Exact Bayesian Moment Based Inference for the Distribution of the Small-Time Movements of an Ito Semimartingale

A. Ronald Gallant * and George Tauchen †

May 3, 2017

Abstract

We modify the Gallant and Tauchen (1996) efficient method of moments (EMM) method to perform exact Bayesian inference, where exact means no reliance on asymptotic approximations. We use this modification to evaluate the empirical plausibility of recent predictions from high frequency financial theory regarding the small-time movements of an Ito semimartingale. The theory indicates that the probability distribution of the small moves should be locally stable around the origin. It makes no predictions regarding large rare jumps, which get filtered out. Our exact Bayesian procedure imposes support conditions on parameters as implied by this theory. The empirical application uses S&P Index options extending over a wide range of moneyness, including deep out of the money puts. The evidence is consistent with a locally stable distribution valid over most of the support of the observed data while mildly failing in the extreme tails, about which the theory makes no prediction. We undertake diagnostic checks on all aspects of the procedure. In particular, we evaluate the distributional assumptions regarding a semi-pivotal statistic, and we test by Monte Carlo that the posterior distribution is properly centered with short credibility intervals. Taken together, our results suggest a more important role than previously thought for pure jump-like models with diminished, if not absent, diffusive component.

Keywords: high-frequency data, activity index, efficient method of moments, semimartingale, specification test, spot variance, stochastic volatility.

JEL classification: C51, C52, G12.

*Department of Economics, Penn State University, University Park PA 16802; email: aronaldg@gmail.com

†Department of Economics, Duke University, Durham NC 27708; email: george.tauchen@duke.edu.

1 Introduction

1.1 Overview

This paper has two closely connected objectives. The first to is to gain further understanding of the distributional aspects of high frequency small-time moves of financial prices. In particular, we aim to evaluate the empirical plausibility of some recent sharp theoretical predictions regarding the probability distribution governing the small moves. As explained further below, this effort directly puts us into a situation where we have a parametric probability model for the data that can be easily simulated, but the density itself is not available in convenient closed form. Of course, this aspect of the paper puts us into the classical indirect estimation context, e.g., simulated method of moments, etc. Of the various available techniques for this context, the EMM approach as reviewed at length in (Gallant and Tauchen, 2010) has certain optimality properties. However, EMM was developed long before new computing techniques made Bayesian inference feasible for more complicated problems than the classical implausibly simplistic setups. Since Bayes is arguably preferred for parametric problems (Berger and Wolpert, 1988, p. 1) our second objective, which follows directly from the first, is to develop the exact Bayesian version of EMM. By exact we mean the term in the usual Bayesian sense of the word exact[1] — no reliance on asymptotic approximations. Our exact Bayesian EMM can be expected to have many applications that could be explored in further work beyond that of this paper .

Our specific interest pertains to the small time structure of the martingale component of a financial time series. By far the dominant presumption in financial econometrics is that the martingale component is locally Brownian motion. This assumption underlies all statistical inference inference theory generating central limit theorems for the various variational measures commonly used in practice. While Brownian motion is perhaps reasonable for a variety of financial prices, there is nothing in the standard arbitrage-free semimartingale theory that

[1] See subsection 1.2 for a precise definition.

dictates Brownian motion, i.e., local Gaussianitiy, which is a very restrictive form for the martingale component. As discussed below, under reasonable assumptions the small-time behavior of the martingale part of any financial series will be either Brownian or infinitely active pure jump. Unlike Brownian motion, pure jump processes are nowhere continuous with infinitely many small jumps on any finite interval of time. They can be thought of as ultra-intense, i.e., infinitely intense, compound Poisson processes; indeed, the analogy is excellent because the class of Lévy processes is just the closure under weak convergence (convergence in distribution) of the family of compound Poisson processes.

The behavior of the small-time moves of the martingale component is governed by the activity index (Jacod and Protter, 2012, p. 67). The index is a higher-order property of the probability distribution of the martingale increments, and here we treat it as a time invariant random variable (or parameter), as common in most of the literature.[2] The index is a nonparametric measure of the characteristics of the small moves in the process; it specifically defines the rate at which the Lévy density diverges to infinity around its pole at zero. The index ranges over $\beta \in [0, 2]$, with low values of the index associated with quiescent low-activity processes and large values of the index associated with highly active processes that in the limit are Brownian motion at $\beta = 2$. Evidently, the index separates the martingale part of financial price processes into equivalence classes labeled by the value of β.

Options are natural candidates for potential pure jump processes. As noted by Andersen et al. (2015) short-maturity deep out-of-the-money put options load mostly on the negative jump intensity and have hardly any exposure to the diffusive spot volatility, and analogous results hold for short-maturity deep out-of-the-money call options. At the same time, short-maturity at-the-money options are largely determined by the current spot value of the stochastic volatility, which has a diffusive Brownian component in many stochastic volatility

[2]Todorov (2016) develops a test for a time varying activity index derived under the null of time-constant index but only implements the method on Monte Carlo data.

models. If an asset's price loads on multiple factors, then its activity index is determined the largest index of the multiple factors, which would be 2 in the presence of a significant Brownian component. In applications, one works with the more convenient representation of an option price in terms of its Black-Scholes implied volatility, which is a smooth nonlinear function of the option price, and the index is preserved under smooth transformations. In other words, the implied volatility will inherit the activity index of the options price itself. Thus, we might expect the implied volatilities of short-maturity deep out-of-the-money put options possibly to be pure jump processes with index β to be determined in estimation, and leave open the possibility that the implied volatilities of close-to-the-money options are either pure jump or diffusive with locally Gaussian behavior of price moves.

We apply the exact Bayesian-EMM method to options data in order to evaluate the empirical plausibility of a recent result (described below) in high frequency theory that under very mild conditions on the financial price process, the probability distribution of the small moves are either Gaussian or locally stable around the origin with index β. The result is representational, in the sense that it means we can always think of the price process as one that is stable plus a second piece that compensates for the large moves. The operative term here is "locally stable." Their result does not state that the probability distribution is globally the classical α-stable distribution used in many areas of science. Perhaps a good way to understand our objective is to contrast it with Garcia et al. (2011), who use a score-based indirect inference procedure to estimate the α parameter of the α-stable distribution. That work requires use a heavy-tailed t-like distribution for an auxiliary model because of the large rare moves that the α-stable generates. We do just the opposite. The data series are truncated to remove the big jumps, which are totally irrelevant in the present context, and we use a type of thin-tailed auxiliary model generated by the SNP sieve estimation technique. We also de-volatilize the data to generate an observed sample reasonably considered iid.

The EMM technique is appealing for model evaluation, because it forces the model under

consideration to confront the entire distribution of the observed data. In our particular case, the model predicts only locally stable data, so we thereby expect model validity to imply that it works well over most of the support of the density of the data but possibly breaks down in the extreme tails. If the model fails to fit the central part of the distribution of the data, then that would be construed as discrediting evidence. Of course the estimate of the index parameter β is quite important as well, because it informs us about the particular equivalence class the price series belongs.

