Understanding High-Frequency Stock Price Dynamics: with an Application on Intraday Values at Risk

Robert Davies
Department of Economics, Duke University,
Durham, NC 27708 (robert.davies@duke.edu)

July 23, 2015

Abstract

This paper develops a method for modeling high-frequency asset prices. It does so by showing how asset prices might be transformed into Lévy processes. Once in the class of Lévy processes this paper develops a novel estimation procedure and a novel test of a model's specification by performing the estimation and testing over a suitably chosen family of weighting functions. An empirical study fits a selection of asset returns to two classes of Lévy processes; and, finally, a detailed empirical exercise develops a flexible method to calculate intraday values at risk up to any within day horizon. A backtest of the intraday values at risk show their coverage to be right in line with the theoretical correct values.

Key Words: characteristic function estimation, high-frequency data, Lévy processes, model specification testing, value at risk

1 Introduction

The literatures on asset pricing, financial econometrics, and financial engineering have all devoted considerable attention to modeling asset returns. Examples include the tempered stable or CGMY models of Rosinski (2007) or Carr et al. (2002) or the many commonly used jump-diffusion models. This paper builds and expands upon these and other works by looking at how one might think of modeling the high-frequency returns of an asset and how to go about testing the fit of such models. In doing so it develops both a novel estimation procedure for asset returns and a novel method of model specification testing.

The paper starts by assuming assets follow a fairly general class of Itô semi-martingales and looks at what assumptions might be needed to simplify such models. In particular it will be shown that if a particular time invariant transformation of the jump measure of the process exists then Itô semi-martingales might be transformed into Lévy processes – which have been extensively studied in the literature. The motivation for doing so, in addition to the fact that Lévy processes have been widely studied, is that Lévy processes have many particularly convenient features that aid in their estimation. One particular useful property of Lévy processes is given by the Lévy-Khintchine theorem which derives a closed-form expression for the characteristic function of any Lévy-process.

Having a method to derive the characteristic functions of asset returns leads the paper immediately into thinking about how to estimate the characteristic function of a process and how the characteristic function of a process might be used to develop a test of model specification fit. It is here that the main theoretical results of the paper are developed. A new method of characteristic function estimation is developed which, while not guaranteeing full maximum-likelihood efficiency, should considerable improve the efficiency of the estimation in practice. Importantly, this method relies on far fewer assumptions than estimation strategies like those in Carrasco and Florens (2002) which do aim for full maximum-likelihood efficiency and it does not require knowledge of the density of the process like the method in Feuerverger and McDunnough (1981).

Having developed a method of characteristic function estimation, this paper proceeds

by examining how the insights from thinking about how to estimation the characteristic function might be used to develop a model specification test. Building on the work of Halbert White and Herman Bierens among others this paper develops a new test of model specification. A proof of the consistency of this test is given along with a detailed Monte Carlo study to show its finite sample performance. Following this, a brief empirical study is presented in which the methods in this paper are used to estimate and test the fit of two Lévy processes for the high-frequency returns on the E-mini S&P 500 futures and a selection of five large market capitalization stocks. The first model being the CGMY or tempered stable distribution from Carr et al. (2002) or Rosinski (2007) and the second being a jump-diffusion model with double-exponentially distributed jump sizes. The jump-diffusion model is strongly rejected for all but one asset at one sampling frequency, but the tempered stable model appears to be a good fit for many of these assets. To the author's knowledge this is the first paper to fit the tempered stable distribution or any pure jump model to the high-frequency returns of a security.

Finally, this paper looks at how the results developed here might be used in practice by developing and testing a method for calculating intra- or within-day values at risk. Since the financial crisis intraday risks have gain a significant amount of attention. Basel III, for example, includes several guidelines on how banks and other financial firms should monitor intraday liquidity risks. Given the growth in day trading it makes sense for firms to monitor the intraday market risks they are taking on as well. This is especially true given that there is evidence that many firms do not fully monitor their day trading activity. (For example, a 2003 SEC report, "Special Study: Report of Examinations of Day-Trading Broker-Dealers", found that "most" firms did not monitor the capital compliance of their day trading activities.) A flexible method for calculating intraday values at risk would be a highly useful tool for firms wishing to monitor the within day market risks they are taking on.

2 Modeling Asset Price Dynamics

We begin our modeling of high-frequency assets by assuming they belong in the class of Itô semi-martingales. Itô semi-martingales are a set of benchmark models for building time-varying stochastic processes and can be expressed as processes of the form

$$X_t = \int_0^t a_t dt + \int_0^t \sigma_s dW_s + J_t \tag{1}$$

where a_t is the drift of the process, σ_t is a time-varying volatility process, and J_t is a jump process.

Given that high-frequency asset returns are often sufficiently close to being mean zero, this paper will drop any drift component in the formulation of these processes. Doing so we can connect drift-less Itô semi-martingales to the class of what Sato defines as *additive processes* in Sato (2013). (Drift-less Itô semi-martingales form a subset of additive processes. See Sato (2013) for details.) Doing so gives us an expression for the characteristic function of such processes.

Theorem 1 (Sato, 2013)**.** *Any additive process (or drift-less Itô semi-martingale) X_t has characteristic function*

$$\phi_{X_t}(u) = \exp\left\{-\frac{1}{2}\sigma_t^2 u^2 + \int_{\mathbb{R}} [e^{iuy} - 1 - iuy\mathbb{1}(|y| \leq 1)] m_t(dy)\right\}. \tag{2}$$

Readers familiar with Lévy processes will hopefully see an immediately resemblance in the above theorem with the Lévy-Khintchine theorem for Lévy processes – with the noticeable difference that the characteristic function of drift-less Itô semi-martingales is time varying. This time variation creates problems for modeling and therefore any usable model needs to remove or model the time-dependencies. A standard approach in modeling asset price returns is to scale processes by their variation. Let us observe what happens to the characteristic function in Theorem 1 when we scale the process by its variation σ_t.

The characteristic function of X_t/σ_t is

$$\begin{aligned}
&\phi_{X_t/\sigma_t}(u) = \phi_{X_t}(u/\sigma_t) \\
&= \exp\left\{-\frac{1}{2}\sigma_t^2(u/\sigma_t)^2 + \int_{\mathbb{R}}[e^{i(u/\sigma_t)y} - 1 - i(u/\sigma_t)y\mathbb{1}(|y| \leq 1)]m_t(dy)\right\} \\
&= \exp\left\{-\frac{1}{2}u^2 + \int_{\mathbb{R}}[e^{iuz} - 1 - iuz\mathbb{1}(|z| \leq 1)]m_t(\sigma_t dz)\right\}
\end{aligned} \tag{3}$$

where we used the change of variables $z = \sigma_t y$. Notice that the diffusive component became time-invariant and we were able to express the time dependency of the jump-component through the time-varying Lévy measure $m_t(\cdot)$ above.

In order to remove the time dependence above a researcher needs to make the crucial assumption that there exists some Lévy measure $\tilde{m}(\cdot)$ such that $m_t(\sigma_t dz) = \tilde{m}(dz)$. If this is the case, then scaling high-frequency returns by their variation would result in a time homogeneous process with characteristic function

$$\phi_{X_t/\sigma_t}(u) = \exp\left\{-\frac{1}{2}u^2 + \int_{\mathbb{R}}[e^{iuz} - 1 - iuz\mathbb{1}(|z| \leq 1)]\tilde{m}(dz)\right\} \tag{4}$$

which by the Lévy-Khintchine theorem is exactly the characteristic function of a Lévy process. Given the characteristic function completely defines a statistical process if X_t/σ_t has a characteristic function defined above in (4) then X_t/σ_t is a Lévy process.

While verifying the assumption of time homogeneity analytically might be possible in some applications, it is not too hard to imagine a researcher lending support to such an assumption empirically by plotting the scaled returns (and possible scaled squared returns) of the process and calculating a sufficiently large number of autocorrelations and partial autocorrelations for the process. If plots of the scaled returns appeared homogeneous and the scaled returns showed little evidence of any autocorrelation then perhaps the assumption of time invariance might be supported.

The next subsection briefly reviews Lévy processes for the interested reader.

2.1 Lévy Processes

Lévy processes are at the core of many models within empirical finance and financial econometrics. Cont and Tankov (2004) provides an excellent review of Lévy processes. A Lévy process is any stochastically continuous càdlàg process with independent and stationary increments. Brownian motion, any time-homogeneous jump diffusion process, the tempered stable processes of Rosinski (2007) or the CGMY processes of Carr et al. (2002) among many others are all Lévy processes. These models have been used extensively to model asset returns, especially high-frequency asset price returns.

Any Lévy process can be defined by three parameters, often termed the Lévy triplet (a, σ, m). While a full treatment of Lévy processes is beyond the scope of this paper, in general, one can think of a as affecting the drift of the process and σ as scaling the diffusive component of the process. The parameter m is a measure that determines the size and intensity of any jumps in the process and is known as the Lévy measure. The Lévy-Khintchine theorem provides an incredibly useful way to analyze Lévy processes as it gives the characteristic function of any Lévy process.

$$\phi(u) = \exp\left\{aiu - \frac{1}{2}\sigma u^2 + \int_{\mathbb{R}} [e^{iuz} - 1 - iuz\mathbb{1}(|z| \leq 1)]m(dz)\right\}. \tag{5}$$

Knowing the Lévy triplet (a, σ, m) of any process then immediately gives one the characteristic function of that process. In the literature one can easily find closed-form expressions for the characteristic functions of many popular Lévy processes. Many of which were derived using the Lévy-Khintchine theorem.

