. If we knew β we could observe X_i directly and obtain the maximum likelihood estimator $\bar{\theta}_{(m)}$ via (5). Instead, in the nuisance parameter case, we only observe $\hat{X}_i$ and obtain $\hat{\theta}_{(m)}$ via (9) and apply Portnoy's (1988) test, (10), using that.

For a given m, this test is just an application of a likelihood-ratio test in a linear exponential family. For the given choice of basis $\phi_{(m)}(x) = (\phi_1(x), ..\phi_m(x))'$, define the m dimensional statistics $\bar{x}_{(m)}$ and $\hat{x}_{(m)}$, by

$$\bar{x}_{(m)} = \left(\frac{\sum_{i=1}^{n} \phi_k(X_i)}{n} \right)_{k=1}^{m} \quad \text{and} \quad \hat{x}_{(m)} = \left(\frac{\sum_{i=1}^{n} \phi_k\left(\hat{X}_i\right)}{n} \right)_{k=1}^{m}.$$

Asymptotic properties will be driven by the difference between the observed sufficient statistic $\hat{x}_{(m)}$ in the correctly specified case with the unobserved $\bar{x}_{(m)}$, pertaining to perfect specification.

The respective maximum likelihood estimators, $\bar{\theta}_{(m)}$ and $\hat{\theta}_{(m)}$ then satisfy

$$\int_0^1 \phi_{(m)}(x)\, p_x\left(\bar{\theta}_{(m)}\right) dx = \bar{x}_{(m)} \quad \text{and} \quad \int_0^1 \phi_{(m)}(x)\, p_x\left(\hat{\theta}_{(m)}\right) dx = \hat{x}_{(m)}. \tag{11}$$

Standard properties of the linear exponential family still apply, specifically the duality between the (sufficient statistic) sample space, say $\mathcal{T}_m$, and the parameter space, say Θ_m. As in Barndorff-Nielsen (1978), consider arbitrary points in both $\mathcal{T}_m$ and Θ_m,

$\omega_{(m)} = \{\omega_1, .., \omega_m\}' \in \mathcal{F}_m$ and $\theta^*_{(m)} = (\theta_1, .., \theta_m) \in \Theta_m$ then the system of m equations

$$\int_0^1 \phi_k(x)\, p_x\left(\theta^*_{(m)}\right) dx = \omega_k, \quad k = 1, .., m, \tag{12}$$

has a unique solution. Denote this solution $\left\langle \theta^*_{(m)} : \omega_{(m)} \right\rangle$, where $\omega_{(m)} = (\omega_1, .., \omega_m)'$. That is solving (12) generates a one-to-one mapping between $\mathcal{T}_m$ and Θ_m.

Here we will be interested in three pairs of points in each space and the mapping between them. As well as the statistics $\bar{x}_{(m)}$ and $\hat{x}_{(m)}$ in $\mathcal{T}_m$, we have $\mu_{(m)} = E_U\left(\phi_{(m)}(x)\right)$. The three solutions to (12) we are interested in are:

$$\left\langle \theta_{(m)} : \mu_{(m)} \right\rangle, \quad \left\langle \bar{\theta}_{(m)} : \bar{x}_{(m)} \right\rangle \quad \text{and} \quad \left\langle \hat{\theta}_{(m)} : \hat{x}_{(m)} \right\rangle, \tag{13}$$

where the latter two are the unfeasible and feasible MLEs defined in (11) and the first represents the population exponential parameter and mean vectors. Note that although these points in (13) depend on the choice of basis ϕ, here we will suppress the dependence for notational brevity, and that under H_0, $\mu_{(m)} = 0_{(m)}$.

In summary, $\theta_{(m)}$ in Θ_m maps from the expectation of the (unobserved) statistic $\bar{x}_{(m)}$, $\mu_{(m)} = E\left[\bar{x}_{(m)}\right]$. The (unfeasible) MLE for $\theta_{(m)}$, if $\bar{x}_{(m)}$ were observed, is $\bar{\theta}_{(m)}$, while for the observed sufficient statistic $\hat{x}_{(m)}$, the (feasible) MLE is $\hat{\theta}_{(m)}$. By exploiting these dualities, we first show that the estimated density $p_x\left(\hat{\theta}_{(m)}\right)$ converges in relative entropy at exactly the same rate as $p_x\left(\bar{\theta}_{(m)}\right)$. The proof of the following theorem is proved in Appendix A.

Theorem 1 *Let $\hat{\theta}_{(m)}$ denote the estimated exponential parameter determined by (9) then under Assumption 1 and for $m, n \to \infty$ with $m^3/n \to 0$,*

$$E_U\left[\ln\left(\frac{1}{p_x\left(\hat{\theta}_{(m)}\right)}\right)\right] = \int_0^1 \ln\left(\frac{1}{p_x\left(\hat{\theta}_{(m)}\right)}\right) dx = O_p\left(\frac{m}{n}\right). \quad \blacksquare$$

According to Theorem 1, in terms of the density estimator, at least, the effect of observing $\left\{\hat{X}_1, .., \hat{X}_n\right\}$ rather than $\{X_1, .., X_n\}$ is asymptotically negligible under Assumption 1. It should not be surprising that the rate of convergence is unaffected when parameters are replaced by $\sqrt{n}$ consistent estimators.

3.2 Properties of the Likelihood Ratio Test

Full implementation proceeds as follows. Let $\hat{X}_i = F\left(Y_i, \hat{\beta}_n\right)$ and $X_i = F\left(Y_i, \beta\right)$ where X_i has uniform distribution and density $u\left(x\right) = 1$, then testing H_0 as in (1) is equivalent to testing

$$H_0 : \theta_{(m)} = 0_{(m)},$$

in the exponential family (2). The likelihood ratio test of Portnoy (1988) applied via the density estimator of Crain (1974) and Barron and Sheu (1991) obtained from the sample $\left\{\hat{X}_1, .., \hat{X}_n\right\}$ is

$$\hat{\lambda}_m = 2\sum_{i=1}^{n} \log\left[\frac{p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)}{p_{\hat{X}_i}\left(0_{(m)}\right)}\right] = 2n\left[\hat{\theta}'_{(m)}\hat{x}_{(m)} - \psi_m\left(\hat{\theta}_{(m)}\right)\right],$$

where $\hat{\theta}_{(m)}$ solves (9). The null hypothesis is rejected for large values of $\hat{\lambda}_m$.

Consider the fixed alternative

$$H_1 : G_i\left(y|\mathcal{G}_i\right) \neq F_i\left(y_i|\mathcal{F}_i; \beta\right) \tag{14}$$

such that the sequence $X_i = F_i\left(Y_i; \beta_*\right)$ is not uniform, identical or independent, in general, and satisfies;

$$\lim_{n\to\infty} n^{-1}\sum\nolimits_{i=1}^{n} \phi_k\left(X_i\right) = \mu_k^1 \neq 0 \text{ for some value(s) of } k \in \mathbb{Z}^+. \tag{15}$$

For every such alternative distribution for Y, then via the unique duality implied by (12) there will be a unique alternative distribution for X on $(0,1)$. Associated with that distribution will be another consistent density estimator given, say $p_x(\theta^1_{(m)})$. In practice, of course, $\theta^1_{(m)}$ will be neither specified nor known. The following Theorem, also proved in Appendix A, gives the asymptotic distribution of the likelihood ratio test statistic under (4) and demonstrates consistency against any fixed alternative, as defined by (14).

Theorem 2 *Suppose that we construct* $\left\{\hat{X}_i\right\}_{i=1}^{n}$ *as described above, that the conditions required in Assumption 1 are met and that* $m, n \to \infty$ *with* $m^3/n \to 0$*, then:*

(i) Under the null hypothesis (1),

$$\hat{\Lambda}_m = \frac{\hat{\lambda}_m - m}{\sqrt{2m}} \to_d N(0,1).$$

(ii) Under a fixed alternative (14), and for any finite κ,

$$\Pr\left[\hat{\Lambda}_m \geq \kappa\right] \to 1. \quad \blacksquare$$

Theorem 2 demonstrates that for any fixed alternative that leads to the PITs being non-uniform on $(0,1)$ the test will consistently reject. Since these asymptotic results arise via convergence to what pertains in the (unfeasible) perfect specification case, the test will have power against the same rate of local alternatives. This implies that $O\left(n^{1/2}\right)$ can be attained when m is grown arbitrarily slowly.

Alternatives which imply that the sequence $\{X_i\}_1^n$ remains marginally uniform - i.e. (15) does not hold- cannot be altogether dismissed. Consequently, a test for independence might also be applied in the event of non-rejection by $\hat{\Lambda}_m$. Such a test is, for instance, detailed in Lin and Wu (2017). Here, however, we will wish to pursue the option of exploiting the density estimator upon which $\hat{\Lambda}_m$ is based, to instead provide corrected (conditional) quantiles for the data. The conditions under which such is possible will then negate the need to test for independence among the $\{X_i\}_1^n$.

4 Numerical Properties

The purpose of this section is to illustrate the properties of the nonparametric likelihood tests and estimators described above. First we explore, numerically, the implication of Theorem 2(i), that as we increase the model dimension m and as the sample size n increases, critical values from the standard normal distribution apply. Both unconditional and conditional model specifications are employed to this end.

We then compare the powers of two versions of the test ($m = 4, 10$) against standard EDF based tests in this field, the Kolmogorov-Smirnov (KS) and Cramer-von Mises (CvM) tests. I.e. those that form the basis for the operational procedures in Andrews (1997) and Bai (2003).

The last set of experiments concern what we may do if the test rejects the null hypothesis. The tests of this paper are based on a consistent density estimator. We can simulate mean-square errors for the estimators of the quantiles of the correct distribution, when an incorrect distribution is used to generate the PITs. Again both conditional and unconditional model specifications are employed. All experiments detailed below were performed using Mathematica 8 and are based on 10000 Monte Carlo replications. All tables of outcomes are presented in Appendix B.

4.1 Numerical properties under the null

Theorem 2 proves that the likelihood ratio test $\hat{\Lambda}_m$ is asymptotically pivotal, specifically standard normal, and consistent against fixed alternatives. Competitor tests, such as KS and CvM these tests are mathematically detailed in Stephens (1976). Outside of the IID case, such tests require either significant adaptation to be applicable, or bootstrap schemes need to be both formally justified and applied. The proposed test requires neither.

Tables B1 and B2, provide rejection frequencies for the test $\hat{\Lambda}_m$ for values of $m = 2, .., 12$, for sample sizes $n = 25, 50, 100, 200$ and for three significance levels, $\alpha = 0.01, 0.05, 0.10$.

We are first interested in testing the unconditional null hypotheses

$$H_0^E : Y \sim Exp\,(\beta) \quad \& \quad H_0^N : Y \sim N\left(\mu, \sigma^2\right).$$

Letting $\bar{y}$ and $\hat{\sigma}^2$ be the estimated mean and variance (i.e. $\hat{\beta}_n = \bar{y}$ for H_0^E and $\hat{\beta}_n = \left(\bar{y}, \hat{\sigma}^2\right)'$ for H_0^N) then the tests are constructed from the mapping to $(0, 1)$, i.e. from $\hat{X}_i = 1 - e^{-Y_i/\bar{y}_n}$ to test H_0^E and $\hat{X}_i = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{Y_i - \bar{y}_n}{\hat{\sigma}_n}\right)\right]$ to test H_0^N. Table B1 provides rejection frequencies for both cases. Values of $\beta = 1$ for H_0^E and $\beta = (0, 1)'$ for H_0^N were chosen to generate the data.

We also generalize H_0^N so as to allow for both heterogeneity and dependence. First let $\mathcal{F}_i = (Y_{i-1}, .., Y_0)$ and test that $\{Y_i\}_1^n$ is generated by a simple autoregression;

$$H_0^{AR1} : Y_i|\mathcal{F}_i \sim N\left(\gamma_0 + \gamma_1 Y_{i-1}, \sigma^2\right),$$

with $Y_0 = 0$. Second let $Z_i \sim IID\,U[0,1]$ and $\mathcal{F}_i = (Z_{i-1}, ..., Z_0)$, and test that $\{Y_i\}_1^n$ is generated by a simple predictive regression;

$$H_0^{PR} : Y_i|\mathcal{F}_i \sim N\left(\gamma_0 + \gamma_1 Z_{i-1}, \sigma^2\right),$$

with $Z_0 = 0$. In both cases $\beta = (\gamma_0, \gamma_1, \sigma^2)'$ and we take $\hat{\beta}_n$ to be the OLS estimator for those parameters. To generate the data we set $\beta = (0.5, 0.5, 1)'$.