1.2 Exact Bayesian Inference

The central idea is that moment equations can be used to construct a "method of moments representation" of the likelihood of a structural model that can be used in otherwise standard Bayesian inference.

In Bayesian inference the parameter ρ of a structural model is formally manipulated as if it were random even though one might regard it as fixed. Thus, the structural model implies a joint distribution $P^o(y, \rho)$ defined over $\mathcal{Y} \times \mathcal{R}$, whose density is the product of a likelihood $p(y \,|\, \rho)$ times a prior $\pi(\rho)$. The joint probability space under consideration is, therefore, $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}^o, P^o)$, where $\mathcal{C}^o$ denotes the Borel subsets of $\mathcal{Y} \times \mathcal{R}$.

One sets forth a set of moment equations $m(y, \rho)$ for use in inference. Supposing, as is the case for this paper, that the data can be summarized as a vector $\tilde{\theta}(y)$, then these moment equations can be written $m(\tilde{\theta}, \rho)$.[3] However, in this subsection, we shall continue to use $m(y, \rho)$. Normalize the moment equations as $Z(y, \rho) = V^{-\frac{1}{2}} m(y, \rho)$, where $V^{-\frac{1}{2}}$ denotes a factorization of the inverse of an estimate $V(y, \rho)$ of the variance-covariance matrix of $m(y, \rho)$.

The random variable $z = Z(y, \rho)$ over $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}^o, P^o)$ has some distribution $\Psi(z)$ with

[3]The EMM literature uses the ordering $(\rho, \tilde{\theta}(y))$ whereas the "Reflections" literature (Gallant, 2016b) uses the ordering (y, ρ). In this subsection we shall use the "Reflections" convention. Later, in Section 4, we revert to the EMM convention and use the ordering (ρ, θ).

a support $\mathcal{Z}$. Let $\mathcal{C}$ be the smallest σ-algebra containing the preimages $C = Z^{-1}(B)$ where B ranges over the Borel subsets of $\mathcal{Z}$. Because the distribution $\Psi(z)$ of $z = Z(y, \rho)$ is determined by P^o the probability measure $P[C = Z^{-1}(B)] = \int_B d\Psi(z)$ over $(\mathcal{Y} \times \mathcal{R}, \mathcal{C})$ will satisfy $P(C) = P^o(C)$ for every $C \in \mathcal{C}$.

Define $\mathcal{C}^*$ to be the smallest σ-algebra that contains all sets in $\mathcal{C}$ plus all sets of the form $R_B = (\mathcal{Y} \times B)$, where B is a Borel subset of $\mathcal{R}$. (Gallant, 2016b, Section 3) proves that there is an extension of $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}, P)$ to a space $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}^*, P^*)$ such that $P^o(C) = P^*(C)$ for all $C \in \mathcal{C}^*$. In particular, $P^o(C) = P(C) = P^*(C)$ for $C \in \mathcal{C}$ and $P^o(R_B) = P^*(R_B)$. The σ-algebras involved satisfy $\mathcal{C} \subset \mathcal{C}^* \subset \mathcal{C}^o$.

If $Z(y, \rho)$ is a semi-pivotal, a sufficient condition for which is that Z is continuous and unbounded in at least one element of y, and has distribution Ψ with density ψ, then y has conditional density $\text{adj}(y, \rho)\psi[Z(y, \rho)]$ defined over $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}^*, P^*)$. The term $\text{adj}(y, \rho)$ is analogous to a Jacobian; it is defined in Gallant (2016c) and set forth explicitly for our application as (22) in Section 4. The density $\text{adj}(y, \rho)\psi[Z(y, \rho)]$ is the 'method of moments representation" of the likelihood.

The key insight that allows us to substitute the "method of moments representation" of the likelihood for $p(y\,|\rho)$ in a Bayesian analysis is the fact that both probability measures P^o and P^* assign the same probability to sets in $\mathcal{C}^*$. Naturally, because $\mathcal{C}^*$ is a subset of $\mathcal{C}^o$, some information is lost. Intuitively this is similar to the information loss that occurs when one divides the range of a continuous variable into intervals and uses a discrete distribution to assign probability to each interval. Both the continuous and discrete distributions assign the same probability to each interval but the discrete distribution cannot assign probability to subintervals. How much information is lost depends on how well one chooses moment conditions.

1.3 Relation to Approximate Bayesian Computation

In this section we relate our proposed estimator to the approximate Bayesian computation (ABC) estimators as summarized and unified in Ng and Forneron (2016). Both the estimator that we propose and what is described in Ng and Forneron (2016) begin with a summary of the data: $\tilde{\theta} = \tilde{\theta}(y)$. With our proposal the summary statistic is further normalized to $z = Z(y, \rho)$, defined in the preceding Subsection 1.2. The ABC statistic may or may not be further normalized; regardless, we denote it similarly as $z = Z(y, \rho)$, realizing, for ABC, that Z might be the identity function $\tilde{\theta}(y) = Z(y, \rho)$.

The crucial difference between ABC and our proposal is subtle: ABC regards z as defined over a probability space $(\mathcal{Y}, \mathcal{B}, P_\rho)$. Our proposal regards z as being defined over the joint probability space $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}^o, P^o)$ defined in the preceding Subsection 1.2.

In consequence, in order to justify an ABC method as a Bayesian method in finite samples, one must verify that the approximating conditional density of z given ρ converges uniformly over the parameter space $\mathcal{R}$ to a density $\psi(z \,|\, \rho)$ and, better, be able to determine the error of the approximation. In general, if the parameter space is not finite, this is an impossible task. We are using simulation methods here so that it is in principle possible to assess the accuracy of the approximation at a finite set of ρ but one cannot as a practical matter do so over a continuum.

With our proposal, all one needs to do to assess the accuracy of a putative density $\psi(z)$ for z is to simulate from the probability space $(\mathcal{Y} \times \mathcal{R}, \mathcal{C}^o, P^o)$. This is straightforward; no issues of uniformity arise. The exact procedure for doing so is described and implemented in Subsection 6.

There are some caveats in order here. The methodology we propose is new so what we suggest next cannot be regarded as definitive: Experience acquired to date indicates that it is important to determine the tails of $\psi(z)$. Other than that, excessive precision does not seem to be required. One's natural predilection is to impose symmetry so that $\int z\, \psi(z)\, dz = 0$ in

deference to the frequentist identification condition. However, it is likely that the condition can be substantially weakened. We guard against reservations one might have due to these caveats by simulating our methodology from beginning to end in Section 6.5. That cross check does not suggest any serious flaws.

1.4 Summary of the Computations

The steps to compute the estimator we propose are distributed over the remainder of the paper as dictated by a logical development. For the reader's convenience, we provide a summary here.

1. A probability space $(\mathcal{Y}, \mathcal{B}, P_\rho)$ determined by a semimartingale with parameter $\rho \in \mathcal{R}$ from which one can simulate is defined in Section 2.

2. The maximum likelihood estimator $\tilde{\theta}$ of an SNP density $f_K(y \,|\, \theta)$ is computed from the data $\{y_t\}_{t=1}^n$ as described in Section 4. Code implementing step 2 and a User's Guide are available at www.aronaldg.org/webfiles/snp.