3 Characteristic Function Estimation

The previous section showed how, under the right assumptions, asset prices might be modeled as Lévy processes and how the characteristic function of any Lévy process could be expressed in closed-form. Since the characteristic function of a process completely describes that process, knowledge of the characteristic function gives us the full dynamics of the

underlying process. What this means from an estimation standpoint is that estimating a model of the characteristic function of a process is equivalent to estimating the process itself. Given this insight the paper proceeds by building a framework for estimating characteristic functions.

Any characteristic function estimation procedure tries to match the theoretical characteristic function $\phi_0(u, \theta)$ under the null of a model to the empirical characteristic function $\phi_n(u)$ of the data – where $\phi_0(.)$ is the class of characteristic functions under the null and θ is the set of parameters to match. For readers unfamiliar with thinking about the characteristic function of a process note that the characteristic function of some process X is simply the Fourier transform of the density of that process. It is defined as $\phi_X(u) \equiv \mathbb{E}[e^{iuX}]$ for any $u \in \mathbb{R}$ where i is the imaginary unit and u indexes the frequency. As such the characteristic function translates the density of the process from the space domain to the frequency domain.

One can generalize most any characteristic function estimation technique as solving the following problem.

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \int_{\mathbb{R}} |\phi_n(u) - \phi_0(u, \theta)|^2 g(u) du. \tag{6}$$

Which one can cast as a GMM procedure using a continuum of moment conditions. (In addition, a procedure that only attempted to match the characteristic function at discrete frequencies, as is sometimes done, could be placed within the framework of equation (6) by choosing $g(u)$ as a set of Dirac mass functions – which might themselves be weighted.)

In general, unlike maximum likelihood estimation (MLE), characteristic function estimation will not achieve the Cramer-Rao efficiency lower bound. However by properly selecting the weighting function, $g(u)$, one can improve the efficiency of the estimation and, given the right assumptions, even achieve full maximum-likelihood efficiency. (See Carrasco and Florens (2000) for a detailed discussion.) The next subsection discusses how to think about choosing the weighting function $g(u)$.

3.1 Weighting Functions

The choice of a weighting function, $g(u)$, in the estimation procedure of equation (6) involves the choice of how to weight the frequencies $u \in \mathbb{R}$ in the characteristic function $\phi(u)$ under consideration. Weighting different frequencies differently involves a trade off between which aspects of a process to weight most heavily. The large and small moves of a process, for example, influence the characteristic function differently at different frequencies. An example of this is the α-stable, tempered-stable, or CGMY processes. Following Todorov and Tauchen (2012) one can show that the activity parameter of these processes (often denoted α) governs the small moves of these processes most heavily and inturn will influence the characteristic function of these processes most heavily at smaller frequencies (i.e., smaller values of u in $\phi(u)$). A researcher interested in estimating this parameter then would do best to give sufficient weight to the smaller frequencies of the characteristic function.

To see this idea more generally, recall that the characteristic function is defined as $\phi_X(u) = \mathbb{E}[e^{iuX}]$. From this definition, it is easy to see that small frequencies (small values of u) will be most influenced by large values of X (large moves of the process), whereas larger frequencies (larger values of u) will be influenced by both large and small moves (large and small values of X). A weighting function then should be tailored to the model and parameters under consideration and the parameters the researcher finds of most interest. A researcher that was most interested in a model governed by smaller frequencies would do best to choose a weighting function that gave more weight to smaller frequencies, whereas a researcher that was most interested in models governed by larger frequencies would do best to choose a weighting function that gave more weight to larger frequencies.

This discussion begs the question of how exactly then to select a weighting function or at the least from which set or class of weighting functions should a particular weighting function be selected. One commonly used class of weighting functions are the Gaussian densities. These functions are centered at zero and decline exponentially in the tails. Such weighting functions are limiting because, while it possible to increase the weight on the tails, the majority of the mass of any Gaussian function will always be centered on the origin. This could potentially be a problem, because, as was discussed, it might be desirable

to weight frequencies away from the origin more heavily than those at the origin which is something the Gaussian weighting functions cannot do.

Another idea in selecting a class of weighting functions comes from a discussion in Heathcote (1977) in which the estimation of the characteristic function of stable processes is discussed. Heathcote (1977) notes that the informational content of the characteristic function of stable processes declines near the origin and in the tails. He suggests using a weighting function that down-weights frequencies near the origin and in the tails, rather than the commonly used Gaussian weighting functions, but goes no further in developing or suggesting such weighting functions. This paper builds on his insights.

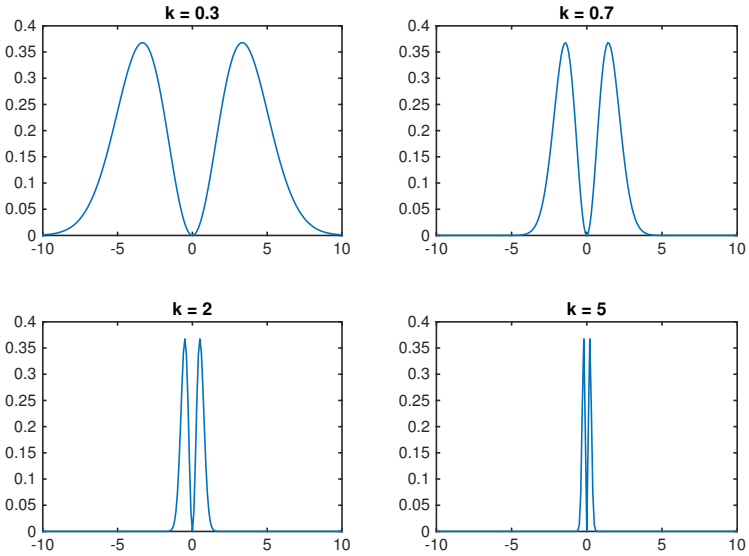
I define weighting functions that down-weight frequencies near the origin and in the tails 'bimodal' weighting functions since they are dual peaked when plotted on the real line. These weighting functions are incredibly flexible in that such weighting functions can be tuned to place weight near the origin, at moderate frequencies, and even at very high frequencies by changing the scale and location of the bimodal peaks of the weighting function. To illustrate such an idea Figure 1 plots a specific example of a such a weighting function and the class of weighting functions used here.

Having decided on a class of weighting functions or at the very least the characteristics of a weighting function that might be desirable the next question is how exactly to choose a particular weighting function from that class. It is here that this paper introduces a particularly novel idea to the characteristic function estimation literature. Rather than choose one particular weighting function $g(u)$ and perform the estimation once, the idea here is to choose a family of weighting functions, in particular one with bimodal properties, and estimate the model once for each weighting function in the family. Denoting the family by $\mathcal{K}$ we might express the idea as estimating the model over each $g_k(u)$ for $k \in \mathcal{K}$. For each $k \in \mathcal{K}$ this will give an estimate $\hat{\theta}_k$ of the parameters. A sufficiently broad family of weighting functions would allow the researcher to estimate the model weighting smaller frequencies, then moderate frequencies, and finally large frequencies most heavily.

Doing so would give the researcher a set of estimators, $\hat{\theta}_k$, corresponding to the weighting functions in the family. An immediate idea on selecting which estimator to prefer is to select

Figure 1: Family of Weighting Functions

$$g_k(u) = (uk)^2 \exp\{-(uk)^2\}$$



the one that results in the 'smallest' estimated asymptotic covariance matrix.[1] Doing so would improve the efficiency of the estimation procedure against a procedure that selected a weighting function in an *ad hoc* manner. Further if a particular $g_k(u)$ was the optimal weighting function in Feuerverger and McDunnough (1981) or Carrasco and Florens (2002) then we would have achieved full maximum likelihood efficiency without knowledge of the density as required in Feuerverger and McDunnough (1981) or the assumptions necessary in Carrasco and Florens (2002).

Needing a parametric class of weighting functions for the family I chose

$$g_k(u) = (uk)^2 \exp\{-(uk)^2\} \quad \text{for } k > 0. \tag{7}$$

The function in (7) has several beneficial features. As can be seen in Figure 1 the weighting function consists of two bi-modal Gaussian-type curves reflected across the origin. The

[1] Here I take 'smallest' to mean the estimated asymptotic covariance with the smallest trace though other partial orderings could be used.

function gives zero weight at the origin[2] and in the limit of the tails. The parameter k controls the thickness and location of the curves. As $k \to 0$ the curves become fatter and further from the origin, whereas when $k \to \infty$ the curves become steeper and closer to the origin. Not knowing the exact location of the maximal informational content of the characteristic function under consideration the researcher can vary k to place more or less weight on different locations of the characteristic function.

The next section uses the parametric class of weighting functions just introduced and the discussion on characteristic functions to build a test of model specification.

4 Specification Testing

The family of weighting functions in (7) that were considered in the estimation procedure also have several features that lend themselves to a framework for model specification testing. Before delving into the usefulness of the family of functions in (7) we need to consider the null of the test and examine the testing procedure.

4.1 Null Hypothesis

Let $\phi_n(u)$ be the empirical characteristic function of the observed data, let $\phi(u)$ be the true unobserved characteristic function of the process under consideration, and let $\phi_0(u, \theta_0)$ be the characteristic function under the null of the model. (We need to match not just the class of characteristic functions $\phi_0(.)$ but also the parameter values θ_0.)