Table B2 provides rejection frequencies for H_0^{AR1}(left) and H_0^{PR} (right). What is demonstrated in Tables B1 and B2 is that for all four cases the procedures described finite sample rejection frequencies do become close to nominal as both n and m increase. That this happens across a range of significance levels illustrates the asymptotic pivotal nature of the tests more clearly than if only a single significance level were chosen.

4.2 Numerical properties under the alternative

No purpose is served by comparing null rejection frequencies with those tests that are not asymptotically pivotal. Instead, table B3 compares the 5% size corrected powers of the base KS and CVM tests. For a single sample size of $n = 100$, tables B3a and B3b compare (size corrected) rejection frequencies for Λ_4 and Λ_{10} against those for the KS and CvM tests for testing H_0^N under alternatives that the data is instead drawn from,

$$H_1^a : Y \sim t_{(v)}, \quad H_1^b : Y \sim \chi^2_{(v)} - v.$$

Tables B3c, B3d and B3e consider alternatives where the moments of the data are not correctly specified, i.e.

$$\begin{aligned} H_1^c &: Y_i|\mathcal{F}_i \sim N\left(vY_{i-1}, 1\right), \\ H_1^d &: Y_i|\mathcal{F}_i \sim N\left(0, 1 + vY_{i-1}^2\right), \\ H_1^e &: Y_i \sim N\left(v \times 1\left(i > \lfloor n/2 \rfloor\right), 1\right), \end{aligned}$$

where $1(.)$ denotes the indicator function. These latter three alternatives represent simplistic variants of common types of mis-specification in econometric or financial data, i.e. mis-specification of a conditional mean, variance or the possibility of a break in the mean (here half way through the sample). Note that these models trivially satisfy Assumption 1, but X_i as defined in (7) will not be IID on $(0,1)$. Lastly, table B3f considers testing H_0^E against the alternative

$$H_1^f : Y \sim \Gamma(1, v) .$$

Note that for each table the left hand panel corresponds to the case where we construct the test imposing the parameter values specified in the null rather than estimating them (i.e. using the test in (6)), whereas the right hand panels do not impose these values.

The outcomes in Table B3 imply the following broad conclusions. The nonparametric likelihood test based $\hat{\Lambda}_4$ is the most powerful almost uniformly, across all alternatives and whether parameters are estimated or not. The observed lack of power of the most commonly used test, KS, is particularly evident, it is consistently the poorest performing test.

Collectively, from these first 3 tables we conclude that $\hat{\Lambda}_{10}$ has size close to nominal and power on average superior to that of the EDF based tests. Its prime advantage, however, is that it is based on an asymptotically pivotal procedure.

4.3 Density estimation under the alternative: location-scale time series

The final shortcoming of EDF based tests of goodness-of-fit, and diagnostics in general, is that rejection of the null hypothesis is not indicative of how the specification could or should be changed. The tests of this paper, however, are based on the consistent nonparametric density estimator of Barron and Sheu (1991). This consistency can readily be extended to the current context of the presence of nuisance parameters in conditional, rather than marginal, densities.

Suppose that the data are generated according to

$$Y_i = \mu_{i|\mathcal{F}_i}(\beta) + \sigma_{i|\mathcal{F}_i}(\beta)\,\varepsilon_i, \quad i = 1, ..., n, \tag{16}$$

where $\mu_{i|\mathcal{F}_i}(\beta) = E[Y_i|\mathcal{F}_i]$ and $\sigma^2_{i|\mathcal{F}_i}(\beta) = E\left[\left(Y_i - \mu_{i|\mathcal{F}_i}(\beta)\right)^2 |\mathcal{F}_i\right]$ represent the conditional mean and variance of Y_i, respectively. These depend on some unknown $k \times 1$ dimensional parameter β. Suppose data is generated according to (16) with $\beta = \beta_0$, some fixed value in $\mathbb{R}^k$. Let $\hat{\beta}_n$ denote any (quasi maximum likelihood) estimator for β, satisfying,

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = O_p(1).$$

That is we require that the estimator for β_0 is consistent under both the null (that the conditional distribution is correctly specified) and the alternative (that the error distribution is different from that specified under the null). We additionally make the following assumption:

Assumption 2: (i) In (16) assume that the process $\{\varepsilon_i\}_1^n$ is IID, $E[\varepsilon_1] = 0$ and $V[\varepsilon_1] = 0$, having density function $g_{\varepsilon_1}(e)$, and

(ii) Let $\mathcal{B}$ denote a closed ball of radius $cn^{-1/2}$, for some finite $c > 0$, centered on β_0 then the conditional mean and variance of Y_i satisfy

$$\begin{aligned} \sup_{i,\beta\in\mathcal{B}} \left|\mu_{i|\mathcal{F}_i}\left(\hat{\beta}_n\right) - \mu_{i|\mathcal{F}_i}(\beta_0)\right| &= O_p\left(n^{-1/2}\right) \quad \text{and} \\ \sup_{i,\beta\in\mathcal{B}} \left|\sigma_{i|\mathcal{F}_i}\left(\hat{\beta}_n\right) - \sigma_{i|\mathcal{F}_i}(\beta_0)\right| &= O_p\left(n^{-1/2}\right). \end{aligned}$$

Assumption 2 is satisfied for a range of parametric and semi-parametric time series models, such as those generated by ARMA and/or GARCH processes. Under this assumption define, for any $\beta \in \mathcal{B}$ the generalized residuals (e.g. see Randles (1984)) by

$$\varepsilon_i(\beta) = \sigma^{-1}_{i|\mathcal{F}_i}(\beta)\left(Y_i - \mu_{i|\mathcal{F}_i}(\beta)\right),$$

so that $\varepsilon_i \sim IID\ \varepsilon_1$, with density $g_{\varepsilon_1}(\varepsilon)$. Now suppose that $f(\varepsilon)$ and $F(\varepsilon)$ are any density and invertible cumulative distribution function (with $f(\varepsilon) = dF(\varepsilon)/d\varepsilon$) satisfying the conditions of Assumption 1 then we immediately obtain the following corollary to Theorem 1.

Corollary 1 Let ε_1 have density $g_{\varepsilon_1}(\varepsilon)$, and define the variables $\hat{X}_i = F\left(\varepsilon_i\left(\hat{\beta}_n\right)\right)$ and $X_i = F(\varepsilon_i(\beta_0))$, so that $X_i \sim IIDX$ on $(0,1)$. Let $\hat{T}_{n,m} \in (0,1)$ be a random variable having density function $p_t\left(\hat{\theta}_{(m)}\right)$ where $\hat{\theta}_{(m)}$ is defined by (9), then whether or not $g_{\varepsilon_1}(\varepsilon) = f(\varepsilon)$,

$$\hat{T}_{n,m} \underset{\mathcal{L}}{\rightarrow} X,$$

as $n, m \rightarrow \infty, m^3/n \rightarrow 0$. I.e. $\hat{T}_{n,m}$ converges in law to the random variable X. ■

Assumption 2 requires that the model is correctly specified, but only up to the conditional mean and variance (not the distribution of the errors ε_i) and that the former may be consistently estimated to order $O_p\left(n^{-1/2}\right)$. Note also that under these conditions the independence of the $\{X_i\}_1^n$ can be assured via standard time series methods, such as consistent lag-length selection in the specification of both conditional mean and variance.

The accuracy of the resulting consistent quantile estimators is explored in the following numerical experiments. Suppose $\{Y_i\}_1^n$ is generated by the $AR(1)$ model,

$$Y_i = \gamma_0 + \gamma_1 Y_{i-1} + \varepsilon_i, \tag{17}$$

and define $\hat{\mu}_{i|\mathcal{F}_i} = \hat{\gamma}_{0,n} + \hat{\gamma}_{1,n} Y_{i-1}$ and $\hat{\sigma}_{i|\mathcal{F}_i} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{\gamma}_0 + \hat{\gamma}_1 Y_{i-1})^2}{n-2}}$, where $\hat{\gamma}_{0,n}$ and $\hat{\gamma}_{1,n}$ are OLS estimators. Apply the series density estimator to the sample

$$\hat{X}_i = \frac{1}{2}\left[1 + \operatorname{erf}\left(\hat{\sigma}_{i|\mathcal{F}_i}^{-1}\left(Y_i - \hat{\mu}_{i|\mathcal{F}_i}\right)\right)\right], \quad i = 1, ..., n,$$

and construct quantile estimators, for each Y_i (from the resulting density estimator $p_t\left(\hat{\theta}_{(m)}\right)$, which has quantiles $q_{\hat{T}_{n,m}}(\pi)$) via

$$\hat{q}_{Y_i}(\pi) = \hat{\mu}_{i|\mathcal{F}_i} + \hat{\sigma}_{i|\mathcal{F}_i}\sqrt{2}\operatorname{erf}^{-1}\left(2q_{\hat{T}_{n,m}}(\pi) - 1\right).$$

Putting $\gamma_0 = 1$ and $\gamma_1 = 0.3$, the mean square errors of $\hat{q}_{Yi}(\pi)$ are presented in Tables B4a (for $\varepsilon_i \sim \bar{t}_{(4)}$, standardized t) and Table B4b (for $\varepsilon_i \sim \bar{\chi}^2_{(4)}$, standardized χ^2) for both $m = 4$ and $m = 10$. The estimated quantiles converge numerically to their population counterparts, albeit slowly in the case of the extreme quantiles in the very skewed case. In small samples increasing the dimension of m has no significant impact on the accuracy. That is although a large value of m is required for the test statistic to be correctly sized, such is not required to accurately estimate (under the conditions of Corollary 1) the quantiles of the process generating the data.

A final implication of Corollary 1 is that it allows for the nonparametric estimator $p_t\left(\hat{\theta}_{(m)}\right)$ to correct the logarithmic predictive score of the fitted model $F\left(\varepsilon_i\left(\hat{\beta}_n\right)\right)$. As in the proof of Corollary 1 let $p_x\left(\theta^1_{(m)}\right)$ be defined as in (A.12), which converges in relative entropy to $u_1(x)$ the density function of $X_i = F(\varepsilon_i(\beta_0))$. Defining the inverse mapping $\varepsilon_i = F^{-1}(X_i;\beta_0)$, then the change of variable formula implies that

$$g_{\varepsilon_1}(\varepsilon) = p_x\left(\theta^1_{(m)}\right) \times f(\varepsilon_1(\beta_0)),$$

so that (in sample) predictive log-scores (see for example, Gneiting and Raftery (2007)) can be corrected, according to

$$\widehat{LS}_i = \ln p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right) + \ln f_{Yi}\left(\hat{\beta}_n\right), \tag{18}$$

where $f_{Yi}\left(\hat{\beta}_n\right) = f\left(\hat{\sigma}^{-1}_{i|\mathcal{F}_n}\left(Y_i - \hat{\mu}_{i|\mathcal{F}_n}\right)\right)$ is the in sample log-score for the original fitted model and $\ln p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)$ is the log-score obtained from the second, nonparametric, stage.

To illustrate, the model in (17) was simulated with standardized $\bar{t}_v$ errors, with $v = 4, 8$. For each case, and for sample sizes from 25 to 200, a (mis-specified) Gaussian $AR(1)$ was estimated and, from the resulting PITs, $p_x\left(\hat{\theta}_{(m)}\right)$ also estimated, for $m = 4, 10$. Monte-Carlo averages for both components of (18) are reported in Table B4c. In addition the log-scores for correctly specified Student t models (first assuming v is known, and second estimating v via profile likelihood) are presented for comparison.

The log-score is not a metric, so for comparative purposes note that a perfectly specified IID student $\bar{t}_4$ has a log-score of -1.682. The gains from the second stage are not huge, however the procedure can recover up to around 15% of the log-score lost by estimating the mis-specified model. As with quantile estimation there is little or no gain from increasing m.

Finally note that (18) offers a useful interpretation of the test, in light of the predictive ability tests of Giacomini and White (2006). The null hypothesis is that $f(\varepsilon_i(\beta_0))$, is correctly specified. Under this null $E\left[\ln p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)\right] = 0$, and the quadratic loss from fitting the model $f(\varepsilon_i(\beta_0))$ compared to the 'corrected' $g_{\varepsilon_i}(\varepsilon)$ can be measured by $\left(\ln p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)\right)^2$. Applying this to out-of-sample PITs and constructing the relevant test of Giacomini and White (2006) would form an alternate predictive specification test. This would be at the cost of dedicating a portion of the sample to estimate the unknown parameters, unlike the proposed test.

4.4 Guidance on implementation

According to Theorem 2 to test conditional specification, Gaussian critical values for the statistic $\hat{\Lambda}_{(m)}$ will be asymptotically correctly sized provided only that $m \to \infty$, while $m^3/n \to 0$. Also under this rate, according to Corollary 1, estimated quantiles for the data based on the statistic $T_{n,m}$, having density $p_t\left(\theta_{(m)}\right)$, will be consistent. However, the rate of local-alternative against which $\hat{\Lambda}_m$ has power, declines with m.