3. From a long simulation of $\{\hat{y}_t\}_{t=1}^N$ distributed as $(\mathcal{Y}, \mathcal{B}, P_\rho)$ compute

$$\begin{aligned} m(\rho, \tilde{\theta}) &= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \log[f_K(\hat{y}_i \,|\, \tilde{\theta})] \\ \mathcal{I}(\rho, \tilde{\theta}) &= \frac{1}{N} \sum_{t=1}^{N} \Big[\frac{\partial}{\partial \theta} \log f_K(\hat{y}_t \,|\, \tilde{\theta})\Big] \Big[\frac{\partial}{\partial \theta} \log f_K(\hat{y}_t \,|\, \tilde{\theta})\Big]' \end{aligned}$$

   as described in Section 4.

4. Compute

$$\begin{aligned} Z(\rho, \tilde{\theta}) &= \sqrt{n} \, \Big[\mathcal{I}(\rho, \tilde{\theta})\Big]^{-1/2} m(\rho, \tilde{\theta}) \\ \text{adj}(\rho, \tilde{\theta}) &= \left| \det \left\{ \frac{\partial}{\partial \theta'} \sqrt{n} \, \Big[\mathcal{I}(\rho, \tilde{\theta})\Big]^{-1/2} m(\rho, \tilde{\theta}) \right\} \right| \end{aligned}$$

   as described in Section 4, where $[(\mathcal{I})^{-1/2}]'[(\mathcal{I})^{-1/2}] = (\mathcal{I})^{-1}$.

5. For prior $\pi(\rho)$ and proposal density $q(\rho^o, \rho^*)$, where ρ^o is contemporaneous and ρ^* future, generate an MCMC chain $\{\rho_t\}_{t=1}^R$ that is extended from $\rho^o = \rho_t$ to ρ_{t+1} by proposing ρ^*, repeating steps 3 and 4 with $\rho = \rho^*$, and setting ρ_{t+1} to ρ^* with probability $\alpha = \min\left(1, \frac{\text{adj}(\rho^*,\tilde{\theta})\,\psi[Z(\rho^*,\tilde{\theta})]\,\pi(\rho^*)\,q(\rho^*,\rho^o)}{\text{adj}(\rho^o,\tilde{\theta})\,\psi[Z(\rho^o,\tilde{\theta})]\,\pi(\rho^o)\,q(\rho^o,\rho^*)}\right)$ else to ρ^o. A more detailed description of the chain is at the end of Section 4. Code implementing steps 3 though 5 and a User's Guide are available at www.aronaldg.org/webfiles/emm.

The MCMC chain is used for Bayesian inference in the usual way: The estimate of ρ is, e.g., the mean of the chain. Credibility intervals are formed from the 2.5% and 97.5% quantiles of the chain.

2 The Setting

2.1 Locally Gaussian Itô semimartingale

The basic model underlying most of financial econometrics is the Brownian Itô semimartingale model given by the following differential equation

$$dX_t = \sigma_t dW_t + dY_t, \tag{1}$$

where σ_t is a process with càdlàg paths, W_t is a Brownian motion, and Y_t is an Itô semimartingale process of pure-jump type (i.e., semimartingale with zero second characteristic, Definition II.2.6 in Jacod and Shiryaev (2003)). The above model is the familiar jump diffusion model with rare jumps embodied in Y_t. One can think of the model as comprised of an ultra-active piece represented by the locally Gaussian part in the Brownian motion and a second very inactive part reflecting the occasional large jump-like moves that are known to be present in financial prices. At high-frequencies, the dominant component of X_t in (1) is the continuous martingale component and at these frequencies the increments of X_t in (1) behave like scaled and independent Gaussian random variables.

The local Gaussianity of the model (1) is crucial for essentially all nonparametric estimation and inference with infill asymptotics in high frequency data analysis. Examples include the jump-robust Bipower Variation of Barndorff-Nielsen and Shephard (2004, 2006) and the many other alternative measures of powers of volatility summarized in the recent book of Jacod and Protter (2012). Another important example is the general approach of Mykland and Zhang (2009) where estimators of functions of volatility are formed by and working as if price moves are Gaussian with constant local volatility over a block of decreasing length. Study of the proof strategy used in all this literature reveals that it proceeds by an initial step for localization and elimination (theoretically) of the jumps (see, e.g., (Jacod and Protter, 2012, 5.3.2 pp. 149–151)) and then zooming down to the local Brownian part and obtaining the sought after central limit theorem. While easy enough to describe, this proof strategy is technically very demanding as revealed by review of the arguments of Jacod and Protter (2012, Sec. 5.3.3, pp. 151–160; Sec. 16.5, pp. 521–551).

2.2 Pure Jump Itô semimartingale

The setup in (1) is a special case of a more general setup in which X_t is a pure-jump type process. Specifically the process X_t follows dynamics

$$dX_t = \sigma_t dS_t + dY_t, \tag{2}$$

where σ_t and Y_t are càdlàg with Y_t of pure-jump type, and S_t is a symmetric stable process with a characteristic function given by

$$\mathbb{E}(e^{iuS_t}) = e^{-|tcu|^\beta}, \tag{3}$$

when $\beta \in (0, 2]$ and $c > 0$ is a scale parameter.

It is important to keep in mind the role of Y_t in (2) above. When $\beta = 2$ in (3), then X_t is the original jump-diffusion specification in (1) above. In that case it is natural to think of Y_t as a low activity large jump process independent of the Brownian motion. On the

other hand, when $\beta < 2$, X_t is of pure-jump type. In this case, Y_t in (2) is by no means independent of S_t and it plays the role of a compensator jump component that cancels out the large moves in the stable process S_t at high frequencies. Y_t can also have dependence not only with S_t but with σ_t as well, and thus X_t does not inherit the tail properties of the stable process S_t. For example X_t can be a tempered stable process, with moments of all orders, and whose tail behavior is very different from that of the stable process. This setup works because the large moves in the stable process are a compound Poisson process which can be canceled in part by moves in the residual jump process Y_t.

While the tail behavior of X_t is not the same as the stable S_t, its local behavior agrees with that of S_t. Under reasonable assumptions, the local Gaussianity of (1) extends as follows:

$$h^{-1/\beta}(X_{t+sh} - X_t) \xrightarrow{\mathcal{L}} \sigma_t \times (S'_{t+s} - S'_t), \quad \text{as } h \to 0 \text{ and } s \in [0,1], \tag{4}$$

for every t and where S'_t is stable of index β as in (3). Put simply, the local behavior of the increments of the process is like that of a stable process in the more general setting of (2).

3 Inference

3.1 Heuristics

The next subsection draws heavily on Todorov and Tauchen (2014), and it is technically demanding, so we first give short overview to keep things in perspective. The main result of is that when financial returns are suitably "de-volatized", i.e., rescaled to adjust for stochastic volatility, and the large jumps removed, then the local behavior of the return is necessarily either Gaussian or β-stable. The result is akin to the classical general central limit theorem with its domains of attraction. The aforementioned paper only develops this distributional prediction but never formally tests it. Instead the paper generates a Kolmogorov-Smirnov type test for local Gaussianity against the alternative of locally stable, and the test statistics

exploits in essential ways the presumed Gaussianity under the null hypothesis.