Since the characteristic function is complex valued we can decompose it into its real and imaginary parts such that $\phi(u) = c(u) + is(u)$. Letting $m(\theta_0, u) = c(u) - c_0(u, \theta_0) + s(u) - s_0(u, \theta_0)$ we can state the null as

$$\begin{aligned} \Omega_0 &\equiv \{\phi(u, x) = \phi_0(u, \theta_0), \forall u \in \mathbb{R}\} \\ &= \{m(\theta_0, u) = 0, \forall u \in \mathbb{R}\}. \end{aligned} \tag{8}$$

[2]This is not a problem or limitation because while the characteristic function of a process might be informative in the neighborhood around the origin, any characteristic function is always 1 at the origin. Therefore the origin itself cannot provide any information on the process under consideration.

Which we test using

$$\Omega_{0,n} = \{m_n(\theta_0, u) = 0, \forall u \in \mathbb{R}\} \tag{9}$$

where $m_n(\theta_0, u) = c_n(u) - c_0(u, \theta_0) + s_n(u) - s_0(u, \theta_0)$.

The following brief lemma shows that when the process under consideration is continuously distributed that we have $\mathbb{P}[m(\theta_0, u) = 0] = 0$ whenever $\phi(u) \neq \phi_0(\theta_0, u)$, thereby justifying the statement of the null in equation (8). (Stating the null as in (8) greatly simplifies the computations in the theorems to follow.)

Lemma 1. *Let X be a continuously distributed random variable and let $\phi(u)$ be the characteristic function of X. Define $m(\theta_0, u) = c(u) - c_0(u, \theta_0) + s(u) - s_0(u, \theta_0)$ as above. When $\phi(u) \neq \phi_0(\theta_0, u)$ we have $\mathbb{P}[m(\theta_0, u) = 0] = 0$.*

4.2 Testing Procedure

In order to test the null in equation (9) the following test statistics are used. Recall the weighting function $g_k(u)$ in equation (7) is indexed by $k \in \mathcal{K}$. For each $k \in \mathcal{K}$ then we can define the statistic

$$Q_{n,k}(\theta) = \int_{\mathbb{R}} m_n(\theta, u) g_k(u) du \tag{10}$$

and test across k's using

$$T_n(\theta) = \sup_k |\sqrt{n} Q_{n,k}(\theta)|. \tag{11}$$

The next subsection will show that having $Q_{n,k}(\theta) = 0$ for all $k \in \mathcal{K}$ is sufficient to conclude $\phi(u) = \phi_0(u)$ and derive the distribution of test statistic $T_n(\theta)$ in equation (11).

4.3 Consistency and Distribution of the Test

The following theorem shows that for the test statistic above

$$Q_{n,k}(\theta) = \int_{\mathbb{R}} m_n(\theta, u) g_k(u) du = 0$$

for all $k \in \mathcal{K}$ is sufficient to conclude $\phi(u) = \phi_0(u)$. It is here that the usefulness of the weighting functions in equation (7) becomes apparent. The weighting functions in equation (7) is analogous to the class of consistent weighting functions in the integrated conditional moment tests of Bierens and Ploberger (1997) or the revealing sets of Stinchcombe and White (1998). (It should be noted however that the weighting function I chose in equation (7) does not fit directly into either class and therefore the consistency of my testing procedure does not follow directly from an application of Bierens and Ploberger (1997) or Stinchcombe and White (1998). In fact, I was unable to come up with a bimodal weighting function that fit within the framework of either paper.)

The class of possible weighting functions is quite large and many alternate choices could have been employed. Recall however that the weighting functions in equation (7) were chosen because they would place weight on the characteristic function in correspondence with its informational content. For the same reason this aided in the estimation procedure it now aids us here in model specification testing. Bierens and Ploberger (1997) require, as does the theorem here, that the set of weighting functions have positive Lebesgue measure. Stinchcombe and White (1998) require that the set be dense. Given that neither is possible in practice, a set of weighting functions chosen based on the informational content of the process in consideration will aid the accuracy of the testing procedure in practice where a dense set of weighting functions or a set with positive Lebesgue measure is unfeasible.

Theorem 2. *Consider the weight functions $g_k(u)$ given in equation (7). Let $\mathcal{K}$ be a compact subset of $\mathbb{R}_+$ with strictly positive Lebesgue measure. Then, if for all $k \in \mathcal{K}$ we have $\int_{\mathbb{R}} m(\theta, u) g_k(u) du = 0$, we can conclude $m(\theta, u) = 0$ for every $u \in \mathbb{R}$ and therefore that $\phi(u) = \phi_0(u)$.*

Given the consistency of the testing procedure above we need a distributional theory to use the test in practice. The theorem below gives this distribution and a method for going about the test. First, however, define $\hat{\theta}_{n,k^*}$ be the chosen estimate of θ_0 where k^* is the particular $k \in \mathcal{K}$ of the chosen estimate.[3] The following theorem requires one set of

[3]Recall we are choosing $\hat{\theta}_{n,k^*}$ as the estimator that leads to the smallest trace of the estimated asymptotic covariance matrix though other choice procedures could be considered.

assumptions.

Assumption A. *(i) $\hat{\theta}_{n,k}$ converges uniformly in $k \in \mathcal{K}$ to θ_0. (ii) The first and second derivatives of $c_0(u)$ and $s_0(u)$ are uniformly bounded by functions integrable with respect to $g_k(u)$. (iii) The data $\{x_j\}$ for $j = 1, \ldots, n$ are independent and identically distributed random variables.*

Assumption A(i) is easy to verify by a uniform law of large numbers since the characteristic function of any process is a bounded function. Assumption A(ii) is easy to verify for most Lévy processes since $\phi(u) = c(u) + is(u)$ is known in closed form. Assumption A(iii) needs to be assumed. However a researcher can check plots of the returns as well as autocorrelograms of the returns and the absolute returns to support such an assumption.

Theorem 3. *Under Ω_0 and Assumption A we have $\sqrt{n}Q_{n,k}(\hat{\theta}_{n,k^*})$ converges to $Z(k)$. Where $Z(k)$ is a Gaussian process defined on $\mathcal{C}(\mathcal{K})$ where $\mathcal{C}$ is the space of continuous functions and the covariance function of $Z(k)$ is the function $\Gamma(k_1, k_2)$ (given in the proof). Moreover, the function*

$$T_n(\hat{\theta}_{n,k^*}) = \sup_{k \in \mathcal{K}} |\sqrt{n}Q_{n,k}(\hat{\theta}_{n,k^*})|$$

converges in distribution to $\sup_{k \in \mathcal{K}} |Z(k)|$.

Theorem 3 gives us a method to perform the testing procedure. First, choose a set $\mathcal{K}$ from which to draw weighting functions $g_k(u)$ for $k \in \mathcal{K}$. Next, perform the estimation as in Section 3 and select a chosen parameter estimate $\hat{\theta}_{n,k^*}$. For each $k \in \mathcal{K}$ calculate $Q_{n,k}(\hat{\theta}_{n,k^*})$ and then $T_n(\hat{\theta}_{n,k^*}) = \sup_{k \in \mathcal{K}} |\sqrt{n}Q_{n,k}(\hat{\theta}_{n,k^*})|$. Next, estimate the covariance function Γ of the Gaussian process $Z(k)$ for the set of $k \in \mathcal{K}$. Finally, using this estimate $\hat{\Gamma}$ stimulate the process $Z(k)$ and estimate the p-value of the statistic $T_n(\hat{\theta}_{n,k^*})$ under the null based on where $T_n(\hat{\theta}_{n,k^*})$ falls in the simulated distribution of $\sup_{k \in \mathcal{K}} |Z(k)|$.

5 Monte Carlo Study

To analyze the performance of the estimation and testing procedures I performed a one thousand replication Monte Carlo study. To do so I simulated data from two common Lévy

processes often used in the literature, a jump-diffusion model and a tempered stable (or CGMY) model. The processes were simulated over $T = 2257$ days with $N = 78$ returns in each day, giving $TN = 176,046$ simulated returns per replication. The number of days corresponds to nine calendar years and the number of returns in each day corresponds to the number of five minute returns in a typical trading day.

5.1 Tempered Stable / CGMY Process

The tempered stable process is a pure-jump Lévy process. A review of the tempered stable process can be found in Cont and Tankov (2004) or a full treatment can be found in Rosinski (2007). The CGMY model in Carr et al. (2002) is also a specialized form of the tempered stable model.

Every tempered stable process has a Lévy jump measure of the form

$$m(x) = c\frac{1}{|x|^{1+\alpha}}e^{-\rho|x|} \tag{12}$$

with parameters $\alpha < 2, c > 0$, and $\rho > 0$. The parameter α controls the activity of the process.[4] When $1 \leq \alpha < 2$ the process is of infinite activity and infinite variation, when $0 \leq \alpha < 1$ the process is of infinite activity, but finite variation, and when $\alpha < 0$ the process is of both finite activity and finite variation and corresponds to a sub-class of compound Poisson processes. When $\alpha = 0$ the tempered stable process coincides with the variance gamma process. The parameter c is essentially a scale parameter although it has an effect on the intensity of the jumps. Finally, the parameter ρ controls the tempering of the tails of the Lévy measure and thereby the large jumps of the process. Larger values of ρ correspond to greater tempering and therefore fewer larger jumps.

I chose $\alpha = 1.2$, $c = 43$, and $\rho = 1.15$ for the model I replicated. These parameters are in line with the estimates from several high-frequency stock price returns.

[4] When $\alpha > 0$ it corresponds to the Blumenthal-Getoor index.