This potential trade-off is evident in the numerical results presented above this Section. Specifically the test is well sized when m is large - for all sample sizes - but has slightly higher power for smaller values of m. The properties of the density estimator are however rather insensitive to the value of m. Further, the relative computational times for the two cases used in Section 4.3 are important. Using Mathematica 8, on average estimating $\hat{\theta}_{(4)}$ took approximately 14% of the time to estimate $\hat{\theta}_{(10)}$, 0.48s against 3.46s.

Together these findings point to a recommended implementation. For the test, $\hat{\Lambda}_m$ then for the kinds of sample sizes considered here a value, $m^* = 10$ can be chosen and then m need only grow arbitrarily slowly,

$$m = \lfloor m^* \times n^{\epsilon} \rfloor,$$

for any $\epsilon > 0$. Given the significantly greater computational burden for larger values of m then a smaller value of m^* can be chosen to implement the density estimator, say

$m^* = 4$. In the following section we exploit this and apply both test and (recursively through the sample) density estimator to some monthly returns data, in order to showcase their properties and usefulness as new financial econometric tools.

5 Application: Monthly S&P500 Returns

Let P_i denote the end of month price for the S&P500 index, adjusted for both stock splits and dividends, obtained from *https://finance.yahoo.com*. The $n = 240$ observations, from January 1998 to December 2017, for monthly returns $Y_i = \ln(P_i/P_{i-1})$ are collated and graphed in Figure C1 in Appendix C, which contains all graphs pertaining to this application. The standardized correlograms for both returns and squared returns (scaled by $\sqrt{n}$) are graphed in Figure C2, along with significance lines at ± 2. Over this period there is no significant correlation in returns at any lag, while only the first five lags are significant for squared returns.

This particular sample has been chosen because it straddles the global financial crisis of 2007/8. It therefore offers the opportunity to not only apply the procedures developed here for their own sake, but also explore their usefulness as applied financial econometric tools in exposing such phenomena.

5.1 Fitting and testing standard models

On the basis of the correlograms described above, the fitted models considered are all nested within the following (atheoretical) $AR(1) - GARCH(1,1)$ model;

$$\begin{aligned} Y_i &= \gamma_0 + \gamma_1 Y_{i-1} + z_i \quad ; \quad z_i = \sigma_i \varepsilon_i, \\ \sigma_i^2 &= \delta_0 + \delta_1 \varepsilon_{i-1}^2 + \delta_2 \sigma_{i-1}^2, \quad i = 1, .., n. \end{aligned}$$

Denote the unrestricted model M_1. Model M_2 imposes $\delta_2 = 0$, M_3 imposes $\delta_1 = \delta_2 = 0$. Model M_4 imposes $\gamma_1 = 0$, M_5 imposes $\gamma_1 = \delta_2 = 0$ and, lastly, M_6 imposes $\gamma_1 = \delta_1 = \delta_2 = 0$. Thus, M_2 is an $AR(1) - ARCH(1)$, M_3 is a simple $AR(1)$, M_4 and M_5 are constant mean $GARCH(1,1)$ and $ARCH(1)$ while M_6 implies the data is IID.

Two assumptions are made for the idiosyncratic error process $\{\varepsilon_i\}_1^n$, i) $\varepsilon_i \sim iidN(0,1)$ and ii) $\varepsilon_i \sim iid\ \bar{t}_{(v)}$, i.e. the standardized (unit variance) student t distribution, with v degrees of freedom. Parameters are estimated via the (quasi) likelihood procedures outlined in Bollerslev (1986, 1987).

We also consider four sub-samples; a) Jan. 1998 to Dec. 2003, b) Jan. 1998 to Dec. 2007, c) Jan. 2008 to Dec. 2017 and d) Jan. 2013 to Dec. 2017 in addition to the entire data set. First we evaluate $\hat{\Lambda}_{10}$ for all six fitted models across all samples. The asymptotic p-value of the outcome of each test statistic is presented in the tables below:

Table 5.1: Asymptotic p-value of $\hat{\Lambda}_{10}$ across subsamples of S&P500 Returns;

Assumption (i): $\varepsilon_t \sim IID\ N(0,1)$

Start	End	n	M_1	M_2	M_3	M_4	M_5	M_6
01/98	12/02	60	.11385	.18462	.12229	.18957	.22602	.16996
01/98	12/07	120	.09947	.01900	.03074	.06776	.01806	.02853
01/98	12/17	240	.28006	.01601	.00018	.28234	.00542	.00025
01/08	12/17	120	.48300	.64644	.30116	.17346	.22931	.20657
01/13	12/17	60	.89109	.89257	.74563	.72437	.70279	.60518

Table 5.2: Asymptotic p-value of $\hat{\Lambda}_{10}$ across subsamples of S&P500 Returns;

Assumption (ii): $\varepsilon_t \sim IID\ \bar{t}_{(v)}$

Start	End	n	M_1	M_2	M_3	M_4	M_5	M_6
01/98	12/02	60	.09788	.10912	.06599	.06788	.15797	.06259
01/98	12/07	120	.08168	.00807	.03791	.08771	.01875	.04226
01/98	12/17	240	.32482	.02057	.00027	.27923	.00994	.00118
01/08	12/17	120	.50373	.60469	.29824	.20443	.18286	.30114
01/13	12/17	60	.91389	.79154	.70613	.74545	.80098	.75838

In small samples, we cannot reject the specification of any of the models at the 5% level. Although none of the estimated values are presented - these are relatively straight forward to reproduce. In all cases the estimated degrees of freedom for $\bar{t}_v$ were in line with those reported in the previous study of Bollerslev (1987). The standardized t specification does not offer a uniformly better conditional specification, as measured by the asymptotic p-value. Moreover, in no case does it change the outcome of the test at any sensible significance level.

On the basis of the results presented in Tables 5.1 and 5.2, we will narrow the focus and consider two Gaussian models, M_1 which is 'correctly specified' across those (sub)samples, and M_6 which is only 'correct' in small samples. Note that, unlike many other predictive or conditional specification tests those of this paper are asymptotically valid in either a recursive or rolling sampling scheme, provided the rolling window width is asymptotic.

Extending the analysis in Table 5.1 we can recursively evaluate the asymptotic p-value for the $\hat{\Lambda}_{10}$ statistic for the Gaussian versions of M_1 and M_6 based on the samples obtained for $t = 1, .., n_1$ where we allow $n_1 = 60, ..., n$. These are plotted in Figure C3. Similarly we can construct rolling five year windows ($R = 60$) of observations $t = 1 + R, ..., R + n_1$ as $n_1 = 1, .., n - R$ and evaluate the the asymptotic p-value of $\hat{\Lambda}_{10}$ applied to each window of observations. These, for M_1 and M_6 are plotted in Figure C4.

When recursively applied $\hat{\Lambda}_{10}$ is insignificant for both models in the smallest samples but becomes significant for both as the sample expands to include the period prior to the financial crisis. Beyond 2009, however, the GARCH model M_1 is then not rejected, while M_6 is always subsequently rejected. Under a rolling scheme the significance of the tests for M_1 and M_6 are more similar, albeit much more volatile, over the whole sample. They diverge only over the window of observations spanning the mid-2000s, when M_6 is rejected consistently, but M_1 is not.

Broadly speaking M_1 does seem to fit the data over most recursive or rolling samples, while the structureless M_6 does not. Here, we will employ the second function of the nonparametric density estimator to construct both predictive interval and probability estimates in order to gauge the value that doing so will add, whether the first stage specification is 'correct' or not.

5.2 Estimated predictive intervals: recursive and rolling sampling schemes

To proceed, take the first 5 years of observations ($n_1 = 60$), then for the $n_1 + 1^{th}$ observation the estimated conditional mean and standard deviation can be denoted $\hat{\mu}_{n_1+1|\mathcal{F}_{n_1+1}}$and $\hat{\sigma}_{n_1+1|\mathcal{F}_{n_1+1}}$. On the basis of the first n_1 observations the series density estimator applied to the in-sample PITs yields quantiles (on $(0,1)$) which we can denote $\hat{q}_{n_1+1}(\pi)$, for $0 < \pi < 1$. An estimator of the predictive quantiles for the $n_1 + 1^{th}$ observation is then

$$\hat{q}_{Y_{n_1+1}}(\pi) = \hat{\mu}_{n_1+1|\mathcal{F}_{n_1+1}} + \hat{\sigma}_{n_1+1|\mathcal{F}_{n_1+1}} \sqrt{2}\,\mathrm{erf}^{-1}\left(2\hat{q}_{n_1+1}(\pi) - 1\right). \tag{19}$$

Allowing $n_1 = 60, 61, ..., n-1$, then from (19) we can construct the sequence of recursive predictive confidence intervals for Y_{n_1+1}, of the form $\left\{\hat{q}_{Y_{n_1+1}}(\pi), \hat{q}_{Y_{n_1+1}}(1-\pi)\right\}$ having nominal $(1 - 2\pi)$ coverage.

It is to be expected that if the models are mis-specified in accordance with the conditions of Corollary 1 then the resulting predictive intervals should have better coverage than intervals formed from the Gaussian distribution, $N\left(\hat{\mu}_{n_1+1|\mathcal{F}_{n_1+1}}, \hat{\sigma}^2_{i+1|\mathcal{F}_{n_1+1}}\right)$. Denote the average width of the recursive confidence interval obtained by fitting model M_j using quantiles from (19) by $\hat{\omega}_{m,M_j}$ and its actual coverage by $\hat{\alpha}_{m,M_j}$ and the equivalent obtained from Gaussian quantiles by $\tilde{\omega}_{G,M_j}$ and $\tilde{\alpha}_{G,M_j}$. The realized coverage and average width of these intervals are presented in Table 5.3, below.

Table 5.3: Coverage (α) and Width (ω) of Recursive S&P500 CIs

$1-2\pi$	$\hat{\alpha}_{4,M_1}$	$\tilde{\alpha}_{G,M_1}$	$\hat{\omega}_{4,M_1}$	$\tilde{\omega}_{G,M_1}$	$\hat{\alpha}_{4,M_6}$	$\tilde{\alpha}_{G,M_6}$	$\hat{\omega}_{4,M_6}$	$\tilde{\omega}_{G,M_6}$
0.50	.5922	.6424	.0548	.0595	.6201	.6480	.0561	.0635
0.60	.7150	.7263	.0692	.0743	.7460	.7598	.0712	.0793
0.70	.7821	.8045	.0865	.0914	.8044	.8268	.0899	.0977
0.80	.8547	.8715	.1086	.1131	.8604	.8771	.1143	.1208
0.90	.9217	.9330	.1413	.1451	.9106	.9274	.1501	.1550

The previous analysis was based upon a recursive predictive scheme. However, the results in Table 5.1 are suggestive that as the sample size becomes large, simpler models are not correctly specified. Instead, therefore, we might also consider a rolling predictive scheme. As above let $R = 60$ denote a fixed window and let $\mathcal{F}_{n_1+1|R}$ denote the truncated information set at time n_1 given (here) only the R observations Y_{n_1-R+1} to Y_{n_1} and let $\hat{\mu}_{n_1+1|\mathcal{F}_{n_1+1|R}}$ and $\hat{\sigma}_{n_1+1|\mathcal{F}_{n_1+1|R}}$ denote the rolling estimators for the conditional mean and variance of Y_{n_1+1}. Rolling predictive quantiles, say $\hat{q}_{n_1+1|R,m}(\pi)$ are then formed analogously to the method described above. The realized coverage and average width of those, and the other intervals, are presented in Table 5.4, where we employ the additional superscript R to signify the employment of the rolling estimation scheme in their construction.

Table 5.4: Coverage (α) and Width (ω) of Rolling S&P500 CIs

$1-2\pi$	$\hat{\alpha}^R_{4,M_1}$	$\tilde{\alpha}^R_{G,M_1}$	$\hat{\omega}^R_{4,M_1}$	$\tilde{\omega}^R_{G,M_1}$	$\hat{\alpha}^R_{4,M_6}$	$\tilde{\alpha}^R_{G,M_6}$	$\hat{\omega}^R_{4,M_6}$	$\tilde{\omega}^R_{G,M_6}$
0.50	.5474	.5754	.0555	.0570	.5698	.5922	.0493	.0579
0.60	.6481	.6481	.0711	.0714	.6759	.7430	.0645	.0724
0.70	.7653	.7653	.0891	.0886	.7709	.7878	.0825	.0891
0.80	.8436	.8492	.1084	.1106	.8324	.8547	.1049	.1102
0.90	.8994	.8994	.1391	.1415	.8927	.8883	.1371	.1414

Across tables 5.3 and 5.4, the Gaussian predictive confidence intervals are both wider and have poorer coverage. This is true even if the Gaussian model is correctly specified, according to the p-value of $\hat{\Lambda}_{10}$. Employing a richer model, i.e. M_1 rather than M_6 and also employing a rolling rather than recursive estimation scheme both lead to narrower and more accurate predictive intervals.