Our goal in the paper is to confront this distributional prediction with the data. This task, of course, entails at the same time estimation of the stability parameter β as well. Since the probability density of the stable distribution is not available in convenient closed form but the stable is easy to simulate, our task fits very neatly into the EMM framework. Indeed, in the empirical section we use an SNP model to represent the reduced form probability distribution of the data and the local stable with index β as the maintained structural model (the data generation process).

It is important to always keep in mind that the distributional prediction to be tested does not imply that the probability distribution of price moves is in the stable class, but rather it says that the fine scale local moves are β–stable-like.[4] Thus in implementation we necessarily need to truncate the huge moves generated in a simulation from the stable distribution. These moves play no role in the theory and are just an artifact of the heavy-tailed nature of the Lévy density of the stable distribution.

Our strategy is to first to use high frequency statistics and econometrics for a fixed time span to indicate the theoretically predicted distribution, i.e., the locally stable. We then proceed in the subsequent section to long span estimation which presumes the data generation process remains constant over time. The long span analysis is needed to overcome the very inherent limitations of a high frequency data set over a fixed interval over time and thereby achieve a reasonable degree of precision.

3.2 Fixed Span

For this subsection, we are interested in the process X_t over an interval of fixed length, say $[0, 1]$, observed over the equidistant grid $0, \frac{1}{n}, ..., 1$ with $n \to \infty$. Under equation (1) the high-frequency increments $\Delta_i^n X = X_{\frac{i}{n}} - X_{\frac{i-1}{n}}$ are approximately Gaussian. In equation (2)

[4]One may view the result as a prior that concentrates on the stable distributions over the space of infinitely divisible distributions.

suggests that the high-frequency increments $\Delta_i^n X = X_{\frac{i}{n}} - X_{\frac{i-1}{n}}$ are approximately β–stable. The idea is to proceed as one might in the locally Gaussian case (1) by using a bipower type variation estimator to estimate the local scale. We estimate σ_t locally and then divide the high-frequency increments by this estimate. It turns out that the procedure is valid in the locally β–stable situation of (2) as well.

One divides the interval $[0,1]$ into blocks, each of which contains k_n increments, for some deterministic sequence $k_n \to \infty$ with $k_n/n \to 0$. On each of the blocks the local estimator of σ_t^2 is given by

$$\widehat{V}_j^n = \frac{\pi}{2}\frac{n}{k_n - 1} \sum_{i=(j-1)k_n+2}^{jk_n} |\Delta_{i-1}^n X||\Delta_i^n X|, \quad j = 1, ..., \lfloor n/k_n \rfloor. \tag{5}$$

$\widehat{V}_j^n$ is the bipower variation estimator proposed by Barndorff-Nielsen and Shephard (2004, 2006) for measuring the quadratic variation of the diffusion component of X_t. While an alternative measure of σ_t is available using a truncated variation, it turns out, however, that the behavior of the two volatility measures differs in the case when S_t is stable with $\beta < 2$. Using the truncated variation in our situation leads to degenerate limits of key components, unlike the case of using the bipower variation estimator in (5) since it leads to the self-scaling discussed below.

The scaling of every high-frequency increment is done after adjusting $\widehat{V}_j^n$ to exclude the contribution of that increment in its formation

$$\widehat{V}_j^n(i) = \begin{cases} \frac{k_n-1}{k_n-3}\widehat{V}_j^n - \frac{\pi}{2}\frac{n}{k_n-3}|\Delta_i^n X||\Delta_{i+1}^n X|, \text{ for } i = (j-1)k_n + 1, \\ \frac{k_n-1}{k_n-3}\widehat{V}_j^n - \frac{\pi}{2}\frac{n}{k_n-3}\left(|\Delta_{i-1}^n X||\Delta_i^n X| + |\Delta_i^n X||\Delta_{i+1}^n X|\right), \\ \qquad \text{for } i = (j-1)k_n + 2, ..., jk_n - 1, \\ \frac{k_n-1}{k_n-3}\widehat{V}_j^n - \frac{\pi}{2}\frac{n}{k_n-3}|\Delta_{i-1}^n X||\Delta_i^n X|, \text{ for } i = jk_n. \end{cases} \tag{6}$$

Now consider

$$\frac{\Delta_i^n X}{\sqrt{\widehat{V}_j^n(i)/n}}, \tag{7}$$

which is the de-volatilized increment. For technical reasons, one uses the first m_n increments in each block, with $m_n \leq k_n$. The case $m_n = k_n$ amounts to using all increments in the block but one needs $m_n < k_n$. Within the first m_n increments on each block one keeps only the high-frequency increments that do not contain big jumps, i.e., a retained increment is of the form

$$\frac{\Delta_i^n X}{\sqrt{\widehat{V}_j^n(i)/n}} \mathbf{1}\left(|\Delta_i^n X| \leq \alpha \sqrt{\widehat{V}_j^n} n^{-\varpi}\right), \tag{8}$$

where $\alpha > 0$ an $\varpi \in (0, 1/2)$; there there is a time-varying jump threshold threshold in our truncation to account for the time-varying σ_t.

The above procedure takes the original n increments and generates a subset of adjusted increments that are asymptotically independently and identically distributed increments, with a distribution that depends upon whether the data generation process is the jump diffusion model (1) or the pure jump model (2). The empirical distribution of these increments is

$$\widehat{F}_n(\tau) = \frac{1}{N^n(\alpha, \varpi)} \sum_{j=1}^{\lfloor n/k_n \rfloor} \sum_{i=(j-1)k_n+1}^{(j-1)k_n+m_n} 1\left\{ \frac{\Delta_i^n X}{\sqrt{\widehat{V}_j^n(i)/n}} \leq \tau \right\} 1_{\left\{|\Delta_i^n X| \leq \alpha \sqrt{\widehat{V}_j^n} n^{-\varpi}\right\}}, \tag{9}$$

where the divisor is the total number of retained increments

$$N^n(\alpha, \varpi) = \sum_{j=1}^{\lfloor n/k_n \rfloor} \sum_{i=(j-1)k_n+1}^{(j-1)k_n+m_n} 1\left(|\Delta_i^n X| \leq \alpha \sqrt{\widehat{V}_j^n} n^{-\varpi}\right), \tag{10}$$

$\widehat{F}_n(\tau)$ is simply the empirical cdf of the de-volatilized increments that do not contain large jumps. In the jump-diffusion case of (1), $\widehat{F}_n(\tau)$ should be approximately the cdf of a standard normal random variable.

The de-volatatization procedure resembles the practice of standardizing increments of the process of fixed length by a measure for volatility constructed from high-frequency data within the interval (after correcting for jumps and leverage effect), see e.g. Andersen et al. (2001). The main difference is that here the length of the increments that are standardized is shrinking and further the volatility estimator is local, i.e., over a shrinking time interval.