5.2 Simulating the Tempered Stable Process

Since the tempered stable and CGMY processes do not have a closed-form density, I needed a method to simulate the tempered stable process I had chosen. To do so I used a method of Fourier inversion to go from the characteristic function of the process to an estimate of the cumulative distribution function (CDF). With the CDF of the process, I could then use the probability integral transform to replicate my tempered stable process as detailed below.

The Fourier inversion of the characteristic function was done using the Gil-Pelaez formula.

$$F(x) = \frac{1}{2} + \frac{i}{2\pi}\int_0^\infty \frac{\phi(u)e^{-iux} - \phi(-u)e^{iux}}{u}du \tag{13}$$

While the CDF of any process is necessarily a monotonically increasing function, estimating the CDF in such a manner will not necessarily lead to a monotonic function. To address this issue I followed Li et al. (2013) and Chernozhukov et al. (2010) and performed a monotonization of the estimated CDF via a rearrangement process as follows.

Let $\hat{F}(.)$ be the estimated CDF of the tempered stable process based on equation (13). For any $\tau \in (0,1)$ define the estimated τ-quantile of the process as

$$\hat{Q}(\tau) = \inf\{x \in \mathbb{R} : \hat{F}(x) \geq \tau\}. \tag{14}$$

From this estimated quantile function we can define a CDF based on inverting the estimated quantile function that is guaranteed to be monotonic. Such a function takes the form

$$\hat{F}^*(x) = \inf\{\tau \in (0,1) : \hat{Q}(\tau) > x\}. \tag{15}$$

In implementing this procedure I used a grid of $x \in [-10, 10]$ with an interval of 10^{-3} and a grid of $\tau \in (0,1)$ with an interval of 10^{-5}.

Given an estimate of the CDF I then numerically simulated uniformly distributed random variables on the unit interval $[0,1]$ and used the integral probability transform to achieve approximations of realizations of the estimated tempered stable process. That is, if $F(.)$ is the CDF of some tempered stable distribution and if $u \sim U[0,1]$ then $x = F^{-1}(u)$ will be

distributed according to that tempered stable process.

5.3 Double-Exponential Jump-Diffusion Process

Jump-diffusion models are quite common in the financial econometrics literature. The jump-diffusion model I used here has jump sizes governed by a double-exponential (or Laplace) distribution. (Along with the normal distribution, the double-exponential distribution is a common distribution in the literature used to model the jump sizes.) The model takes the form

$$dX_t = \sigma dW_t + dJ_t$$

where W_t is a Brownian motion component and J_t is a compound Poisson jump process with intensity parameter λ and with jump sizes that follow a double-exponential (or Laplace) distribution with parameter β.

In performing the Monte Carlo study I found the parameter estimates for the jump-diffusion model were estimated with a much smaller bias and and a smaller mean absolute deviation from their true values if the parameter β for the double-exponential distribution and the jump intensity λ were pre-estimated and these pre-estimated parameters were used to create bounds on the parameter estimates in the later characteristic function estimation procedure. The pre-estimation was done by isolating the jumps using a jump detection procedure and then estimating the intensity λ and jump size parameter β of these isolated jumps. The jump detection was done by removing returns that were smaller in absolute value than $\alpha\Delta^{0.49}\sqrt{BV_t}$ where BV_t was a local estimator of the bipower variance of the process. (I found setting $\alpha = 3.75$ worked best in the Monte Carlo study.)

The parameters for the jump-diffusion model were chosen so that 80% of the quadratic variation would come from the diffusive component and 20% from the jump component and that there should be eight expected jumps each calendar year. These led to the following parameters choices $\sigma = 0.8944$, $\lambda = 0.0317$, and $\beta = 2.51$.

Table 1: Monte Carlo Results (Rejection Rates)

Null: Tempered Stable

	Rejection Rates		
True Model	10%	5%	1%
Tempered Stable	8.7%	5.7%	1.9%
Jump-Diffusion	100.0%	99.3%	97.3%

Null: Jump-Diffusion

	Rejection Rates		
True Model	10%	5%	1%
Tempered Stable	82.9%	82.1%	81.9%
Jump-Diffusion	10.0%	5.9%	1.2%

NOTE: Study based on 1000 replications. See Section 5 for the details of the study and the models used.

5.4 Results

Based on 1000 Monte Carlo replications Table 2 reports the quantiles of the parameter estimates and Table 1 reports the rejection rates of the testing procedure.

When the true model and null of the test are the same, the rejection rates for both models are very close to the theoretical values they should be, lending support to the testing procedure in this paper. When the true model and the null of the test differ the rejection rate are quite high showing the power of the test. When the true model is the simulated jump-diffusion process and the null of the test is a tempered stable model, the rejection rates vary from 100% at the 10% level to 97.3% at the 1% level. When the true model is the simulated tempered stable process and the null of the test is a jump-diffusion model, the rejection rates range from 82.9% at the 10% level to 81.9% at the 1% level. These results lend significant support to the power of the testing procedure in this paper.

In addition, the parameter estimates of the Monte Carlo study in Table 2 are all very close to their true values lending support to the accuracy of the estimation procedure.

Table 2: Monte Carlo Results (Parameter Estimates)

Tempered Stable

	True	Estimated Quantiles							
	Value	1st	5th	10th	Median	90th	95th	99th	MAD
α	1.2	0.98	1.09	1.12	1.26	1.50	1.56	1.63	0.13
c	43	14.20	19.09	22.85	38.17	49.99	53.21	64.75	9.40
ρ	1.15	0.57	0.71	0.80	1.08	1.25	1.30	1.41	0.16

Jump-Diffusion

	True	Estimated Quantiles							
	Value	1st	5th	10th	Median	90th	95th	99th	MAD
σ	0.8944	0.891	0.892	0.892	0.895	0.897	0.898	1.000	0.004
λ	0.0317	0.010	0.023	0.025	0.031	0.037	0.039	0.041	0.004
β	2.5100	1.888	2.062	2.116	2.537	2.998	3.056	3.433	0.261

NOTE: Study based on 1000 replications. MAD is the mean absolute deviation of the parameter estimates from their true values.

6 Empirical Results

As an empirical exercise I estimated and performed a specification test on the high-frequency returns of the E-mini S&P 500 futures (ES) and a selection of five large market capitalization stocks (AAPL, GE, JNJ, WMT, and CVX) under the null of a CGMY or tempered stable distribution and the null of a jump-diffusion distribution. With both distributions taking the functional forms described in the Monte Carlo study of Sections 5.1 and 5.3.

I used one minute pricing data on the E-mini S&P futures (ES) from the years 2007 to 2014 and, for the stocks, I used one minute pricing data from the trades and quotes (TAQ) database spanning the years 2005 to 2013. While ES contracts are traded 23 hours a day, to match the asset market of the stocks I only used data from 9:30am till 4:00pm. All pricing data are transaction prices and when no trades occurred prices are back-fitted from the most recent trade. In addition, I excluded any NYSE holidays and partial trading days. I performed the estimation and testing procedures over sampling frequencies of five, ten, fifteen, and twenty minutes.

Before performing the estimation and testing procedures I first scaled the returns by a estimate of their diurnal and stochastic variations. This was motivated by Section 2 in which

it was discussed how certain time varying stochastic processes might be transformed into Lévy processes by scaling these processes by their variation. It is also a standard practice in estimating high-frequency asset price returns. The diurnal variation has been explored extensively in the literature, see Andersen and Bollerslev (1997) for an initial treatment. For each recurring daily high frequency interval I estimated the variation at this interval and then used the mean across days as my estimate of the diurnal variation.

To estimate the stochastic variation I followed Todorov and Tauchen (2014) and used a local block estimator of the bipower variation. Let $r_{t,i} = p_{t,i} - p_{t,i-1}$ be the within day return where t indexes the day and i indexes the within day interval and $p_{t,i}$ is the log-price of the asset. With this notation the estimator of the block bipower variation can be expressed as

$$BV_{t,j} = \frac{\pi}{2} \frac{n}{k_n - 1} \sum_{i=(j-1)k_n+2}^{jk_n} |r_{t,i}||r_{t,i-1}| \tag{16}$$

where n is the number of daily intervals, k_n is the block size, and j indexes the blocks.[5]

Both the diurnal variation and the stochastic variation were estimated separately for each sampling interval. So, for example, the one minute returns were scaled by estimates of the variations based on one minute return data and the twenty minute returns were scaled by estimates of the variations based on twenty minute return data. The rational for doing so was to keep the entire estimation and testing process separate across sampling frequencies. In practice, most researchers first select a sampling frequency and then attempt to estimate and test their model. I wanted to keep the procedure as closely in line with common research practices as possible.

The results of the estimation and testing procedures are given in Table 3. This table lists the p-values for the null of a CGMY or tempered stable process and the null of a double-exponential jump-diffusion process. The p-values were calculated based on 10,000 replications of the estimated covariances of the test statistics following Section 4. Parameter estimates and their standard errors are available upon request. Based on the p-values of the specification test, the jump-diffusion model is strongly rejected at all sampling frequencies

[5]To match the block sizes in Todorov and Tauchen (2014) I set $k_n = \lfloor 3.1n^{0.49} \rfloor$.