This general finding is borne out in the average predictive log-scores for both the fitted Gaussian model and the non-parametric density estimator applied to the recursive, then rolling out-of-sample PITs. These are presented in Table 5.5:

Table 5.5: Realized average log-scores,

Model	M_1	M_1	M_6	M_6
	$\ln f_{Y_{n_1+1}}\left(\hat{\beta}_{n_1}\right)$	$\ln p_{\hat{X}_{n_1+1}}\left(\theta_{(4)}\right)$	$\ln f_{Y_{n_1+1}}\left(\hat{\beta}_{n_1}\right)$	$\ln p_{\hat{X}_{n_1+1}}\left(\theta_{(4)}\right)$
Recursive	1.95807	0.05200	1.78688	0.06486
Rolling	1.96761	0.02895	1.75993	0.03572

Both in terms of the interval prediction and evaluation of the probability forecast as a whole, the contribution of the second stage series density estimator is positive. Although it is difficult to draw general conclusions it seems clear that, in practice, assuming Gaussianity implies less certain inference. This would then have significance for practitioners using such to evaluate the riskiness of any investment in this index.

5.2.1 Value at Risk and the Probability of Negative Returns

To that end, one obvious application of a method which provides consistent estimation of the quantiles of the distribution of a financial variable is to calculate the value at risk. See Duffie and Pan (1999) for an overview. Denoting the π-quantile of returns for time n_1+1 by $q_{n_1+1}(\pi)$, then the VaR at time n_1, say VaR_{π,n_1+1}, is

$$VaR_{\pi,n_1+1} = (1 - e^{q_{n_1+1}(\pi)})P_{n_1},$$

where P_{n_1} is the value of the asset at time n_1.

Here P_{n_1} will be the S&P500 price at time n_1, and we can estimate the quantiles via either $\hat{q}_{n_1+1,4}(\pi)$ or $\hat{q}_{n_1+1|R,4}(\pi)$ for the recursive and rolling schemes and compare with that obtained from Gaussian quantiles, $\hat{q}_{n_1+1,G}(\pi)$ or $\hat{q}_{n_1+1|R,G}(\pi)$. Figure C4 plots the relative difference between the predictive VaRs obtained from the estimated and Gaussian quantiles, i.e.

$$\hat{\upsilon}_{n_1} = \frac{(1 - e^{\hat{q}_{n_1+1,4}(\pi)})}{(1 - e^{\hat{q}_{n_1+1,G}(\pi)})} \quad \& \quad \hat{\upsilon}_{n_1|R} = \frac{(1 - e^{\hat{q}_{n_1+1|R,4}(\pi)})}{(1 - e^{\hat{q}_{n_1+1|R,G}(\pi)})}$$

It is seen above that the second nonparametric stage is able to correct mis-specified predictive quantiles (as in Table B4) and offers superior predictive intervals (as in

Tables 5.3 and 5.4). Consequently significant deviation from one of $\hat{\upsilon}_{n_1}$ implies that performing VaR calculations using Gaussian quantiles will likely be inaccurate.

In figure C4 the relative quantiles are plotted for M_1 (solid line) and M_6 (dashed) and both Recursive (Fig. C4a) and Rolling (Fig. C4b) estimation schemes. Naturally one would expect the relative quantiles (nonparametric to Gaussian), and hence VaR, to diverge for the simpler, less well-specified model. Although the divergence is smaller for the better specified M_6 it remains present throughout the sample and correlates well to the p-values of the $\hat{\Lambda}_{10}$ test, particularly when applied recursively.

Finally, as well as producing predictive quantiles, the density estimator $p_t\left(\hat{\theta}_{(4)}\right)$ can be used to produce predictive probabilities for particular (sets of) outcomes of the returns. Here we will predict the probability of a negative return. To do so, let

$$\hat{\pi}^0_{n_1+1,G} = \int_{-\infty}^{0} \Xi\left(\hat{\mu}_{n_1+1|\mathcal{F}_{n_1}}, \hat{\sigma}^2_{n_1+1|\mathcal{F}_{n_1}}, y\right) dy,$$

where $\Xi\left(\mu, \sigma^2, y\right)$ denotes the CDF of a $N\left(\mu, \sigma^2\right)$ random variable, evaluated at y. Thus $\hat{\pi}^0_{n_1+1,G}$ represents the Gaussian estimator, at time n_1, of the probability of a negative outcome for $Y_{n_{1+1}}$.

The nonparametric estimator for this quantity is instead,

$$\hat{\pi}^0_{n_1+1,4} = \int_{0}^{\hat{\pi}^0_{n_1+1,G}} p_t\left(\hat{\theta}_{(4)}\right) dt.$$

Again these probability estimators can be recursively constructed through the sample as $n_1 = 60, .., 239$ and are plotted in Figure C5. Figure C5a plots the percentage $\left(\hat{\pi}^0_{n_1+1,m}\right) / \hat{\pi}^0_{n_1+1,G}$, constructed from both M_1 and M_6. As was done for predictive VaR, we can also construct two rolling sequences of predicted probabilities of negative returns. Doing so, then the percentage difference between them, again for both models M_1 and M_6, is plotted in Figure C5b.

Taking Figures C5a and C5b together it is clear that there is a divergence of predictive intervals and probabilities obtained from a Gaussian model and those obtained by correcting via the nonparametric density estimator. This becomes particularly pronounced around 2008 onwards. That is, uncorrected Gaussian models would imply a significantly different assessment of risk than the corrected predictive model would have.

6 Conclusions

This paper has extended the nonparametric likelihood ratio based tests introduced in Marsh (2007) to cover specifications involving estimated parameters in the context of conditional, dynamic models. Doing so yields a straight forward two stage process. In the context of standard time series models this allows fitting of a parametric model, allied to a specified error distribution, in the first stage, with the nonparametric estimation of the density function of the PITs, in the second.

The second stage provides tests which help overcome the three potential shortcomings of EDF based tests, i.e. that they are not pivotal, have low power, and offer little direction in case of rejection. Instead the tests of this paper are shown to be asymptotically standard normal, there is good evidence that they have a power advantages, particularly over the basis for the most conditional specification tests, the Kolmogorov-Smirnov. In the event of rejection these results can be used to improve the predictive ability of the original fitted model, through consistent quantile estimation or via an improvement of a loss function such as the log-score.

The procedure of this paper is seen to have relevance in empirical financial research. Applied to monthly S&P500 data, the usefulness of being able to correct the quantiles of a predictive density are manifest. This is particularly true for the management of risk - both Value at Risk calculations and estimators for the probability of a negative return - based on the estimator begin to diverge significantly from assumed Gaussian equivalents, around and from the time of the financial crisis of 2007/8.

Acknowledgements: This work has greatly benefited from comments received from Francesco Bravo, Giovanni Forchini, Les Godfrey, David Harvey, Lorenzo Trapani and Peter Phillips as participants at seminars given at the Universities of Southampton, York, Nottingham, Surrey and the Tinbergen Institute, Amsterdam.

REFERENCES

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097-1128.

Bai, J. (2003) Testing parametric conditional distributions of dynamic Models, *Review of Economics and Statistics* 85, 531-549.

Barron, A.R. and C-H. Sheu (1991) Approximation of density functions by sequences of exponential families. *Annals of Statistics* 19, 1347-1369.

Barndorff-Nielsen O. (1978) *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *'Journal of Econometrics* 31, 307-327.

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542-547.

Bontemps, C. and N. Meddahi (2012) Testing distributional assumptions: a GMM approach. *Journal of Applied Econometrics* 27, 978-1012.

Chen, Y.-T. (2012) A simple approach to standardized-residuals-based higher-moment tests. *Journal of Empirical Finance* 19, 427-453.

Claeskens, G. and N.L. Hjort (2004) Goodness of fit via nonparametric likelihood ratios. *Scandinavian Journal of Statistics* 31, 487-513.

Corradi, V. and N. R. Swanson (2006a) Predictive Density Evaluation. In G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting,* Vol. 1, pp. 197-284, Elsevier.

Corradi, V. and N. R. Swanson (2006b) Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779-806.

Corradi, V. and N. R. Swanson (2012), "A survey of recent advances in forecast accuracy comparison testing, with an extension to stochastic dominance," in X. C. Chen and N. R. Swanson (eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr.* Berlin: Springer.

Crain, B.R. (1974), "Estimation of distributions using orthogonal expansions," *Annals of Statistics* 2, 454–463.

Diebold, F.X., T.A. Gunther and A.S. Tay (1998), "Evaluating density forecasts with applications to financial risk management," *International Economic Review* 39, 863-83.

Domowitz, I and H. White (1982) Misspecified models with dependent observations. *Journal of Econometrics* 20, 35-58.

Duffie, D. and J. Pan (1999) An Overview of Value at Risk. *Journal of Derivatives* 4, 749.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.

Gneiting, T. and A.E. Raftery (2007) Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association* 102, 359-378.

Giacomini, R. and H. White (2006) Tests of conditional predictive ability. *Econometrica* 74, 1545-1578.

Kallenberg, W.C. and T. Ledwina (1997) Data-driven smooth tests when the hypothesis is composite. *Journal of the American Statistical Association* 92, 1094-1104.

Ledwina, T. (1994) Data-driven version of the Neyman smooth test of fit. *Journal of the American Statistical Association* 89, 1000-1005.

Marsh, P. (2007) Goodness of fit tests via exponential series density estimation. *Computational Statistics and Data Analysis* 51, 2428–2441.

Marsh, P. (2018) Nonparametric series density testing and estimation. *Statistical Methods and Applications*, forthcoming.

Portnoy, S. (1988) Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Annals of Statistics* 16, 356-366.

Randles, R. (1984) On tests applied to residuals. *Journal of the American Statistical Association* 79, 349-354.

Rossi, B. and T. Sekhposyan (2015) Alternative tests for correct specification of conditional predictive densities. Working paper, ICREA - Universität Pompeu Fabra.

Stephens, M.A. (1976) Asymptotic results for goodness of fit statistics with unknown parameters. *Annals of Statistics* 4, 357–369.

APPENDIX A (Proofs of theorems and corollary)

In order to avoid any ambiguity throughout this appendix the order of magnitude symbol $O(.)$ is defined by,

$$a_{n,m} = O\left(b_{n,m}\right) \Longleftrightarrow \lim_{m,n\to\infty\quad ;\quad m^3/n\to 0} \frac{a_{n,m}}{b_{n,m}} \leq c_1 < \infty,$$

and analogously for the probabilistic versions $O_p(.)$ and $o_p(.)$. If the quantity under scrutiny does not depend upon the dimension m then the condition $m^3/n \to 0$ becomes redundant.

Proof of Theorem 1:

Convergence of the density estimator is established by showing that the effect of having to estimate unknown parameters is asymptotically negligible, under Assumption 1. To proceed denote the following two $m \times 1$ vectors;

$$\begin{aligned}
\hat{x}_{(m)} &= n^{-1}\sum_{i=1}^{n}\left(\cos\left(2\pi k\hat{X}_i\right), \sin\left(2\pi k\hat{X}_i\right), .., \cos\left(\pi m\hat{X}_i\right), \sin\left(\pi m\hat{X}_i\right)\right)' \quad \text{and} \\
\bar{x}_{(m)} &= n^{-1}\sum_{i=1}^{n}\left(\cos\left(2\pi kX_i\right), \sin\left(2\pi kX_i\right), .., \cos\left(\pi mX_i\right), \sin\left(\pi mX_i\right)\right)',
\end{aligned}$$

and let

$$E\left(\cos\left[2\pi kX_i\right]\right) = \mu_{k,C} \quad \text{and} \quad E\left(\sin\left[2\pi kX_i\right]\right) = \mu_{k,S},$$

for all i, so that we can write the mean of $\bar{x}_{(m)}$ as

$$\mu_{(m)}^{P} = E\left[\bar{x}_{(m)}^{P}\right] = \left(\mu_{1,C}, \mu_{1,S}, \ldots.., \mu_{m/2,C}, \mu_{m/2,S}\right)'.$$

Notice that since $\cos(.)$ and $\sin(.)$ are bounded functions then both $\mu_{k,C}$ and $\mu_{k,S}$ are bounded for all k,and as $k \to \infty$.