Both these differences are crucial for deriving the limit theory of $\widehat{F}_n(\tau)$.

The limit result for $\widehat{F}_n(\tau)$ obtained in Todorov and Tauchen (2014) subsumes both data generation processes (1) and (2), but for technical reasons (see below) it is presumes that $\beta \in (1, 2]$. Their result is that when the tuning parameters k_n and m_n grow at appropriate rates

$$\widehat{F}_n(\tau) \xrightarrow{\mathbb{P}} F_\beta(\tau), \quad \text{as } n \to \infty, \tag{11}$$

where the above convergence is uniform in τ over compact subsets of $\mathbb{R}$; $F_\beta(\tau)$ is the cdf of $\sqrt{\frac{2}{\pi}}\frac{S_\beta}{\mathbb{E}|S_\beta|}$ (S_β is a realization of a β-stable random variable (with $t = 1$ in (3)), and $F_2(\tau)$ equals the cdf of a standard normal variable $\Phi(\tau)$. The limit result in (11) shows that when S_t is stable with $1 < \beta < 2$, $\widehat{F}_n(\tau)$ estimates the cdf of a β-stable random variable around the origin, while when $\beta = 2$ it estimates the cdf of the Gaussian.

The technical restriction $\beta > 1$ pertains to a possible drift, but it is innocuous in the present context. Stable processes with $\beta < 1$ are of finite variation with trajectories that appear nearly flat over long stretches of time interspersed with occasional huge moves that are so large that the first moment does not exist. Such processes would be inadequate representations of the high-activity component of conventional financial time series. Stable processes with $1 < \beta < 2$ are of infinite variation with finite first moment, just like Brownian motion. Note role that the implied first moment plays in in the statement of the limiting result (11).

As a final note on the local nature of the distributional prediction to be tested, the convergence in (11) can trivially be extended from compact sets around the origin to uniform convergence over the entire line using some standard tools of probability theory. But the uniform convergence is of theoretical interest at most and is not operational, in the sense being useful for empirical practice. For the distribution to hold over arbitrarily large intervals would require arbitrarily large sample sizes well beyond those that would ever be available.

3.3 Empirical Implications

To summarize, the preceding indicates that on any particular fixed interval of time the ensemble of de-volatized returns (8) can be expected to be distributed either as locally Gaussian under the jump diffusion model (1), or locally β-stable under model (2) when $1 < \beta < 2$. To get a visual handle on the implied probability distributions, Figure 1 shows the probability distribution of the β–stable for $\beta = 1.10$ along with a reference Gaussian scaled so that both distributions have an absolute first moment unity $\mathcal{E}(|Z|) = 1$; the distribution is more like the Cauchy, which is the case $\beta = 1$. Note that $\beta > 1$ so the first moment exists, and with this scaling the variance of the reference Gaussian is $\pi/2$. The shape is very non-Gaussian for $\beta = 1.10$ but obviously it would be more Gaussian-like as β gets closer to 2.

4 Bayesian Inference with EMM

We are in a situation that can be characterized as follows. A Levy process determines the distribution of a certain transform of high frequency data. This distribution, called the structural model henceforth, has parameter ρ. In general ρ is a vector but here it has only one element: $\rho = (\rho_1) = (\beta)$. Theory provides strong prior information regarding ρ. Specifically, the support of the prior must include $1 < \rho_1 \leq 2$. The prior we shall use distributes its mass uniformly over $(1, 2]$. Denote this prior by $\pi(\rho)$. We can accurately simulate from the structural model. The likelihood is not known in closed form and standard approximations such as inversion of the characteristic function are not numerically stable. This situation indicates the desirability of a Bayesian method of moments approach. Our preference is for exact Bayesian inference rather than approximate methods.

Let $\mathcal{E}_\rho$ denote expectation with respect to the structural model. Because we can accurately simulate from the structural model, we can compute an expectation to any desired accuracy using $\mathcal{E}_\rho(g) \doteq \frac{1}{N}\sum_{i=1}^{N} g(\hat{y}_i)$, where $\{\hat{y}_i\}_{t=1}^{N}$ a simulation of length N generated from

the structural model with the parameter set to ρ. Thus, the ability to accurately simulate provides a means to implement a Bayesian method of moments approach.

We construct the moments to be used in the analysis as follows. Let $f(y \mid \theta)$ be a density that well approximates the density of the observed data $\{y_i\}_{i=1}^n$. Let

$$\tilde{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \log f(y_i \mid \theta), \tag{12}$$

where Θ is the parameter space of the density $f(y \mid \theta)$. We regard $\tilde{\theta}$ as a summary of the data; the data enters our analysis only through this summary. Define

$$m(\rho, \theta) = \mathcal{E}_\rho \left\{ \frac{\partial}{\partial \theta} \log f(y \mid \theta) \right\} \tag{13}$$

Our moment equations are $m(\rho, \tilde{\theta})$. These moment equations satisfy $m(\rho, \tilde{\theta}) = O_p(n^{-\frac{1}{2}})$ when ρ is set to the value of the structural model that generated the data regardless of how well $f(y \mid \theta)$ approximates the data (Gallant and Tauchen, 1996).

One would expect that better approximations will achieve better results in applications. In the event that regularity conditions, e.g., Chernozhukov and Hong (2003), are in place such that the posterior from the methods proposed below has an asymptotic distribution, the results of Gallant and Tauchen (1996) imply that better approximations yield shorter credibility intervals.

We shall use the SNP density for $f(y \mid \theta)$ and shall denote it as $f_K(y \mid \theta)$. It has the form of a location-scale transform of an innovation v

$$y = Rv + \mu, \tag{14}$$

where R is an upper triangular matrix; v and μ are vectors in general but in our application v, μ, and R are scalars. Nonetheless, for generality, we will retain the multivariate representation whereby v and μ are vectors of length M and R is an $M \times M$ matrix. The density function of the innovation is

$$h_K(v) = \frac{[\mathcal{P}_K(v)]^2 \phi(v)}{\int [\mathcal{P}_K(u)]^2 \phi(u)\, du}, \tag{15}$$

where $\mathcal{P}_K(v)$ is a polynomial in v of degree K and $\phi(v)$ is the standard normal $\phi(v) = (2\pi)^{-\frac{1}{2}M} e^{-\frac{1}{2}v'v}$. A standard representation of a multivariate polynomial $\mathcal{P}_K(v)$ is

$$\mathcal{P}_K(v) = \sum_{|\lambda|=0}^{K} a_\lambda v^\lambda, \tag{16}$$

where λ is a multi-index and $|\cdot|$ denotes the degree of a multi-index. A multi-index is a vector λ of dimension M with non-negative integer elements; $v^\lambda = \prod_{i=1}^{M} v_i^{\lambda_i}$; and $|\lambda| = \sum_{i=1}^{M} \lambda_i$. The parameters θ of $f_K(y \,|\, \theta)$ are, from (14), the elements of μ and the elements of the upper triangle of R, and, from (16), the scalars $\{a_\lambda\}_{\lambda \le K}$. Because $[\mathcal{P}_K(v)]^2 / \int [\mathcal{P}_K(u)]^2 \phi(u) du$ is a homogeneous function of the $\{a_\lambda\}_{\lambda \le K}$, $h_K(v)$ can only be determined to within a scalar multiple. To achieve a unique representation, the constant term a_0 of the polynomial $\mathcal{P}_K(v)$ is put to one. With this normalization, $h_K(v)$ has the interpretation of a series expansion whose leading term is the normal density $\phi(v)$ and whose higher order terms induce departures from normality. To avoid division by zero when evaluating (15) or its score vector one can add a small $\epsilon > 0$ to the denominator of (15).