Table 3: Empirical Results (p-values)

Null: Tempered Stable

	Sampling Frequency			
	5min	10min	15min	20min
ES	0.52	0.33	1.00	0.62
AAPL	1.00	1.00	1.00	1.00
GE	0.00	0.44	0.45	0.53
JNJ	0.00	0.27	0.94	1.00
WMT	0.00	0.18	0.99	0.99
CVX	1.00	1.00	1.00	1.00

Null: Jump-Diffusion

	Sampling Frequency			
	5min	10min	15min	20min
ES	0.00	0.00	0.00	0.00
AAPL	0.00	0.00	0.00	0.00
GE	0.00	0.00	0.00	0.00
JNJ	0.00	0.00	0.00	0.55
WMT	0.00	0.00	0.00	0.00
CVX	0.00	0.00	0.00	0.00

NOTE: The p-values are for the specification test outlined in Section 4. They are based on 10,000 replications of the estimated covariances. The ES data are data from the E-mini S&P 500 futures and span the years 2007 to 2014. The data on the stocks AAPL, GE, JNJ, WMT, and CVX come from the TAQ database and span the years 2005 to 2013. See Section 6 for a more details on the data.

and for all assets (with the sole exception of the twenty-minute returns of JNJ). However, with the exception of the five-minute returns of GE, JNJ, and WMT, I was unable to reject the null of a tempered stable model at the 1%, 5%, or even 10% level for all of the assets and for all the sampling frequencies considered. As far as the author is aware, this is the first time the high-frequency returns of a security have been successfully fitted to a pure jump model.

7 Empirical Application: Intraday Values at Risk

As mentioned in the introduction intraday risks are becoming an increasingly important topic of concern following the recent financial crisis. The Basel III guidelines for example give a great deal of concern to the intraday liquidity risks financial firms take on. Given the growth in day trading activities, it makes sense to monitor intraday market risks as well,

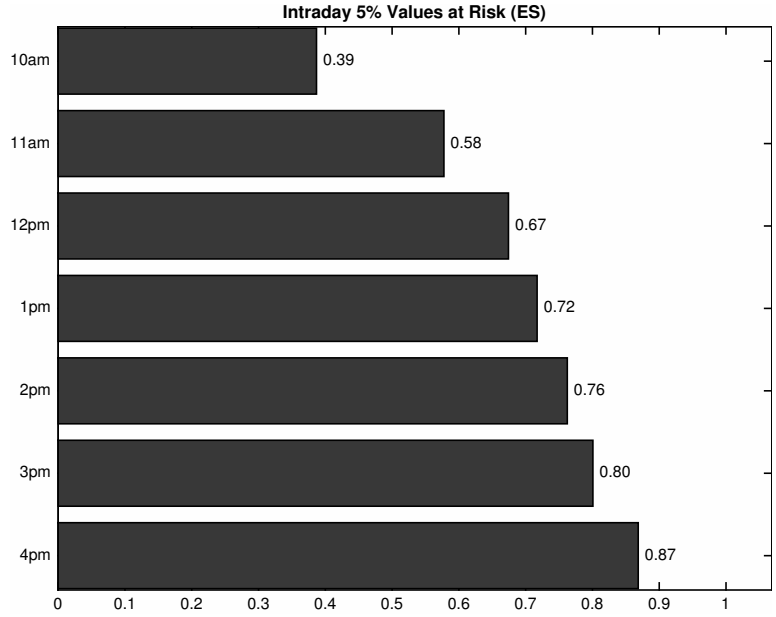
especially given that there exists evidence that firms do not fully monitor their intraday trading activities. (See, for example, the 2003 SEC report, "Special Study: Report of Examinations of Day-Trading Broker-Dealers", in which it was found that most firms did not monitor the capital compliance of their day trading activities.) A flexible intraday value at risk would be an appropriate tool for monitoring intraday market risks and its derivation using the methods developed in this paper is the subject of this empirical exercise.

For readers unfamiliar with the value at risk, the value at risk is defined as the smallest number l such that the losses L will be no larger than l with probability $1-\alpha$, where $\alpha \in (0,1)$ is a given confidence level. One may define such a concept as

$$VaR_\alpha(L) \equiv \inf\{l \in \mathbb{R} : \mathbb{P}(L > l) \leq 1-\alpha\}. \tag{17}$$

Since I estimate and simulate asset price returns on a high frequency scale the methods here can easily be used to calculate the value at risk at any high frequency interval during the day. One simply needs to look at the accumulated high frequency returns up to whatever horizon is of interest. With this idea in mind I forecasted intraday values at risk for a position in the E-mini S&P futures (ES) using five minute simulated return intervals and then calculated the 1%, 5%, and 10% values at risk for positions throughout the day. (The next subsection provides the details on how these calculations were performed.) Figure 2 shows an example of one set of such forecasted intraday values at risk. The figure plots the 5% values at risk for a position in ES that opened at 9:35am on the morning of November 11th, 2014 and would be held until the times listed on the vertical axis. The values at risk on the horizonal axis are in terms of the percentage of the position that could be lost. So, for example, a trader that opened a position in ES at 9:35am and held it until 2pm that day would, with 95% confidence, face potential losses of no more than 0.76% of his or her investment.

Figure 2: Intraday Values at Risk



NOTE: The value at risk is in percentage terms for a position that opened at 9:35am on November 11, 2014 up until the point on the vertical axis. See Section 7 for details on the calculation.

7.1 Estimation

Recall that in the estimation of the high-frequency returns I first scaled the returns by an estimate of their diurnal and stochastic variations. Given the strong support for using the tempered stable distribution to model the high-frequency returns of ES in Section 6, I used that model here to simulate the scaled high-frequency returns. Because the model estimates and simulates returns that have been scaled by their variation to use the model in practice I needed to simulate not just the increments of the estimated tempered stable process that was estimated, but also replace the variation that had been scaled out. To do so I needed to forecast and simulate an estimate of the block bipower volatility as discussed in Section 6 and replace the diurnal variation.

To simulate the tempered stable process I used the same methods as in the Monte Carlo study of Section 5.2. This involved a Fourier inversion of the tempered stable characteristic function to estimate the cumulative distribution function (CDF). While Fourier inversion

can often perform poorly in estimation procedures, my simulation only required one Fourier inversion rather than the multitudes often required in most estimation procedures. The results of the estimation in Section 6 gave me an estimate of the diurnal variation. Given the coarseness of these estimates I smoothed the diurnal variation using a two sided moving average model with two leads and lags. To forecast the stochastic variation required more care and is detailed below.

7.2 An HAR Model for the Block Bipower Volatility

To simulate and forecast the block bipower volatility I used a modification of Corsi's HAR model in Corsi (2009). To avoid confusion define the block bipower variation as in equation (16) and the bipower volatility as its square-root, *i.e.*, $BVol_{t,j} = \sqrt{BV_{t,j}}$.

Letting B equal the number of daily blocks of the block bipower variation in equation (16) of Section 6. I modeled the log bipower volatility as following the heterogeneous autoregressive model.

$$\begin{aligned} \log BVol_{t,j+1} = \beta_0 + \beta_1 \log BVol_{t,j} + \cdots + \beta_B \log BVol_{t-1,j+1} \\ + \beta_{B+1} \log BVol^{(w)}_{t,j} + \beta_{B+2} \log BVol^{(m)}_{t,j} + \epsilon_{t,j+1} \end{aligned} \tag{18}$$

where

$$\log BVol^{(w)}_{t,j} = \frac{1}{5B} \sum_{\tau,b=t-5,j}^{t,j} \log BVol_{\tau,b} \tag{19}$$

and

$$\log BVol^{(m)}_{t,j} = \frac{1}{22B} \sum_{\tau,b=t-22,j}^{t,j} \log BVol_{\tau,b}. \tag{20}$$

The motivation for the model in equation (18) follows Corsi (2009). Short-term traders should influence the bipower volatility over the day and therefore there should be information in the preceding blocks of the bipower volatility over the previous day, whereas medium-term and long-term traders might only affect the bipower volatility in levels over the weekly and monthly horizon. Using one full trading day's worth of preceding blocks of the bipower volatility helps to capture possible diurnal patterns in the bipower volatility as well.

In addition, I estimated the model in logs to avoid negativity issues in forecasting, however due to the convexity of the log-transform this creates a bias in the forecasts of the bipower volatility because

$$\mathbb{E}_{t,j}[BVol_{t,j+1}] \neq \exp\{\mathbb{E}_{t,j}[\log BVol_{t,j+1}]\}. \tag{21}$$

Bollerslev et al. (2009) provides evidence that the log bipower volatility of the S&P 500 might be approximately normal. If this is the case then the bipower volatility should be approximately log-normally distributed and we can correct for the bias by subtracting an estimate of $\sigma^2/2$ from the log bipower volatility, where σ^2 is the variation of $\epsilon_{t,j+1}$ in equation (18). That is, if the bipower volatility is log-normally distributed then,

$$\mathbb{E}_{t,j}[BVol_{t,j+1}] = \exp\{\mathbb{E}_{t,j}[\log BVol_{t,j+1} - \hat{\sigma}^2/2]\}. \tag{22}$$

Given this insight I needed to estimate the conditional variance of log bipower volatility as well. The estimate I used is simply the mean of the squared residuals from the estimation in equation (18). A more thorough analysis of the block bipower volatility might look at ways to account for heterogeneity or autocorrelation in the residuals, but given the scope of this paper I did not. (Parameter estimates for the HAR model estimated here are available upon request.)

7.3 Backtesting

I performed an extensive backtest to test the performance of my calculated values at risk. To do so I calculated intraday values at risk for a position in ES over a seven year period making sure that no future information would ever be used in the forecasts.