The Euclidean distance between $\hat{x}_{(m)}$ and $\bar{x}_{(m)}$ satisfies

$$\left|\hat{x}_{(m)} - \bar{x}_{(m)}\right| = \sqrt{\sum_{k=1}^{m/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{C}_{k,i} - C_{k,i}\right)\right)^2 + \sum_{k=1}^{m/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{S}_{k,i} - S_{k,i}\right)\right)^2},$$

where we have denoted, $\hat{C}_{k,i} = \cos\left(2\pi k\hat{X}_i\right)$, $C_{k,i} = \cos\left(2\pi kX_i\right)$, $\hat{S}_{k,i} = \sin\left(2\pi k\hat{X}_i\right)$ and $S_{k,i} = \sin\left(2\pi kX_i\right)$

Since both $\cos\left(2\pi kx\right)$ and $\sin\left(2\pi kx\right)$ are also continuously differentiable in x and $\hat{X}_i - X_i = e_i = O_p\left(n^{-1/2}\right)$ with $|e_i| \leq 1$, then $e_i^j = O_p\left(n^{-j/2}\right)$ and, for any k, expansion of $\hat{C}_{k,i}$ (and $\hat{S}_{k,i}$) around X_i yields,

$$\hat{C}_{k,i} - C_{k,i} = O_p\left(n^{-1/2}\right),$$

and similar for $\hat{S}_{k,i}$. Boundedness of the sin and cosine functions implies

$$\frac{1}{n}\sum_{i=1}^{n}\hat{C}_{k,i} - \frac{1}{n}\sum_{i=1}^{n}C_{k,i} = O_p\left(n^{-1/2}\right),$$

also. Defining,

$$d_C^* = \sup_{k\in Z^+}\frac{1}{n}\sum_{i=1}^{n}\hat{C}_{k,i} - \frac{1}{n}\sum_{i=1}^{n}C_{k,i} \quad \& \quad d_S^* = \sup_{k\in Z^+}\frac{1}{n}\sum_{i=1}^{n}\hat{S}_{k,i} - \frac{1}{n}\sum_{i=1}^{n}S_{k,i},$$

then both d_C^* and d_S^* are $O_p\left(n^{-1/2}\right)$ and so

$$\left|\hat{x}_{(m)} - \bar{x}_{(m)}\right| \leq \sqrt{\sum_{k=1}^{m/2}d_C^{*2} + \sum_{k=1}^{m/2}d_S^{*2}} = O_p\left(\sqrt{\frac{m}{n}}\right). \tag{A.1}$$

Now consider the Euclidean distance between $\hat{x}_{(m)}$ and $\mu_{(m)}$ which, via the triangle inequality, satisfies,

$$\left|\hat{x}_{(m)} - \mu_{(m)}\right| \leq \left|\hat{x}_{(m)} - \bar{x}_{(m)}\right| + \left|\bar{x}_{(m)} - \mu_{(m)}\right|. \tag{A.2}$$

The first term in (A.2) is $O_p\left(\sqrt{\frac{m}{n}}\right)$. For the second, we have

$$\left|\bar{x}_{(m)} - \mu_{(m)}\right| \leq \sqrt{\sum_{k=1}^{m/2}\left(\frac{\sum_{i=1}^{n}C_{k,i}}{n} - \mu_{k,c}^T\right)^2 + \sqrt{\sum_{k=1}^{m/2}\left(\frac{\sum_{i=1}^{n}S_{k,i}}{n} - \mu_{k,c}^T\right)^2}}$$

Again the boundedness of $\cos(.)$ and $\sin(.)$ implies that both $\mu_{k,C}$ and $\mu_{k,S}$ are bounded for all $k \in \mathbb{Z}^+$, and thus a standard Weak Law of Large Numbers applies;

$$\frac{\sum_{i=1}^{n} C_{k,i}}{n} - \mu_{k,c}^{T} = O_p\left(n^{-1/2}\right) \quad \& \quad \frac{\sum_{i=1}^{n} S_{k,i}}{n} - \mu_{k,s}^{T} = O_p\left(n^{-1/2}\right). \tag{A.3}$$

As a consequence we find, $\left|\bar{x}_{(m)} - \mu_{(m)}\right| = O_p\left(\sqrt{\frac{m}{n}}\right)$ as $m, n \to \infty$, also. Together these results imply that,

$$\left|\hat{x}_{(m)} - \mu_{(m)}\right| \leq \left|\bar{x}_{(m)} - \mu_{(m)}\right| + \left|\hat{x}_{(m)} - \bar{x}_{(m)}\right| = O_p\left(\sqrt{\frac{m}{n}}\right), \tag{A.4}$$

which follows from (A.1) and noting the same order of magnitude applies for the first distance, as in equation 6.5 of Barron and Sheu (1991), so that the order of magnitude of $\left|\hat{x}_{(m)} - \mu_{(m)}\right|$ is the same as that of $\left|\hat{x}_{(m)} - \bar{x}_{(m)}\right|$. Thus, asymptotically, the effect of having to estimate β is negligible, under Assumption 1.

Extending the decomposition of the Kullback-Leibler divergence of Barron and Sheu (1991, eq. 6.9) we have,

$$\begin{aligned} E_U\left[\ln\left(\frac{u(x)}{p_x(\hat{\theta}_{(m)})}\right)\right] &= E_U\left[\ln\left(\frac{u(x)}{p_x(\theta_{(m)})}\right)\right] + E_U\left[\ln\left(\frac{p_x(\theta_{(m)})}{p_x(\bar{\theta}_{(m)})}\right)\right] \\ &\quad + E_U\left[\ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right)\right]. \end{aligned} \tag{A.5}$$

Since here we choose $u(x) = 1$ (the X_i are IID Uniform under correct specification) and $\theta_{(m)} = 0_{(m)}$, then this immediately simplifies to

$$E_U\left[\ln\left(\frac{1}{p_x(\hat{\theta}_{(m)})}\right)\right] = E_U\left[\ln\left(\frac{p_x(0_{(m)})}{p_x(\bar{\theta}_{(m)})}\right)\right] + E_U\left[\ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right)\right]$$

By construction $\log[u(x)] = 0 \in W_2^{\infty}$ then from Barron and Sheu (1991, Theorem 1) the first two terms in (A.5) are, respectively, $lim_{r\to\infty} O(m^{-2r})$ and $O_p(m/n)$. For the third term, application of the last part of Lemma 5 of Barron and Sheu (1991), which holds for any two values in $\mathcal{T}_m \subset \mathbb{R}^m$, here uniquely defined by equations (5) and (9), implies that

$$E_U\left[\ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right)\right] = \int_0^1 \ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right) u(x)dx = O_p\left(\left|\hat{x}_{(m)} - \bar{x}_{(m)}\right|^2\right) = O_p\left(\frac{m}{n}\right),$$

and hence

$$E_U\left[\ln\left(\frac{1}{p_x(\hat{\theta}_{(m)})}\right)\right] = O_p\left(\frac{m}{n}\right),$$

as required. ■

Proof of Theorem 2:

Consider the problem of testing $H_0 : \theta_{(m)} = 0_{(m)}$ against the alternative $H_1 : \theta_{(m)} \neq 0_{(m)}$ when $n, m \to \infty$ and $m^3/n \to 0$.

Part (i): To proceed we have defined,

$$\hat{\lambda}_m = 2n\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} - \left(\psi_m\left(\hat{\theta}_{(m)}\right) - \psi_m\left(0_{(m)}\right)\right)\right] = 2n\left(\hat{\theta}_{(m)} - 0_{(m)}\right)\hat{x}_{(m)},$$

where $\hat{\theta}_{(m)}$ solves (9), or equivalently,

$$\psi_m'\left(\hat{\theta}_{(m)}\right) = \left.\frac{\partial\psi_m\left(\theta_{(m)}\right)}{\partial\theta_{(m)}}\right|_{\theta_{(m)}=\hat{\theta}_{(m)}} = \hat{x}_{(m)}.$$

Since $\psi_m(.)$ is the cumulant function then the value $\theta^0_{(m)} = 0_{(m)}$ can be defined by,

$$\psi_m'\left(\theta_{(m)}\right)\big|_{\theta_{(m)}=0_{(m)}} = \mu_{(m)} = E(\bar{x}_{(m)}).$$

Since the exponential log-likelihood is strictly convex, the mapping

$$\theta_{(m)}(\eta) : \psi_m'\left(\theta_{(m)}\right) = \mu_{(m)}$$

is one-to-one between the parameter space $\Theta_m \subset \mathbb{R}^m$ and sample space $\mathcal{F}_m \subset \mathbb{R}^m$ and application of Barron and Sheu (1991, eq. 5.6) and also (A.4) gives,

$$O_p\left(\left|\hat{\theta}_{(m)} - 0_{(m)}\right|\right) = O_p\left(\left|\hat{x}_{(m)} - \mu_{(m)}\right|\right) = O_p\left(\sqrt{\frac{m}{n}}\right). \tag{A.6}$$

As a consequence of both (A.6) and (A.4) we have that,

$$O_p\left(\left|\hat{\theta}_{(m)} - 0_{(m)}\right|\right) = O_p\left(\left|\bar{\theta}_{(m)} - 0_{(m)}\right|\right) \quad \& \quad O_p\left(\left|\hat{x}_{(m)} - \mu_{(m)}\right|\right) = O_p\left(\left|\bar{x}_{(m)} - \mu_{(m)}\right|\right),$$

and note that the expansions given in the proofs of Theorems 3.1 and 3.2 of Portnoy (1988) apply for any two pairs of values, here $\left(\bar{\theta}_{(m)}, 0_{(m)}\right)$ and $\left(\bar{x}_{(m)}, \mu_{(m)}\right)$.

To continue, denote expectations under the null hypothesis as $E_{0_{(m)}}[.]$ and let the $m \times 1$ vector $U_{\theta_{(m)}}$ have density function

$$p_{\theta_{(m)}}(u_1, .., u_m) = \exp\left\{\sum_{k=1}^{m} \theta_k u_k - \psi_m\left(\theta_{(m)}\right)\right\}.$$

Analagous to Portnoy (1988, eq. 3.5 and 3.6), we have the following two expansions;

$$|\hat{\theta}_{(m)} - 0_{(m)}|^2 = \left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} - \frac{1}{2} E_{0_{(m)}}\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' U_{(m)}\right]^2 + O_p\left(\frac{m^2}{n^2}\right), \tag{A.7}$$

and

$$\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} = |\hat{x}_{(m)}|^2 - \frac{1}{2} E_{0_{(m)}}\left[\left(\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' U_{(m)}\right)^2 \hat{x}'_{(m)} U_{(m)}\right] + O_p\left(\frac{m^2}{n^2}\right). \tag{A.8}$$

Subtracting (A.8) from (A.7) restates equation (3.7) of Portnoy (1988) and hence

$$|\hat{\theta}_{(m)} - 0_{(m)} - \hat{x}_{(m)}| = O_p\left(\frac{m}{n}\right).$$

From the definition of the likelihood ratio test we therefore have,

$$\begin{aligned} \hat{\lambda}_m &= 2n\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} - \left(\psi_m\left(\hat{\theta}_{(m)}\right) - \psi_m\left(\theta^0_{(m)}\right)\right)\right] \\ &= n\left[|\hat{x}_{(m)}|^2 - |\hat{\theta}_{(m)} - 0_{(m)} - \hat{x}_{(m)}|^2 + \frac{1}{6} E_{U_0}\left(\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' U\right)^3\right] + O_p\left(\frac{m^2}{n}\right), \end{aligned} \tag{A.9}$$

as in Portnoy (1988, eq. 3.12). Let $\bar{e} = \hat{x}_{(m)} - \bar{x}_{(m)}$, then from the proof of Theorem 1, we have

$$|\hat{x}_{(m)}|^2 = |\bar{x}_{(m)} + \bar{e}|^2 \leq |\bar{x}_{(m)}|^2 + |\bar{e}|^2 = |\bar{x}_{(m)}|^2 + O_p\left(\frac{m}{n}\right). \tag{A.10}$$

Note that for the given trigonometric basis we have $E\left[\bar{x}_{(m)}\right] = 0_{(m)}$. Now define the $m \times 1$ random variable $V_{0_{(m)}} = \psi''_m\left(0_{(m)}\right)^{-1/2}(\bar{x})$, and denote its density $p_{\theta^V_{(m)}}\left(v_{(m)}\right)$, so that $E[V_m] = 0_{(m)}$ and $Var[V_m] = I_m$. Since the likelihood ratio statistic is parameterization invariant the likelihood ratio test based on observations on V_m will be identical to that based on $\bar{x}_{(m)}$.