The SNP model is a sieve, which is to say that it is a parametric model with a variable number of parameters that is dense for some class of functions in some statistically relevant norm. See Gallant and Nychka (1987) for regularity conditions, the most stringent of which are that that the class of density functions that can be represented by an SNP density with $K = \infty$ must possess a moment generating function and must be dominated by a Sobelev norm. Usually this domination condition has the effect of limiting the parameters of structural models whose density can be represented to a compact set. In applications, the SNP density behaves much the same as a kernel density estimator (Fenton and Gallant, 1996) and using the Bayes information criterion (BIC) to choose K is a data driven rule that guarantees good performance (Coppejans and Gallant, 2002).

The first step of our proposed Bayesian estimator is to use

$$f_K(y \,|\, \theta), \quad \theta \in \Theta, \tag{17}$$

to summarize the data $\{y_i\}_{i=1}^n$ by computing

$$\tilde{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{t=1}^{n} \log[f_K(y_t \,|\, \theta)]. \tag{18}$$

The data $\{y_i\}_{i=1}^n$ are the retained increments given by (8). The value of K is chosen using the BIC criterion. The software we use to implement the first step is available together with a User's Guide at www.aronaldg.org/webfiles/snp.

We compute $m(\rho, \theta)$ given by (13) by averaging over a long simulation

$$m(\rho, \theta) \doteq \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \log[f_K(\hat{y}_i \,|\, \theta)]. \tag{19}$$

Our approach is to use simulations that are long enough that the error in the approximation of (13) by (19) is negligible.

Define

$$\mathcal{I}(\rho, \theta) = \frac{1}{N} \sum_{t=1}^{N} \Big[\frac{\partial}{\partial \theta} \log f(\hat{y}_t \,|\, \theta)\Big] \Big[\frac{\partial}{\partial \theta} \log f(\hat{y}_t \,|\, \theta)\Big]' \tag{20}$$

and consider

$$Z(\rho, \tilde{\theta}) = \sqrt{n} \, \Big[\mathcal{I}(\rho, \tilde{\theta})\Big]^{-1/2} m(\rho, \tilde{\theta}), \tag{21}$$

where $[(\mathcal{I})^{-1/2}]'[(\mathcal{I})^{-1/2}] = (\mathcal{I})^{-1}$. Now $Z(\rho, \tilde{\theta})$ follows some distribution $\Psi(z)$ determined by the structural model and the prior; denote the corresponding density by $\psi(z)$. Under regularity conditions, $\text{adj}(\rho, \tilde{\theta})\psi[Z(\rho, \tilde{\theta})]$ can be used as a likelihood for Bayesian inference Gallant (2016b,c). The adjustment, defined in Gallant (2016c, 2015), in this instance is the Jacobian term

$$\text{adj}(\rho, \tilde{\theta}) = \left| \det \left\{ \frac{\partial}{\partial \theta'} \sqrt{n} \, \Big[\mathcal{I}(\rho, \tilde{\theta})\Big]^{-1/2} m(\rho, \tilde{\theta}) \right\} \right| \tag{22}$$

In Section 6 we verify by simulation that in our application the normal density function $\phi(z)$ can be used for $\psi(z)$.

Note, in passing, that the reason $\sqrt{n}$ appears in the above expressions is not due to asymptotic considerations but due to a presumption that $\phi(z)$ expects z to have the identity as its variance-covariance matrix.

The regularity conditions are mild, the most important of which is that $Z(\rho, \tilde{\theta})$ must satisfy a semi pivotal condition, a sufficient condition for which is that $Z(\rho, \tilde{\theta})$ be continuous in at least one element of $\tilde{\theta}$ and be unbounded from above and below in that element. That $Z(\rho, \tilde{\theta})$ actually be pivotal is also sufficient, but this is harder to verify.

Using $\text{adj}(\rho, \tilde{\theta})\psi[Z(\rho, \tilde{\theta})]$ as a likelihood together with a proper prior is exact Bayesian inference (Gallant, 2016b,c). It is not approximate Bayesian inference. The analogy with frequentist inference is nearly exact. In frequentist inference method of moments and maximum likelihood confidence interval procedures using exact finite sample distributions can lead to confidence intervals with different expected lengths but both are, nonetheless, valid 95% frequentist confidence intervals. The same is true here, using $\text{adj}(\rho, \tilde{\theta})\psi[Z(\rho, \tilde{\theta})]$ instead of the likelihood of the structural model can lead to 95% credibility intervals of different lengths. Nonetheless, both are valid Bayesian credibility intervals. The difference in lengths arises because Bayesian method of moments uses less information that classical Bayes; on this see Gallant (2016b,c). When the regularity conditions of Chernozhukov and Hong (2003) are in place, the results of Gallant and Long (1997) imply that lengths become the same as $K \to \infty$.

Note, in passing, that for θ of dimension M, $\log[\text{adj}(\rho, \tilde{\theta})]$ increases with n at the rate $\frac{M}{2}\log(n)$ whereas $\log\{\phi[Z(\rho, \tilde{\theta})]\}$ increases at the rate n. Therefore, when $\psi = \phi$ the adjustment becomes irrelevant for large n when ρ and θ are restricted to compact sets, which is a typical regularity condition in an asymptotic analysis. Nonetheless, in the work reported here, the adjustment term is included in keeping with our desire for exact rather than approximate Bayesian inference.

The second step of our proposed Bayesian estimator is to run an MCMC chain. The chain is a sample from the posterior distribution that can be used for Bayesian inference. Specifically, the mean $\bar{\rho}$ over the chain or that value $\hat{\rho}$ in the chain that maximizes $\text{adj}(\rho, \tilde{\theta})\phi[Z(\rho, \tilde{\theta})]\,\pi(\rho)$ can be used as the estimate of ρ. The sample standard deviations over

the chain can be used as an estimate of the standard deviations of the posterior distribution of ρ. On this see Gamerman and Lopes (2006). Credibility intervals can be formed from lower and upper quantiles of the chain; e.g., the 2.5% and 97.5% to get a 95% credibility interval.