An important difference here was that I estimated the tempered stable distribution on returns that had been scaled by forecasted estimates of the diurnal and stochastic variations rather than realized estimates of these variations. Doing so is important because the distribution of the high-frequency returns scaled by the forecasts of the variations might not

Table 4: Coverage (Intraday Values at Risk)

	Value at Risk			
period	**1%**	**5%**	**10%**	N
all-periods	98.12%	94.64%	90.87%	4458
morning	99.06%	95.83%	92.19%	1486
midday	97.24%	93.27%	89.43%	1486
afternoon	98.05%	94.82%	90.98%	1486

NOTE: This table lists the percentage of time the actual realized losses were smaller than the forecasted value at risk. See Section 7.3 for the starting and closing times of the holding periods. The calculations are for a position in the E-mini S&P futures (ES). N is the number of forecasted intraday values at risk. The calculations span the years 2009 till 2014.

follow the same distribution as the high-frequency returns scaled by the realized variations. The reason for doing so is that in calculating the value at risk I would need to use forecasts of the variations and not their realized estimates as these would be unavailable in practice.

The dataset for the E-mini S&P futures (ES) spanned the years 2007 to 2014. I used a one year rolling window on all the estimates and updated the parameters for the tempered stable model, the HAR model of block bipower variation, and the estimate of the diurnal variation every month. To give an example of how this would work consider calculating an intraday value at risk on January 11th, 2014. The parameters for all the models would have been estimated on data from January 1, 2013 till December 31, 2013. The block bipower variation though would be updated in real time so that in forecasting the block bipower the forecast would be based on data up until January 11th, but the parameters would not be updated again until February 1st.

Within each day I calculated three intraday values at risk: a morning period lasting from 9:35am till 11:45am, a midday period lasting from 11:45am till 1:55pm, and an afternoon period lasting from 1:55pm till 4:00pm. For each period I calculated a 1%, 5%, and 10% value at risk. Finally, I compared each value at risk with the actual realized losses or gains over that period to check the coverage of my estimates. With the rolling window I was able to calculate values at risk starting in 2009 until 2014. This gave 1486 days worth of calculations and 4458 total value at risk calculations. Table 4 lists the coverage results of this study. The coverage is remarkably close to the theoretical values. Taking all the within day periods together, the 1% value at risk has a coverage of 98.12%, the 5% value at risk

has a coverage of 94.64%, and the 10% value at risk has a coverage of 90.87%. The coverage results are similar looking at the within day periods separately.

8 Conclusion

This paper outlined a method for modeling high-frequencies asset price dynamics. It did so by assuming asset prices followed a class of Itô semi-martingales and showed how if the existence of a certain transformation of the jump measure of these Itô semi-martingales existed that by scaling asset returns by their variation that asset returns could be modeled as Lévy processes. Once in the class of Lévy processes this paper showed how the characteristic function of these assets could be derived in closed form and, given knowledge of the characteristic function, this paper developed a novel estimation technique and a novel test of a model's specification. It did so by estimating and testing the characteristic function of the model over a family of weighting functions where the weighting functions were chosen to provide flexibility in terms of the informational content of the characteristic functions of the models under consideration. A Monte Carlo study lent strong support to the consistency of the estimation strategy and the size and power of the specification test.

Having developed a novel estimation and testing procedure this paper turned to the estimation and model specification testing of six assets in an empirical study. One particular class of models, the CGMY model of Carr et al. (2002) or equivalently the tempered stable models in Rosinski (2007), was found to fit these assets quite strongly. As far as the author is aware, this is the first time the high-frequency returns of an asset have been successfully fitted to a pure jump model (as the tempered stable or CGMY models are). Finally, an empirical exercise was performed on the high-frequency returns from the E-mini S&P futures (ES) to calculate intraday values at risk. The exercise developed a method to forecast the intraday value at risk of an asset at any within day horizon and a backtest of the estimated values at risk showed their coverage to be right in line with the theoretically expected values.

A Proofs

Proof of Lemma 1. Let X be a continuously distributed random variable and assume $\phi(u) \neq \phi_0(\theta_0, u)$ as in the statement of the lemma. Notice $m(\theta_0, u) = c(u) - c_0(u, \theta_0) + s(u) - s_0(u, \theta_0) = \cos(uX) - c_0(u, \theta_0) + \sin(uX) - s_0(u, \theta_0)$.

We need to show that the set of events in which $\cos(uX) - c_0(u, \theta_0) + \sin(uX) - s_0(u, \theta_0) = 0$ has measure zero. Since $c_0(u, \theta_0)$ and $s_0(u, \theta_0)$ are given by the model, they are non-random real numbers. When $\phi(u) = \phi_0(u, \theta_0)$ we have $\mathbb{E}[\cos(uX)] = c_0(u, \theta_0)$ and $\mathbb{E}[\sin(uX)] = s_0(u, \theta_0)$, however since X is continuously distributed when $\phi(u) \neq \phi_0(u, \theta_0)$ we will have $\cos(uX) = c_0(u, \theta_0)$ and $\sin(uX) = s_0(u, \theta_0)$ only on a set of measure zero.

We will next show that $\mathbb{P}[\cos(uX) = -\sin(uX)] = 0$ so that even if $c_0(u, \theta_0) = -s_0(u, \theta_0)$ we would have $m(\theta_0, u) = 0$ only with probability zero. Since $\cos(x) = -sin(x)$ only on a countable set of $x \in \mathbb{R}$ and X is a continuously distributed random variable, we see that for any given $u \in \mathbb{R}$ that $\cos(uX) = -\sin(uX)$ only on a countable set, and therefore $\mathbb{P}[\cos(uX) = -\sin(uX)] = 0$.

Using a similar line of reasoning, one can easily show that $m(\theta_0, u) = 0$ only with probability zero for any other possible combination. □

Proof of Theorem 2. Let $m(\theta, u)$ be given as in the theorem for some fixed θ. We will show that if $\int_{\mathbb{R}} m(\theta, u) g_k(u) du = 0$ for all $k \in \mathcal{K}$, then $m(\theta, u) = 0$ for a.e. $u \in \mathbb{R}$.

Let S be a given random variable on $\mathbb{R}$ such that $S \neq 0$ and define $U \equiv m(\theta, S)$. The proof will follow Bierens (1990) by showing that $\mathbb{P}(\mathbb{E}[U|S] = 0) < 1$ implies that the set $\mathcal{S} = \{k \in \mathcal{K} : \mathbb{E}[U g_k(S)] = 0\}$ has Lebesgue measure zero. Assume $\mathbb{P}(\mathbb{E}[U|S] = 0) < 1$. Theorem 2 in Bierens (1982) states that $\mathbb{P}(\mathbb{E}[U|S] = 0) < 1$ if and only if there exists some non-negative integer m such that $\mathbb{E}[US^m] \neq 0$. Applying this theorem, let m be such that $\mathbb{E}[US^m] \neq 0$. We will use this to show first that if $\mathbb{E}[U g_k(S)] = 0$, in a neighborhood of $\mathbb{E}[U g_k(S)] = 0$, we will have $\mathbb{E}[U g_k(S)] \neq 0$.

For any given $p \in \mathbb{N}$ it can be shown

$$\begin{aligned}\left(\frac{d}{dk}\right)^p \mathbb{E}[Ug_k(S)] &= \left(\frac{d}{dS}\right)^p \mathbb{E}[U(Sk)^2 \exp\{-(Sk)^2\}] \\ &= \sum_{j=N}^{\infty} \frac{(-1)^j}{j!} \left(\prod_{i=0}^{p-1} 2j+2-i\right) \mathbb{E}[Uk^{2j+2-p}S^{2j+2}]\end{aligned} \tag{23}$$

where $N = 0$ if $p = 0, 1, 2$ and $N = \lceil p/2 \rceil - 1$ otherwise. Notice that when p is odd that $2N + 2 - p = 0$ so that $k^{2N+2-p} = 1$. This implies for an odd p

$$\left(\frac{d}{dk}\right)^p \mathbb{E}[Ug_k(S)] \to \frac{(-1)^N}{N!} \left(\prod_{i=0}^{N-1} 2N+2-i\right) \mathbb{E}[US^{2N+2}] \text{ as } k \to 0. \tag{24}$$

If we chose p such that $m = 2N + 2$ we will have $\mathbb{E}[Ug_k(S)] \neq 0$ in a neighborhood of $\mathbb{E}[Ug_k(S)] = 0$ since the limit above shows that the p-th derivative of $\mathbb{E}[Ug_k(S)]$ will be nonzero.

Now let $k_0 \neq 0$ be such that $\mathbb{E}[Ug(k_0S)] = 0$. We want to show that in a neighborhood of $k = k_0$ that $\mathbb{E}[Ug(kS)] \neq 0$. Let $k \equiv k_0 + \epsilon$ for some small $\epsilon \neq 0$. Notice we can decompose $g(kS)$ as

$$\begin{aligned} g(kS) = g[(k_0+\epsilon)S] &= [(k_0+\epsilon)S]^2 \exp\{-[(k_0+\epsilon)S]^2\} \\ &= g(k_0S)\exp\{-2k_0\epsilon S^2 - (\epsilon S)^2\} + (2k_0\epsilon S^2)\exp\{-[(k_0+\epsilon)S]^2\} \\ &+ g(\epsilon S)\exp\{-(k_0S)^2\}\exp\{-2k_0\epsilon S^2\}. \end{aligned} \tag{25}$$

Consider the above decomposition as it relates to $\mathbb{E}[Ug(kS)]$. First, since $\mathbb{E}[Ug(k_0S)] = 0$, we will have $\mathbb{E}[Ug(k_0S)\exp\{-2k_0\epsilon S^2 - (\epsilon S)^2\}] = 0$. Second, consider $\mathbb{E}[U(2k_0\epsilon S^2)\exp\{-[(k_0+\epsilon)S]^2\}]$. In a neighborhood of $\epsilon = 0$ we will have $(2k_0\epsilon S^2)\exp\{-[(k_0+\epsilon)S]^2\} \neq 0$ since $k_0 \neq 0$ and $S \neq 0$. Because of this, the behavior of this second component around zero then will depend on value of U. Third, and finally, since

$$\mathbb{P}(\mathbb{E}[U\exp\{-(k_0S)^2\}\exp\{-2k_0\epsilon S^2\}|S] = 0) = \mathbb{P}(\mathbb{E}[U|S] = 0) < 1$$

if we replace U by $U\exp\{-(k_0S)^2\}\exp\{-2k_0\epsilon S^2\}$ in (24) we see as $\to 0$ that

$$\mathbb{E}[U\exp\{-(k_0S)^2\}\exp\{-2k_0\epsilon S^2\}g(\epsilon S)] \neq 0 \tag{26}$$

in a neighborhood of $\epsilon = 0$. Therefore while the first part of the decomposition of $\mathbb{E}[Ug(kS)]$ is zero and the second is undetermined, by the third part we know $\mathbb{E}[Ug(kS)] \neq 0$ in a neighborhood $k = k_0$.