Rather than defining a new triple of values, analogous to those in (13), in both the parameter space Θ_m (note that in particular the hypothesized value would no longer satisfy $\theta^0_{(m)} = 0_{(m)}$) and sample space $\mathcal{F}_m$ we will instead, and without any loss of generality assume a parameterization in which both $E\left[\bar{x}_{(m)}\right] = 0$ and now also $V\left[\bar{x}_{(m)}\right] = I_m$. Note, however, that it is the unobserved $\bar{x}_{(m)}$ which is assumed to be standardized not the observed $\hat{x}_{(m)}$.In this parameterization the asymptotic distribution of first $|\bar{x}_{(m)}|^2$ and hence $|\hat{x}_{(m)}|^2$ (via (A.10)) and then via (A.9) for $\hat{\Lambda}_m = \frac{\hat{\lambda}_m - m}{\sqrt{2m}}$ follows exactly as in Portnoy (1988, Theorem 4.1). ■

Part (ii): Under any fixed alternative the density of $X_i = F_i\left(Y_i|\mathcal{F}_i;\beta_*\right)$ is not uniform, nor even independent or identically distributed. However, since $\phi_k(x)$ is bounded then even under H_1,

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n \phi_k\left(\hat{X}_i\right)}{n} = \mu_k \leq M < \infty.$$

Consequently, let $\theta^1_{(m)}$ be the unique solution to,

$$\int_0^1 \phi_k(x)\, p_x\left(\theta^1_{(m)}\right) dx = \mu_k, \quad k = 1,..,m. \tag{A.11}$$

The uniqueness of solutions to (A.11) imply $\theta^1_{(m)} \neq 0_{(m)}$.

To take the least favorable case, define

$$\theta^1_{(m)} = \left(\theta^1_1, \theta^1_2, .., \theta^1_m\right)'$$

and suppose that $\theta^1_k \neq 0$ for some finite k but that $\theta^1_j = 0$ for all $j \neq k$. The series density estimator is consistent for $\theta^1_{(m)}$, under H_1, in that $\left|\hat{\theta}_{(m)} - \theta^1_{(m)}\right| = O_p\left(\sqrt{\frac{m}{n}}\right)$, analogous to (A.6) above, and so we can write,

$$n\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} = n\left[\left(\hat{\theta}_{(m)} - \theta^1_{(m)}\right)' \hat{x}_{(m)} + \left(\theta^1_k\right) \frac{1}{n} \sum_{i=1}^n \phi_k\left(\hat{X}_i\right)\right].$$

Since also $\psi_m\left(0_{(m)}\right)=0$, we can write the likelihood ratio as

$$\begin{aligned}
\hat{\lambda}_m &= 2n\left[\left(\hat{\theta}_{(m)}-0_{(m)}\right)'\hat{x}_{(m)}-\psi_m\left(\hat{\theta}_{(m)}\right)\right] \\
&= 2n\left[\left(\hat{\theta}_{(m)}-\theta^1_{(m)}\right)'\hat{x}_{(m)}-\left(\psi_m\left(\hat{\theta}_{(m)}\right)-\psi_m\left(\theta^1_{(m)}\right)\right)\right] \\
&\quad +2n\left[\left(\theta^1_k-\theta^0_k\right)\frac{1}{n}\sum_{i=1}^{n}\phi_k\left(\hat{X}_i\right)-\psi_m\left(\theta^1_{(m)}\right)\right] \\
&= \hat{\lambda}^1_m+2n\left[\left(\theta^1_k-\theta^0_k\right)\frac{1}{n}\sum_{i=1}^{n}\phi_k\left(\hat{X}_i\right)-\psi_m\left(\theta^1_{(m)}\right)\right],
\end{aligned}$$

where $\hat{\lambda}^1_m$ is the likelihood ratio for testing $H_1:\theta_{(m)}=\theta^1_{(m)}$.

Thus, under H_1, we can write

$$\hat{\Lambda}_m=\frac{\hat{\lambda}_m-m}{\sqrt{2m}}=\frac{\hat{\lambda}^1_m-m}{\sqrt{2m}}+\frac{2n\left[\left(\theta^1_k-\theta^0_k\right)\frac{1}{n}\sum_{i=1}^{n}\phi_k\left(\hat{X}_i\right)-\psi_m\left(\theta^1_{(m)}\right)\right]}{\sqrt{2m}}.$$

Immediate from Part (i) of this theorem is that as $m,n\to\infty$, with $m^3/n\to 0$,

$$\frac{\hat{\lambda}^1_m-m}{\sqrt{2m}}\to_d N(0,1),$$

i.e. $\left(\hat{\lambda}^1_m-m\right)/\sqrt{2m}$ is $O_p(1)$. However, since $\psi_m(.)$ is a uniquely defined cumulant function then

$$\psi_m\left(\theta^1_{(m)}\right)\neq 0,$$

while $\frac{1}{n}\sum_{i=1}^{n}\phi_k\left(\hat{X}_i\right)=O_p(1)$ and since $m^3/n\to 0$,

$$\hat{\Lambda}_m=O_p(1)+O_p\left(\frac{n}{\sqrt{m}}\right)\to\infty,$$

and hence $\Pr\left[\hat{\Lambda}_m>\kappa\right]\to 1$, as required. ■

Proof of Corollary 1:

Under Assumptions 1 and 2, we immediately find

$$\sup_{i,\beta\in\mathcal{B}}\left|\varepsilon_i\left(\hat{\beta}_n\right)-\varepsilon_i(\beta)\right|=O_p\left(n^{-1/2}\right),$$

and since $F(e)$ is continuously differentiable and monotone then, similar to the proof of Theorem 1, we have

$$\hat{X}_i=X_i+\hat{e}_i,$$

where $0 < \hat{e}_i < 1$ and $\hat{e}_i = O_p\left(n^{-1/2}\right)$ and $X_i \sim IID\, X$. Let the density of X be $u_1(x)$ and define the unique $m \times 1$ vector $\theta^1_{(m)}$ by

$$\int_0^1 \phi_k(x)\ p_x\left(\theta^1_{(m)}\right) dx = \int_0^1 \phi_k(x)\, u_1(x)\, dx. \tag{A.12}$$

Similar to the proof of Theorem 1 we can decompose the Kullback-Leibler divergence between $p_x(\hat{\theta}_{(m)})$ and $u_1(x)$, with

$$\begin{aligned} E_U\left[\ln\left(\frac{u_1(x)}{p_x(\hat{\theta}_{(m)})}\right)\right] &= E_U\left[\ln\left(\frac{u_1(x)}{p_x(\theta^1_{(m)})}\right)\right] + E_U\left[\ln\left(\frac{p_x(\theta^1_{(m)})}{p_x(\bar{\theta}_{(m)})}\right)\right] \\ &\quad + E_U\left[\ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right)\right], \end{aligned}$$

where $\bar{\theta}_{(m)}$ is defined in (13). Notice that the approximation error represented by the first term does not vanish in this case since, in general $u_1(x) \neq 1$. None-the-less, from Barron and Sheu (1991) and the proof of Theorem 1, we have

$$E_U\left[\ln\left(\frac{u_1(x)}{p_x(\hat{\theta}_{(m)})}\right)\right] = O\left(m^{-2r}\right) + O_p\left(\frac{m}{n}\right) + O_p\left(\frac{m}{n}\right) = O_p\left(m^{-2r} + \frac{m}{n}\right),$$

i.e. $p_x(\hat{\theta}_{(m)})$ converges in relative entropy to $u_1(x)$ which is sufficient for convergence in law, so $\hat{T}_{m,n} \rightarrow_{\mathcal{L}} X$, as required. $\blacksquare$

APPENDIX B (Tables of outcomes of Monte Carlo simulations)

Table B1: Sizes of tests for both H_0^E and H_0^N.

n	25			25			50			50		
	H_0^E			H_0^N			H_0^E			H_0^N		
α / m	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
2	.053	.033	.017	.012	.007	.002	.059	.040	.017	.012	.005	.001
4	.071	.042	.020	.039	.023	.011	.067	.044	.021	.042	.029	.009
6	.073	.037	.007	.046	.026	.011	.072	.045	.016	.045	.027	.009
8	.081	.056	.020	.054	.027	.004	.077	.047	.020	.056	.031	.008
10	.092	.057	.018	.058	.030	.005	.096	.055	.020	.065	.037	.009
12	.091	.055	.014	.065	.034	.005	.094	.055	.018	.063	.038	.007

n	100			100			200			200		
	H_0^E			H_0^N			H_0^E			H_0^N		
α / m	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
2	.055	.037	.017	.013	.006	.001	.065	.038	.020	.012	.006	.002
4	.070	.042	.008	.050	.035	.009	.070	.041	.015	.049	.031	.010
6	.088	.051	.016	.063	.041	.011	.085	.047	.017	.065	.041	.011
8	.093	.055	.016	.078	.045	.013	.093	.048	.016	.077	.043	.010
10	.098	.054	.014	.080	.047	.015	.098	.055	.012	.085	.045	.017
12	.096	.053	.014	.092	.045	.018	.096	.052	.013	.094	.046	.016

Table B2: Sizes of tests for both $H_0^{AR(1)}$ and H_0^{PR}.

n	25			25			50			50		
	H_0^{AR1}			H_0^{PR}			H_0^{AR1}			H_0^{PR}		
α / m	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
2	.044	.028	.009	.028	.017	.006	.044	.028	.014	.029	.016	.006
4	.063	.035	.014	.045	.026	.011	.066	.044	.017	.047	.031	.013
6	.079	.053	.019	.061	.037	.013	.079	.046	.015	.056	.038	.017
8	.088	.061	.023	.082	.057	.016	.079	.047	.016	.065	.039	.017
10	.104	.066	.026	.121	.084	.031	.083	.055	.019	.065	.042	.013
12	.123	.081	.030	.128	.087	.038	.083	.055	.017	.074	.044	.014

n	100			100			200			200		
	H_0^{AR1}			H_0^{PR}			H_0^{AR1}			H_0^{PR}		
α / m	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
2	.049	.029	.011	.031	.019	.006	.052	.033	.015	.029	.017	.007
4	.064	.043	.015	.049	.030	.013	.063	.041	.015	.053	.033	.012
6	.075	.046	.020	.061	.038	.015	.069	.044	.018	.060	.033	.010
8	.072	.042	.015	.066	.036	.013	.082	.046	.019	.069	.043	.014
10	.078	.045	.018	.071	.042	.014	.083	.045	.014	.082	.046	.013
12	.087	.045	.013	.086	.047	.015	.095	.049	.013	.093	.048	.011

Table B3: Rejection frequencies under various alternatives.

Table B3a: Power $H_0 : Y \sim N(0,1)$ vs. $H_1 : Y \sim t_{(v)}$.

v	4	6	8	10	12	4	6	8	10	12
$\hat{\Lambda}_4$	.923	.710	.388	.260	.118	.601	.297	.158	.130	.099
$\hat{\Lambda}_{10}$	.854	.569	.257	.165	.087	.494	.241	.133	.111	.081
KS	.614	.206	.091	.055	.049	.217	.114	.075	.059	.052
CvM	.722	.309	.165	.092	.061	.296	.132	.087	.075	.066

Table B3b: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim \chi^2_{(v)} - v$.

v	12	20	28	36	44	12	20	28	36	44
$\hat{\Lambda}_4$	.855	.664	.571	.469	.417	.574	.269	.185	.151	.111
$\hat{\Lambda}_{10}$	.796	.641	.546	.427	.377	.388	.189	.158	.111	.096
KS	.717	.568	.443	.388	.350	.238	.151	.106	.093	.075
CvM	.837	.663	.563	.463	.403	.274	.176	.131	.100	.091

Table B3c: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim N(vY_{i-1}, 1)$.

v	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_4$	.701	.588	.385	.163	.090	.912	.736	.507	.274	.099
$\hat{\Lambda}_{10}$	.688	.483	.351	.141	.071	.847	.683	.461	.235	.091
KS	.592	.458	.254	.091	.053	.579	.359	.207	.122	.058
CvM	.690	.585	.362	.140	.066	.648	.448	.273	.162	.083

Table B3d: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim N\left(0, 1 + vY_{i-1}^2\right)$.

v	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_4$	.730	.519	.278	.120	.079	.871	.739	.483	.228	.106
$\hat{\Lambda}_{10}$	.704	.503	.263	.113	.074	.864	.740	.460	.225	.094
KS	.568	.361	.161	.063	.052	.509	.350	.201	.112	.080
CvM	.709	.497	.255	.109	.075	.511	.352	.185	.115	.073

Table B3e: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim N\left(v 1_{t>\lfloor T/2 \rfloor}, 1\right)$.

v	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_4$	.754	.553	.346	.200	.081	.655	.491	.277	.143	.081
$\hat{\Lambda}_{10}$	.738	.525	.311	.173	.064	.592	.442	.256	.139	.066
KS	.256	.189	.127	.088	.052	.542	.349	.185	.078	.059
CvM	.362	.291	.164	.103	.066	.601	.445	.260	.130	.078

Table B3f: Power $H_0 : Y_i \sim Exp[1]$ vs. $H_1 : Y_i \sim \Gamma(v,1)$.

v	1.10	1.15	1.20	1.25	1.30	1.10	1.15	1.20	1.25	1.30
$\hat{\Lambda}_4$	.113	.125	.226	.302	.432	.189	.302	.595	.770	.886
$\hat{\Lambda}_{10}$	.103	.106	.179	.277	.398	.177	.285	.550	.712	.825
KS	.066	.069	.136	.200	.252	.096	.193	.404	.616	.747
CvM	.094	.099	.179	.237	.343	.174	.280	.551	.732	.853

Table B4: MSEs of estimated quantiles and mean predictive log-scores.