To compute an MCMC chain one needs a proposal density. The proposal density $q(\rho^{old}, \rho^{new})$ that we use is move-one-at-a-time-random-walk. Specifically, we randomly choose an index $\hat{i}$ of the parameter vector $\rho = (\rho_1, \rho_2, ..., \rho_p)$ and sample u from $\phi(u)$. (In our application $\rho = (\beta)$ with $p = 1$.) Then ρ^{new} is equal to ρ^{old} excepting that $\rho_{\hat{i}}^{new} = \rho_{\hat{i}}^{old} + \sigma_{\hat{i}} u$. The scale parameters $\sigma_i, i = 1, \ldots, p$ are chosen such that the proportion of times ρ' is set to ρ^o instead of ρ^* in the algorithm described next paragraph is between 0.4 to 0.8 for each $i, i = 1, \ldots, p$. These proportions are referred to as the rejection rates.

The MCMC chain is computed as follows: Given a current ρ^o obtain the next ρ' by:

1. Draw ρ^* according to $q(\rho^o, \rho^*)$.
2. Simulate $\{\hat{y}_t\}_{t=1}^N$ from the structural model with parameter set to ρ^*. If the simulation fails, set $\rho' = \rho^o$ and return to step 1 with the new ρ^o set to ρ'.[5]
3. Compute $m(\rho^*, \tilde{\theta})$ using (19), $\mathcal{I}(\rho^*, \tilde{\theta})$ using (20), $Z(\rho^*, \tilde{\theta})$ using (21), and $\text{adj}(\rho^*, \tilde{\theta})$ using (22).
4. Compute $\alpha = \min\left(1, \frac{\text{adj}_{(\rho^*, \tilde{\theta})}\, \psi[Z(\rho^*, \tilde{\theta})]\, \pi(\rho^*)\, q(\rho^*, \rho^o)}{\text{adj}_{(\rho^o, \tilde{\theta})}\, \psi[Z(\rho^o, \tilde{\theta})]\, \pi(\rho^o)\, q(\rho^o, \rho^*)}\right)$.
5. With probability α, set $\rho' = \rho^*$, otherwise set $\rho' = \rho^o$. Return to step 1 with the new ρ^o set to ρ'.

The software we use to implement the second step is available together with a User's Guide at www.aronaldg.org/webfiles/emm.

[5]The implication of returning to step 1 when simulation fails at step 2 is that simulation success is part of the support of the prior. The possibility of simulation failure does not arise in our application for $\rho \in (1, 2]$.

5 Data

We use the E-mini futures price series and seven option-implied volatilities developed by Andersen et al. (2015) who in turn employ the core data set of Andersen et al. (2013). Options-implied volatilities, instead of raw options prices, are used so as to create standardized series with a fixed date to maturity for given fixed levels of moneyness. Since the Black-Scholes implied volatility function is sufficiently smooth, then, as well known, the small-scale fine structure of the options prices are preserved in the implied volatilities. The issues, and in particular the invariance of the index β, are discussed at further length in Andersen et al. (2015).

The E-mini S&P 500 futures contract (commodity ticker symbol ES) and its options are traded exclusively on the CME GLOBEX electronic platform. The markets for the ES futures and options are among the deepest and most liquid worldwide. The sample period is January 3, 2007, to March 22, 2011, covering 1062 trading days, and the underlying data are at the 15-second frequency. For each 15-second interval there are seven implied volatilities, IV1-IV7 for the fixed maturity of 30 calendar days.[6] The implied volatilities correspond to values of the the moneyness measure m, defined as follows: Letting σ_{BS} denote the ATM implied Black-Scholes volatility and F the forward price corresponding to tenor (maturity) τ, then

$$m = \frac{\ln(K/F)}{\sigma_{BS}\sqrt{\tau}}\,.$$

The implied (Black-Scholes) volatility measures are labeled IV1-IV7 and match to the following levels of moneyness, $m = -4, -3, -2, -1$, 0, 0.5, and 1, which covers deep out of the money ($m = -4$ to just in the money options at $m = 1$. To alleviate the effects of trading friction noise (some staleness as well as price uncertainty associated with the bid-ask spread), Andersen et al. (2015) aggregate the 15-second observations by pre-averaging over

[6]The seven implied volatility measures are constructed in the manner described in the Appendix of Andersen et al. (2015).

non-overlapping five-minute blocks using the triangular weighting scheme of Podolskij and Vetter (2009).

To summarize, our data set here consists of seven option-implied volatilities, IV1–IV7, and the log-returns on the E-mini futures, 5-minute level, January 3, 2007, to March 22, 2011, covering 1062 trading days. All data are 5-minute observations based on pre-averaged 15-second data. The data were de-volatized exactly as described in Subsection 3.2, and then at the very end rescaled to have absolute first moment unity, which puts all series on comparable scales and implies the theoretical value of c in (3) is thereby $c = \sqrt{2/\pi}$.

6 Empirical Results

Before undertaking estimation, we relate our Bayesian EMM method to other approaches for estimating the index now in the literature. The material of (Ait-Sahalia and Jacod, 2014, Chapter 12, and references therein) mainly concerns estimating the second activity index in the presence of a Brownian motion with dominating index 2, which is a statistically challenging problem with a very slow rate of convergence. Here we are interested in making inference about the dominate index not constrained to equal 2.0. There are several proposed power variation estimators for the dominate index that work only off scaling laws. Andersen et al. (2015) implement one such estimator on the same data to used here. However, the data have been truncated to exclude the uninteresting very large rare jumps. It turns out that the truncation can upward bias the power variation estimator, as is evident in Figure 4. Interestingly, the range of implied post-truncation index values, 1.65–2.00, seen in the figure is the range of values reported in Andersen et al. (2015), and these values well above our posterior medians reported below. Todorov (2015) develops an alternative estimator, which is rather complex, with a need for determination of five or more tuning parameters (the tuning parameters determine the number moments used in estimation). It also entails a sandwich for the asymptotic variance of the estimator, while in the frequentist context EMM

is asymptotically fully efficient. More important, our procedure circumvents problems with truncation by mimicking the cutoff in the simulations, and the method entails diagnostic checks on the distributional implications the model has on the data.

6.1 The Auxiliary Model

Following Section 4 above, the first step of the proposed procedure is estimate the SNP model $f_K(y\,|\,\theta)$ as described in equation (17) by quasi maximum likelihood (18). Because of serial independence in the de-volatilized data, we need not make the coefficients of the polynomial expansion in (16) state dependent (time varying). In doing the quasi maximum likelihood estimation, there was overwhelming evidence from the BIC criterion (not shown) that all of the eight series considered here are non-Gaussian. Following established protocols for the BIC, in the upward SNP expansion we started at $K_z = 4$ and only used even powers of the polynomial to enforce symmetry. In several cases the expansion stopped at a polynomial of degree 4 or 6, while in others it indicated $K_z = 8$, or possibly higher where numerical instabilities can set in. To maintain equivalence we set $K_z = 8$ across all eight series. Thus, there was mild over-expansion in some cases. But where a lower order polynomial might suffice, over-expansion just moves $f_K(y\,|,\theta)$ a bit closer to an encompassing model with little computational cost.