Recall $\mathcal{S} = \{k \in \mathcal{K} : \mathbb{E}[Ug_k(S)] = 0\}$. The above result implies $\inf_{k\in\mathcal{S},k\neq k_0} |k - k_0| > 0$ if $k_0 \in \mathcal{S}$ and hence that $\mathcal{S}$ is countable. Since a countable set has Lebesgue measure zero the result follows. □

Proof of Theorem 3. The proof of Theorem 3 follows by a series of lemmas. First define

$$Z_n(k) = \sqrt{n}[Q_{n,k}(\theta_0) - (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k}(\theta_0)]. \tag{27}$$

Lemma 2. *Under Ω_0 and Assumption A we have that for any $k_1, k_2 \in \mathcal{K}$, $(Z_n(k_1), Z_n(k_2))$ converges to a multivariate normal with covariance Γ where the elements of Γ have the form $\Gamma(k_1, k_2)$ given in the proof below.*

Proof. Since

$$Z_n(k) = \sqrt{n}[Q_{n,k}(\theta_0) - (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k}(\theta_0)]$$

we have

$$\begin{aligned}
\Gamma(k_1, k_2) &= \operatorname{cov}[Q_{n,k_1}, Q_{n,k_2}] \\
&\quad + \operatorname{cov}[(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}, (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}] \\
&\quad - \operatorname{cov}[(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}, Q_{n,k_2}] \\
&\quad - \operatorname{cov}[Q_{n,k_1}, (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}].
\end{aligned} \tag{28}$$

(Where the dependence on θ_0 has been suppressed above.)

Consider first $\text{cov}[Q_{n,k_1}, Q_{n,k_2}]$. Notice for any $k \in \mathcal{K}$ we have

$$\begin{aligned} Q_{n,k} &= \int [\{c_n(u) - c_0(u,\theta)\} + \{s_n(u) - s_0(u,\theta)\}] g_k(u) du \\ &= \frac{1}{n} \sum_{j=1}^{n} \int [\cos(x_j u) - c_0(u,\theta) + \sin(x_j u) - s_0(u,\theta)] g_k(u) du. \end{aligned} \tag{29}$$

From (29) we can see

$$\begin{aligned} \text{cov}[Q_{n,k_1}, Q_{n,k_2}] = \int \int \{ &\text{cov}[\cos(xu), \cos(xv)] \\ &+ 2\text{cov}[\cos(xu), \sin(xv)] + \text{cov}[\sin(xu), \sin(xv)]\} g_{k_1}(u) g_{k_2}(v) du dv. \end{aligned} \tag{30}$$

Next, consider $\text{cov}[(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}, (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}]$. Define

$$I_{n,k}(\theta) \equiv \int |\hat{\phi}_n(u) - \phi_0(u,\theta)|^2 g_k(u) du \tag{31}$$

and let $d = \dim(\theta)$. Then for any $k \in \mathcal{K}$ we have

$$(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k} = \sum_{i=1}^{d} (\hat{\theta}^i_{n,k^*} - \theta^i_0) \nabla_{\theta^i} Q_{n,k} \tag{32}$$

where θ^i is the i-th element of θ. Since $\hat{\theta}^i_{n,k^*} - \theta^i_0 = -[\nabla_{\theta^i \theta'^i} I_{n,k^*}]^{-1} [\nabla_{\theta^i} I_{n,k^*}]$ we see $\text{cov}[\hat{\theta}^i_{n,k^*} - \theta^i_0, \hat{\theta}^i_{n,k^*} - \theta^i_0] = (\Lambda^{ii}_{k^*})^{-1} \Sigma^{ij}_{k^*} (\Lambda^{jj}_{k^*})^{-1}$ where $\mathbb{E}[\nabla_{\theta\theta'} I_{n,k}] \equiv \Lambda_k$,

$$\begin{aligned} \Sigma_k = \int \int \{ &\text{cov}[\cos(xu), \cos(xv)] \nabla_\theta c_0(u) \nabla_\theta c_0(v) \\ &+ 2\text{cov}[\cos(xu), \sin(xv)] \nabla_\theta c_0(u) \nabla_\theta s_0(v) \\ &+ \text{cov}[\sin(xu), \sin(xv)] \nabla_\theta s_0(u) \nabla_\theta s_0(v)\} g_k(u) g_k(v) du dv, \end{aligned} \tag{33}$$

and $\Lambda^{ii}_{k^*}$ is the ii-th element of Λ and $\Sigma^{jj}_{k^*}$ is the ij-th element of Σ_{k^*}.

This implies

$$\begin{aligned} &\text{cov}[(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}, (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}] \\ &\qquad = \sum_{i=1}^{d} \sum_{j=1}^{d} a^i_{k_1} a^j_{k_2} (\Lambda^{ii}_{k^*})^{-1} \Sigma^{ij}_{k^*} (\Lambda^{jj}_{k^*})^{-1} \end{aligned} \tag{34}$$

where $a_k \equiv \mathbb{E}[\nabla_\theta Q_{n,k}]$ and a_k^i is the i-th element of a_k.

Now consider $\text{cov}[(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}, Q_{n,k_2}]$. Notice,

$$\begin{aligned}
\text{cov}[(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}, Q_{n,k_2}] &= \sum_{i=1}^{d} \text{cov}[(\hat{\theta}^{i*}_{n,k^*} - \theta_0^i)\nabla_{\theta^i} Q_{n,k_1}, Q_{n,k_2}] \\
&= \sum_{i=1}^{d} a^i_{k_1} \text{cov}[\hat{\theta}^i_{n,k^*} - \theta_0^i, Q_{n,k_2}] \\
&= \sum_{i=1}^{d} -a^i_{k_1} (\Lambda^{ii}_{k^*})^{-1} \text{cov}[\nabla_{\theta^i} I_{n,k^*}, Q_{n,k_2}].
\end{aligned}$$

Since,

$$\begin{aligned}
\nabla_{\theta^i} I_{n,k^*} = -2\frac{1}{n}\sum_{j=1}^{n} \int [\{\cos(x_j u) - c_0(u)\}\nabla_{\theta^i} c_0(u) \\
+ \{\sin(x_j u) - s_0(u)\}\nabla_{\theta^i} s_0(u)] g_{k^*}(u) du
\end{aligned}$$

we see

$$\begin{aligned}
\text{cov}[\nabla_{\theta^i} I_{n,k^*}, Q_{n,k_2}] = -2 \int\int [&\text{cov}\{\cos(xu), \cos(xv)\}\nabla_{\theta^i} c_0(u) \\
&+ \text{cov}\{\cos(xu), \sin(xv)\}\nabla_{\theta^i} c_0(u) \\
&+ \text{cov}\{\sin(xu), \cos(xv)\}\nabla_{\theta^i} s_0(u) \\
&+ \text{cov}\{\sin(xu), \sin(xv)\}\nabla_{\theta^i} s_0(u)] g_{k^*}(u) g_{k_2}(v) du dv.
\end{aligned}$$

By a similar argument we see

$$\text{cov}[Q_{n,k_1}, (\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}] = \sum_{i=1}^{d} -\text{cov}[Q_{n,k_1}, \nabla_{\theta^{i*}} I_{n,k^*}] a^i_{k_2} (\Lambda^{ii}_{k^*})^{-1}$$

and

$$\begin{aligned}
\text{cov}[Q_{n,k_1}, \nabla_{\theta^i} I_{n,k^*}] = -2 \int\int [&\text{cov}\{\cos(xu), \cos(xv)\}\nabla_{\theta^i} c_0(v) \\
&+ \text{cov}\{\cos(xu), \sin(xv)\}\nabla_{\theta^i} c_0(v) \\
&+ \text{cov}\{\sin(xu), \cos(xv)\}\nabla_{\theta^i} s_0(v) \\
&+ \text{cov}\{\sin(xu), \sin(xv)\}\nabla_{\theta^i} s_0(v)] g_{k_1}(u) g_{k^*}(v) du dv.
\end{aligned}$$

□

Lemma 3. *Given Assumption* (A) *and under* Ω_0 *we have*

$$\sup_{k\in\mathcal{K}} |\sqrt{n}Q_{n,k}(\hat{\theta}_{n,k^*}) - Z_n(k)| \xrightarrow{p} 0.$$