Table B4a: MSE of estimated quantiles, $\hat{q}_{Y^T}$, for fitted $AR(1)$ with $\bar{t}_{(4)}$ errors

m	4	4	4	4	10	10	10	10
n / π	25	50	100	200	25	50	100	200
0.05	.0454	.0287	.0191	.0125	.0837	.0563	.0335	.0216
0.25	.0433	.0248	.0131	.0073	.0503	.0264	.0149	.0080
0.50	.0329	.0168	.0079	.0036	.0397	.0173	.0094	.0045
0.75	.0421	.0249	.0132	.0068	.0512	.0266	.0140	.0076
0.95	.0457	.0296	.0198	.0124	.0819	.0557	.0332	.0218

Table B4b: MSE of estimated quantiles, $\hat{q}_{Y^T}$, for fitted $AR(1)$ with $\bar{\chi}^2_{(4)}$ errors

m	4	4	4	4	10	10	10	10
n / π	25	50	100	200	25	50	100	200
0.05	.1742	.1583	.1470	.1416	.1039	.0711	.0515	.0372
0.25	.0337	.0205	.0123	.0087	.0406	.0227	.0140	.0081
0.50	.0434	.0280	.0186	.0141	.0449	.0251	.0147	.0093
0.75	.0748	.0475	.0328	.0252	.0759	.0475	.0282	.0188
0.95	.3110	.2914	.2638	.2491	.2531	.2070	.1773	.1578

Table B4c: Average in sample Log-Scores for fitted $AR(1)$ with $\bar{t}_{(4)}$ and $\bar{t}_{(8)}$ errors.

The mis-specified Gaussian model is $f_y(\beta)$, the correct model is $g_y(\beta; v)$.

	n	25	50	100	200	25	50	100	200
Log-Score	v	4	4	4	4	8	8	8	8
$\ln p_{\hat{X}_{n+1}}\left(\hat{\theta}_{(4)}\right)$		.0935	.0801	.0714	.0618	.0799	.0492	.0374	.0281
$\ln p_{\hat{X}_{n+1}}\left(\hat{\theta}_{(10)}\right)$		.1019	.0871	.0704	.0674	.0883	.0668	.0455	.0347
$\ln f_{Y_{n+1}}\left(\hat{\beta}_n\right)$		-2.56	-2.39	-2.28	-2.11	-2.04	-1.95	-1.90	-1.62
$\ln g_{Y_{n+1}}\left(\hat{\beta}_n; v\right)$		-1.47	-1.43	-1.36	-1.35	-1.53	-1.42	-1.27	-1.26
$\ln g_{Y_{n+1}}\left(\hat{\beta}_n; \hat{v}\right)$		-1.82	-1.61	-1.55	-1.49	-1.81	-1.67	-1.50	-1.46

APPENDIX C (Application graphs)

Figure C1: S&P500 Monthly Returns; Jan 1998 to December 2017

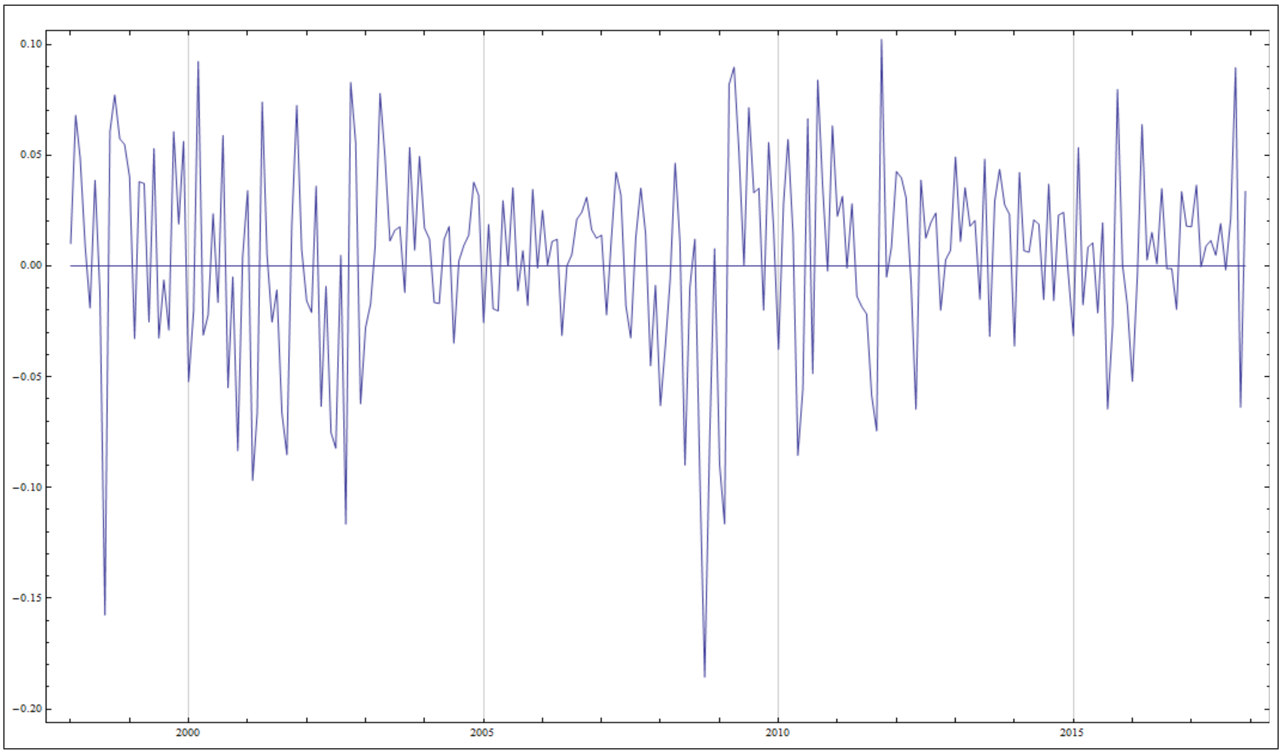

Figure C2: Correlogram for Returns “+” and for Squared Returns “O”

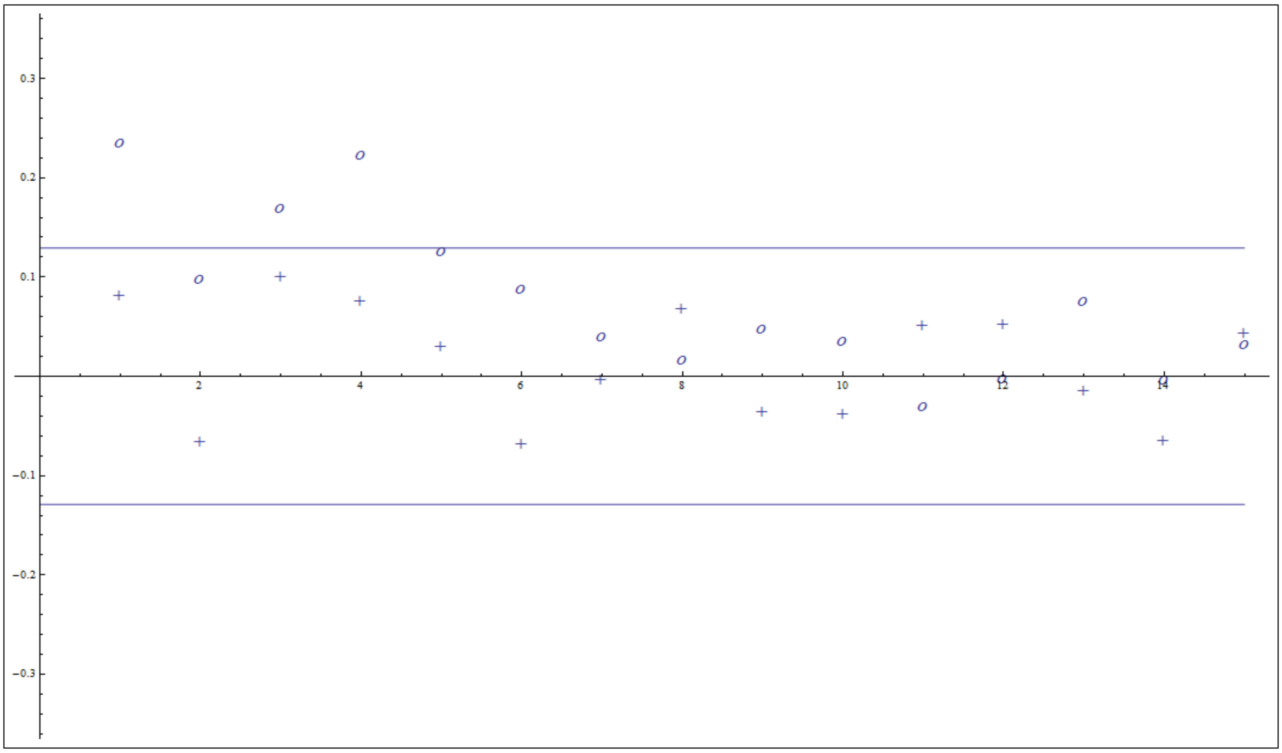

Figure C3: Observed p-values for $\hat{\Lambda}_{10}$. Solid line for M_6 dashed line for M_1.

Figure C3a: Recursive Sampling Scheme

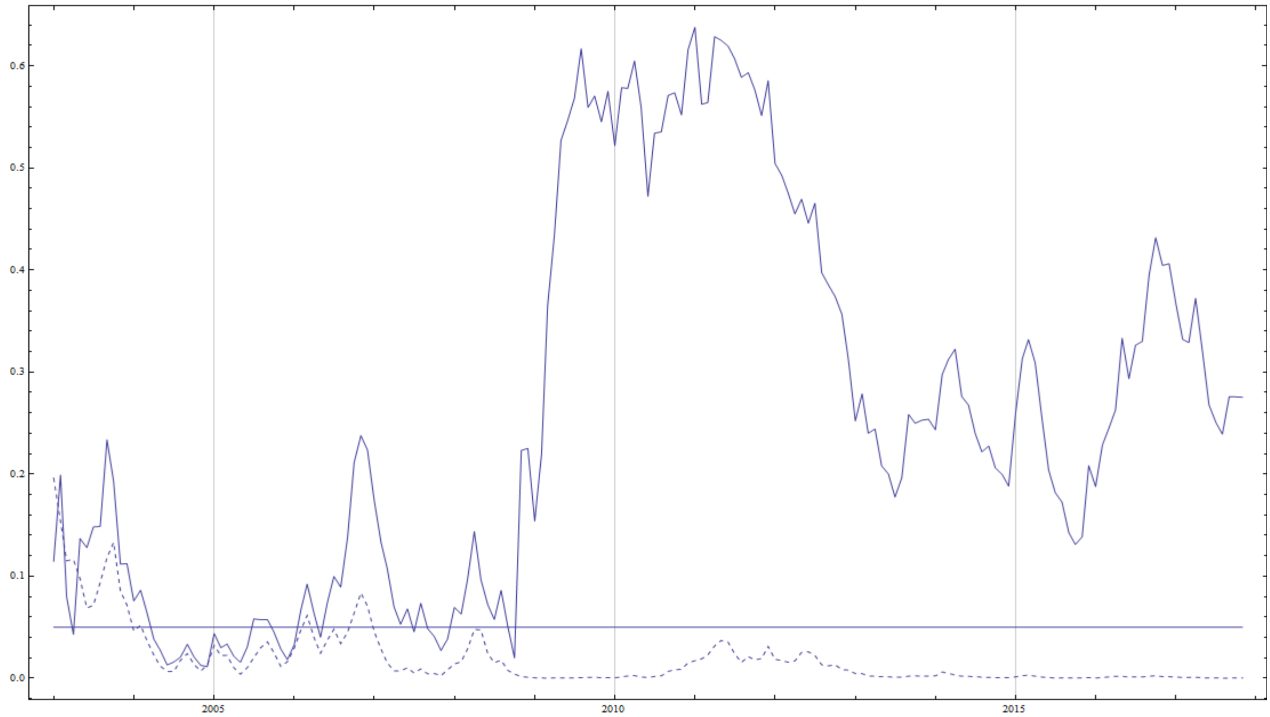

Figure C3b: Rolling Sampling Scheme (5 year rolling window)

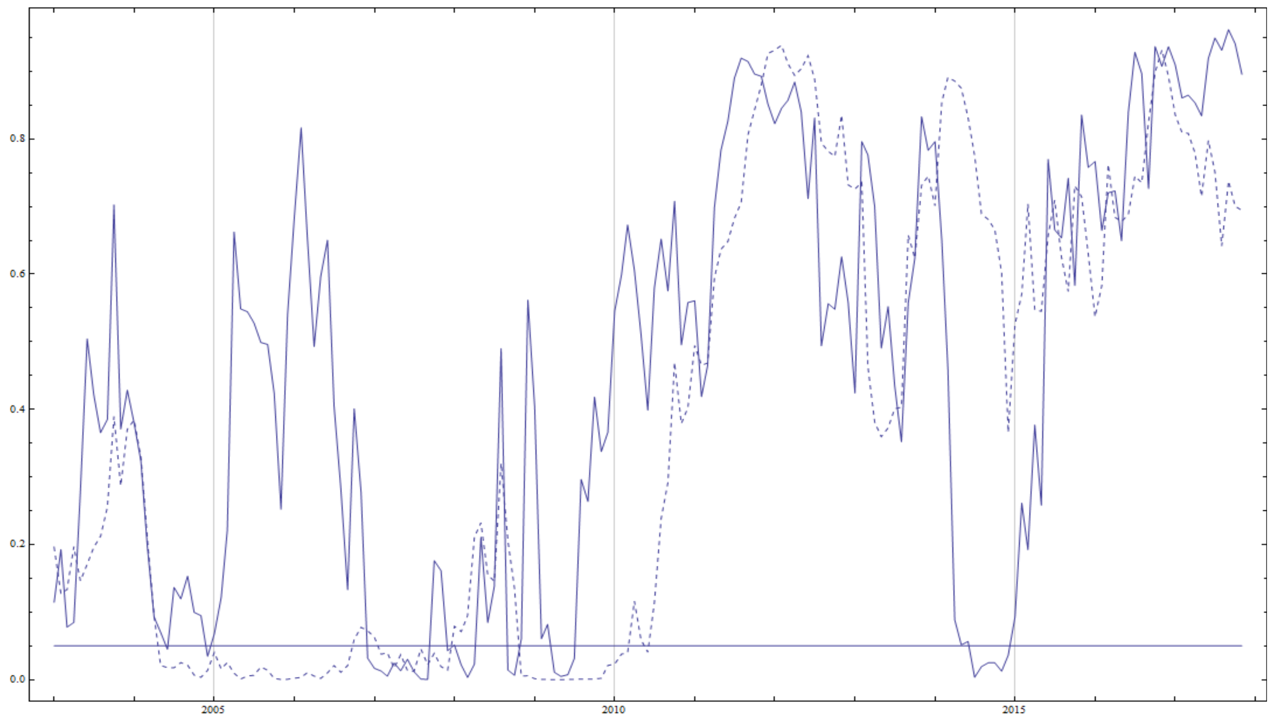

Figure C4: Relative Estimates of Predicted Value at Risk, $VaR_{0,05}$;

Jan. 2003 to Nov 2017, M_1 solid and M_6 dashed.

Figure C4a: Recursive Predictive Scheme

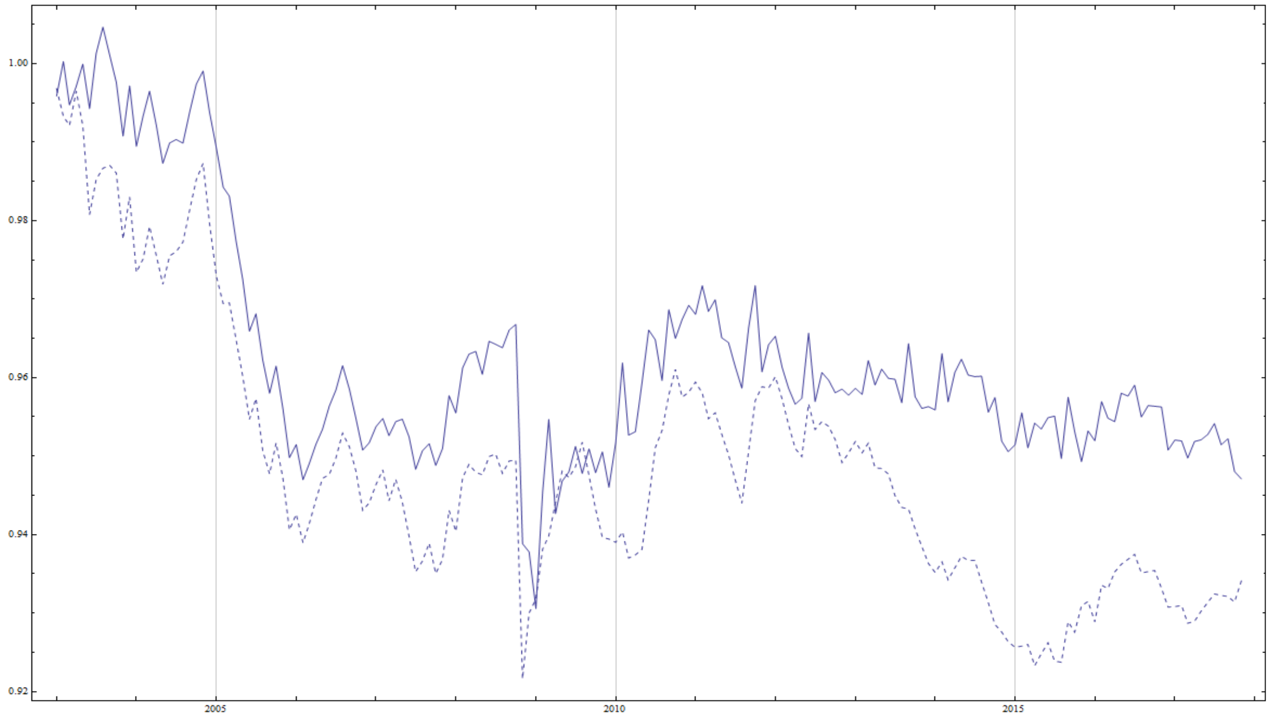

Figure C4b: Rolling Predictive Scheme (5 year rolling window)

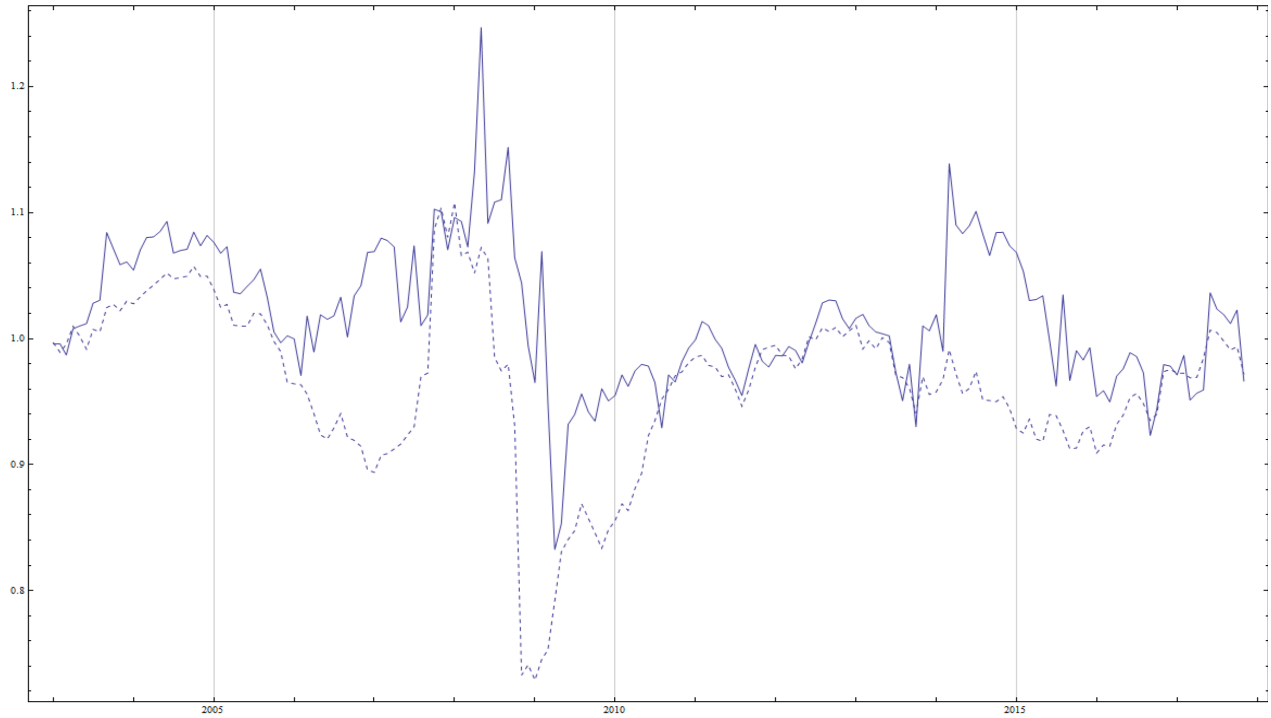

Figure C5: Ratios of the nonparametric to the Gaussian predictors of the probability of a negative return, Jan. 2003-Nov. 2017. Solid line for M_1, dashed for M_6.

Figure C5a: Recursive sampling scheme.

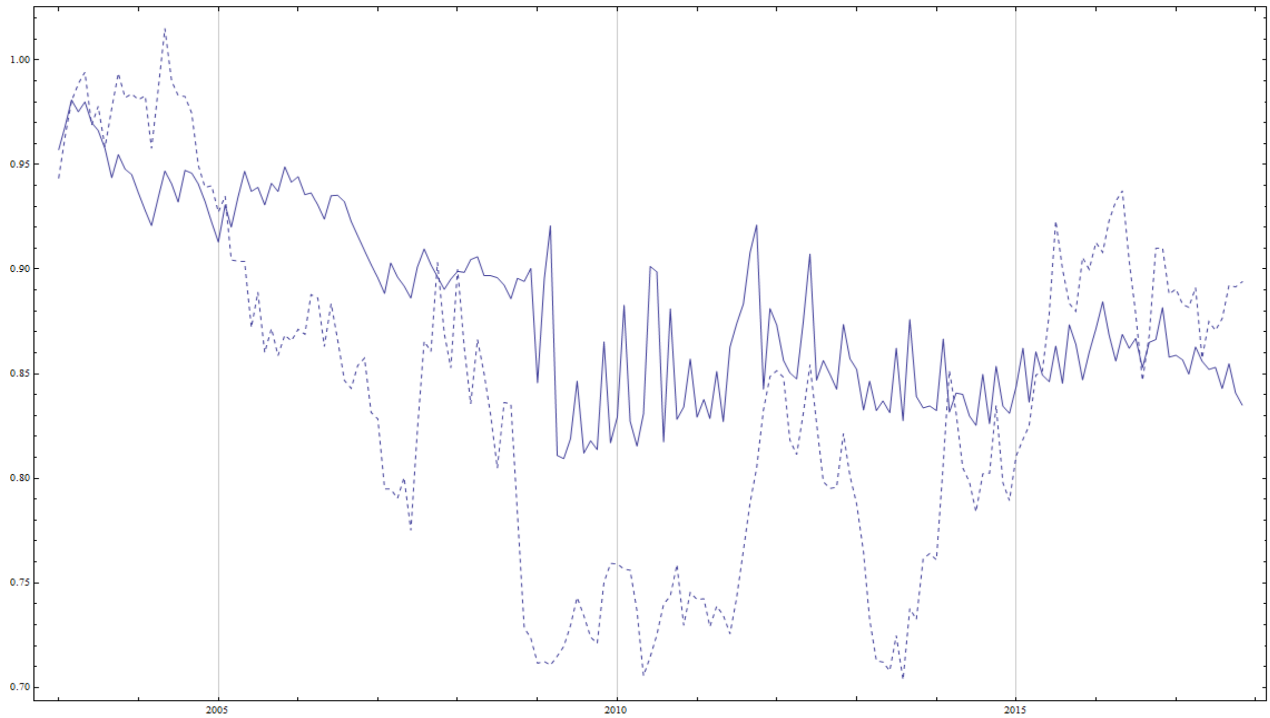

Figure C5b: Rolling Sampling Scheme (5 year rolling window)