Proof. By the mean value theorem we can write

$$\sqrt{n}Q_{n,k}(\hat{\theta}_{n,k^*}) = \sqrt{n}Q_{n,k}(\theta_0) + \sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k}(\bar{\theta}). \tag{35}$$

Where

$$\nabla_\theta Q_{n,k}(\theta) = -\int [\nabla_\theta c_0(u,\theta) + \nabla_\theta s_0(u,\theta)] g_k(u) du \equiv a_k(\theta) \tag{36}$$

and $\bar{\theta} \in N_\delta(\theta_0)$ where $\delta = |\hat{\theta}_{n,k^*} - \theta_0|$. Notice $a_k(\theta)$ does not depend on n, so that $\nabla_\theta Q_{n,k}(\bar{\theta}) \xrightarrow{p} a_k(\theta_0) \equiv a_k$. Further, the uniform convergence of $\hat{\theta}_{n,k^*}$ to θ_0 implies

$$\sup_{k\in\mathcal{K}} |\nabla_\theta Q_{n,k}(\bar{\theta}) - a_k| \xrightarrow{p} 0. \tag{37}$$

By a central limit theorem we see that $\sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)$ converges uniformly to a mean zero random variable with variance-covariance matrix $\Lambda_{k^*}^{-1}(\theta_0)^\top \Sigma_{k^*}(\theta_0)\Lambda_{k^*}^{-1}(\theta_0)$. (The definitions of $\Lambda_k(\theta)$ and $\Sigma(\theta)$ can be found in Lemma 1) Also notice,

$$\begin{aligned}\sqrt{n}Q_{n,k}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int [\cos(x_j u) - c_0(\theta_0, u) \\ + \sin(x_j u) - s_0(\theta_0, u)] g_k(u) du.\end{aligned} \tag{38}$$

Since the sine and cosine functions are bounded and the characteristic function of any random variable is bounded we know by a central limit theorem that $\sqrt{n}Q_{n,k}(\theta_0)$ converges uniformly to a mean zero normal random variable with variance equal to

$$\begin{aligned}V_k \equiv \int\int (\text{cov}[\cos(xu), \cos(xv)] + 2\text{cov}[\cos(xu), \sin(xv) \\ + \text{cov}[\sin(xu), \sin(xv)]) g_k(u) g_k(v) du dv.\end{aligned} \tag{39}$$

Combining the above results the lemma follows. □

The next lemma shows that the process $Z_n(k)$ is tight in $\mathcal{C}(\mathcal{K})$

Lemma 4. *Given Assumption* (A) *and under* Ω_0*,* Z_n *is tight.*

Proof. Following Theorem 8.1 in Billingsley (1968) it suffices to prove two conditions:

(C.1) for each $\delta > 0$ and an arbitrary $k_0 \in \mathcal{K}$, there exists an $\epsilon > 0$ such that $\sup_n \mathbb{P}(|Z_n(k_0)| > \epsilon) \leq \delta$; and,

(C.2) for each $\delta > 0$ and $\epsilon > 0$, there exists an $\xi > 0$ such that

$$\sup_n \mathbb{P}\left(\sup_{|k_1-k_2|<\xi} |Z_n(k_1) - Z_n(k_2)| \geq \epsilon\right) \leq \delta.$$

Condition (C.1) follows from the fact that $Z_n(k_0) \xrightarrow{d} N(0, \Gamma)$ (cf., Lemma 2).

To prove condition (C.2) first observe

$$\begin{aligned} |Z_n(k_1) - Z_n(k_2)| \leq |\sqrt{n}Q_{n,k_1}(\theta_0) - \sqrt{n}Q_{n,k_2}(\theta_0)| + \\ |\sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}(\theta_0) - \sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}(\theta_0)|. \end{aligned} \tag{40}$$

Since

$$\begin{aligned} \sqrt{n}Q_{n,k_1}(\theta_0) - \sqrt{n}Q_{n,k_2}(\theta_0) = \frac{1}{n}\sum_{j=1}^{n} \int [\cos(x_j u) - c_0(\theta_0, u) \\ + \sin(x_j u) - s_0(\theta_0, u)][g_{k_1}(u) - g_{k_2}(u)]du, \end{aligned} \tag{41}$$

and the sine function, the cosine function, and any characteristic function is bounded in $[-1, 1]$, in order to bound $|\sqrt{n}Q_{n,k_1}(\theta_0) - \sqrt{n}Q_{n,k_2}(\theta_0)|$ we need to bound $|g_{k_1}(u) - g_{k_2}(u)|$.

Since $|g_{k_1}(u) - g_{k_2}(u)| \leq |g_{k_1}(u)| + |g_{k_2}(u)|$ it suffices to prove $g_k(u)$ is bounded for any $k \in \mathcal{K}$. Recall $g_k(u) = (ku)^2 \exp\{-(ku)^2\}$. A quick inspection of this function reveals it to lie on $[0, e^{-1}]$ for any $ku \in \mathbb{R}$. Given $u \in \mathbb{R}$ and $k \in \mathcal{K} \subset \mathbb{R}$ we can conclude $g_k(u)$ is bounded for any u or k.

Since $\sqrt{n}(\hat{\theta}_{n,k} - \theta_0)$ converges uniformly for any $k \in \mathcal{K}$, we know for some $\eta > 0$ and some $N_{k^*}(\eta)$ that for any $n > N_{k^*}(\eta)$ we will have $|\sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)| < \eta$. This implies for

$n > N_{k^*}(\eta)$ that

$$\begin{aligned}
&|\sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}(\theta_0) - \sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}(\theta_0)| \\
&< |\sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_1}(\theta_0)| + |\sqrt{n}(\hat{\theta}_{n,k^*} - \theta_0)^\top \nabla_\theta Q_{n,k_2}(\theta_0)| \qquad (42) \\
&\leq |\eta| \left(|\nabla_\theta Q_{n,k_1}(\theta_0)| + |\nabla_\theta Q_{n,k_2}(\theta_0)| \right).
\end{aligned}$$

Recall for $k = \{k_1, k_2\}$ that

$$\nabla_\theta Q_{n,k}(\theta_0) = -\int [\nabla_\theta c_0(u, \theta_0) + \nabla_\theta s_0(u, \theta_0)] g_k(u) du.$$

Since $c_0(u, \theta_0)$ and $s_0(u, \theta_0)$ are the real and imaginary parts of the tempered stable characteristic function respectively, and the tempered stable characteristic function is uniformly continuous, the derivatives $\nabla_\theta c_0(u, \theta_0)$ and $\nabla_\theta s_0(u, \theta_0)$ must be bounded. Given this fact, and the result shown earlier that $g_k(u)$ is bounded, we can conclude (42) is bounded.

By bounding the terms in (42) and (41) condition (C.2) follows. □

By an application of Lemma 2 and Lemma 3 it follows that for an arbitrary set $\{k_1, \ldots, k_q\}$ in $\mathcal{K}$ that the vector $(Z_n(k_1), \ldots, Z_n(k_q))$ converges in distribution to $(Z(k_1), \ldots, Z(k_q))$. Together with Lemma 4 this implies Z_n converges weakly in $\mathcal{K}$ to the process Z (cf., Billingsley, 1968, p. 47). Since $|.|$ and $\sup_{k \in \mathcal{K}}(.)$ are continuous mappings from $\mathcal{C}(\mathcal{K})$ to $\mathcal{C}(\mathcal{K})$ the proof follows by Lemma 2 and Theorem 5.1 in Billingsley (1968). □

Acknowledgments

The author is incredibly grateful for all the extremely helpful advice from George Tauchen and is greatly indebted to Tim Bollerslev, Jia Li, and Andrew Patton as well as the entire financial econometrics group at Duke University for their helpful discussions and advice.

References

Andersen, T. G. and Bollerslev, T. (1997), "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4, 115–158.

Bierens, H. J. (1982), "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

— (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

Bierens, H. J. and Ploberger, W. (1997), "Asympototic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Billingsley, P. (1968), *Convergence of Probability Measures*.

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. (2009), "A discrete-time model for daily S&P500 returns and realized variations : Jumps and leverage effects," *Journal of Econometrics*, 150, 151–166.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002), "The Fine Structure of Asset Returns : An Empirical Investigation," *Journal of Business*, 75, 305–332.

Carrasco, M. and Florens, J.-P. (2000), "Generalization of GMM To a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834.

Carrasco, M. and Florens, J.-p. (2002), "Efficient GMM Estimation Using the Empirical Characteristic Function," *Working Paper*.

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2010), "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Cont, R. and Tankov, P. (2004), *Financial Modelling With Jump Processes*.

Corsi, F. (2009), "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196.

Feuerverger, A. and McDunnough, P. (1981), "On the Efficiency of Empirical Characteristic Function Procedures," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 43, 20–27.

Heathcote, C. R. (1977), "The integrated squared error estimation of parameters," *Biometrika*, 64, 255–264.

Li, J., Todorov, V., and Tauchen, G. (2013), "Estimating the Volatility Occupation Time via Regularized Laplace Inversion," *Working Paper*.

Rosinski, J. (2007), "Tempering stable processes," *Stochastic Processes and their Applications*, 117, 677–707.

Sato, K.-I. (2013), *Levy Processes and Infinitely Divisible Distributions*.

Stinchcombe, M. B. and White, H. (1998), "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295–325.

Todorov, V. and Tauchen, G. (2012), "The Realized Laplace Transform of Volatility," *Econometrica*, 80, 1105–1127.

— (2014), "Limit theorems for the empirical distribution function of scaled increments of Itô semimartingales at high frequencies," *The Annals of Applied Probability*, 24, 1850–1888.