6.2 Structural Model Simulations

For the simulations under the model, we used the locally stable distribution exactly as predicted by the high frequency theory described in Subsections 3.1–3.2 above. For any candidate parameter $\rho = (\beta)$ we simulated long realizations of β-stable deviates with characteristic function (3) above. To maintain comparability, all observed data after the de-volitization described in Subsection 3.2 were rescaled to have sample absolute first moment of unity before entering the SNP step above, which makes the theoretical value of the scale parameter $c = \sqrt{2/\pi}$ across the eight data series.

In generating simulations under the model, it is imperative to truncate the extreme values generated by the stable distribution, and the reason for doing so is fundamental to the nature of the theory being tested. The main result being evaluated here is that the high frequency financial increments are locally stable (or Gaussian). It makes no prediction that the data should be globally stable (or Gaussian), at least for any sample size ever available in financial econometrics. Without the truncation, the theory would be pushed to a limit far beyond where it can reasonably be expected to apply. In the implementation, we generated simulated data from a truncated stable[7] with the truncation point set to 4.5. The truncation has effect — it has to — since the stable is capable of generating extremely large values unlike any close to resembling an observed financial return, but some experimentation revealed insensitivity to the chosen value. This robustness is to be expected, because the stable is known to generate huge values with very low frequency, so reasonable truncations will all essentially exclude the same set of values.

6.3 Distribution of the Semi-Pivotal

The parameter space of ρ is $(1, 2]$. The SNP density $f_K(y \,|\, \theta)$ has four even polynomial coefficients and one location parameter that are unconstrained and one positive scale parameter whence $\theta \in \mathbb{R}^5 \times \mathbb{R}^+$. Therefore, the semi-pivotal $Z(\rho, \tilde{\theta})$ given by (21) is defined over $(1, 2] \times \mathbb{R}^5 \times \mathbb{R}^+$. Let $p(y \,|\, \rho)$ denote the density of the observed data; $\pi(\rho)$ is the prior, which, recall, is uniform over $(1, 2]$. The density that determines the distribution of Z is $p(y_1, \ldots, y_n, \rho) = [\prod_{i=1}^n p(y_i \,|\, \rho)]\, \pi(\rho)$ (Gallant, 2016b).

To simulate Z one first draws $\hat{\rho}$ from the uniform over $(1, 2]$ then draws two samples $(y_1, \ldots, y_n)$ and $(\hat{y}_1, \ldots, \hat{y}_N)$ from $p(y \,|\, \hat{\rho})$ as described in Subsection 6.2. We used $n = N = 55000$. One computes $\tilde{\theta}$ from $(y_1, \ldots, y_n)$ using (12). Next one computes $\hat{z} = Z(\hat{\rho}, \tilde{\theta})$ from $(\hat{y}_1, \ldots, \hat{y}_N)$ using (21) via the intermediate computations (19) and (20). Then $\hat{z}$ is a draw

[7]For the stable itself we used the standard algorithm involving trigonometric functions; see (Gennady and Taqqu, 2000, Proposition 1.7.1, p. 41).

from ψ. We drew $R = 120000$ such $\hat{z}$.

The tails of ψ are the most important determinant of the behavior of our proposed estimator of $\rho = (\beta)$ (Gallant, 2016a). Figure 2 plots the quantiles of $\{\hat{z}_i\}_{i=1}^{R}$ (y-axis) against the quantiles of the standard normal (x-axis) at p-values from 0.001 through 0.999 in increments of 0.001. Figure 3 is the same with quantiles of a 3 d.f. Student's t-distribution instead of normal quantiles. It is clear from the plots that $\psi(z)$ is neither fat-tailed nor thin-tailed relative to the standard normal density $\phi(z)$ and that a choice of $\psi(z) = \phi(z)$ is reasonable.

6.4 Findings

Table 1 displays the Bayesian interval estimates of the index β; recall, $\rho = (\beta)$. The estimates appear to be rather sharp, given the narrowness of the individual intervals.

As discussed in the introduction, we expect far out-of-the-money options to load heavily on jump factors because of the embedded value associated with taking jump-like moves back into the money. We see just that in the very low estimates of β for the three most out-of-the-money option categories, IV1–IV3, which is consistent with being pure jump processes with relatively larger jumps. For IV4, the estimate of β is midway between unity and 2, suggesting it also is a pure jump process but with smaller jumps than IV1–IV3, presumably because of the relatively more important role that more diffusive-like factors in the determination of its value. For the near-the money options, IV5–IV7 and the S&P index future, the value of the estimated index β is much closer to the upper bound of 2. For processes like these, the jump intensity is so high and the moves so small that the sample trajectories become nearly indistinguishable from those of Brownian motion. The estimates suggest these assets presumably load on factors that are essentially diffusive, or nearly so.

It bears noting that our estimates of the index stand in contrast to those of Andersen et al. (2015), among others. We find that for the options data the estimates range from just over

1.0 to 1.88 across the moneyness m, while those reported by Andersen et al. (2015) for options data range from 1.64 to 2.00 across the moneyness. The resolution of the discrepancies lies in characteristics of the estimators. The latter set of estimates were constructions based on the behavior of the power variations computed at two different sampling frequencies without using much information regarding the distributional aspects of the data, and in particular to the truncation for large jumps. Figure 4 shows the large sample limit of the power variation estimator applied to truncated data for various candidate values of the index. As can be seen from the figure, the truncation imparts an upward bias, and the range of the asymptotic limits of the power variation estimator seen in the figure is about the same of the range of the estimates reported in the aforementioned study. By way of contrast, our EMM approach perforce must accommodate the truncation in the data simulator, and it is thereby unaffected by any biases due to truncation of large jumps. Below, we validate by Monte Carlo that the EMM is indeed unbiased. Prior to this work, most financial econometricians thought that, in view of the general vibrancy of financial prices, the activity index of any price process was nearly always 2.0 (Brownian) or no less than, say 1.60 or 1.70. Our results using a full information technique indicates that for these data the evidence suggests much lower values of the index than previously thought for the prices of far out of the money options.

Figure 5 makes clear the distributional aspects of our model assessments. The figure shows Q–Q plots of the model-implied quantiles against the quantiles of the observed data for each of the moneyness categories. The model-implied quantiles are computed from a random sample of size 1000 drawn from the model density at the estimated median value of the index β for each moneyness level; the data quantiles are computed from random sample of the observed data also of size 1000. We use the shorter random samples just to keep the number of plot points low enough so the figure can be rendered by printers now in use. The units of the axes are absolute first moments, which for Gaussian data would be $\sqrt{\pi/2} = 1.25$, so $(-3, 3)$ is akin to a range of ± 3.75 standard deviations. As seen from the figures, the

model- and data-implied quantiles are in very close agreement within the range $(-3, 3)$, just as the theory implies. Outside that range the agreement is less close, also to be expected given that the theory implies only locally, not globally, stable data.

6.5 Cross Checks

Two key aspects of our findings deserve careful scrutiny.

First, in all cases, the initial SNP estimate of the auxiliary model indicates non-Gaussian data with highly significant Hermite polynomial coefficients. Once such an auxiliary model is put in place, then the EMM procedure has to generate an estimate of the index β that is less than 2.0 because that is the only way to account for the non-Gaussian character of the auxiliary model. Selection of the auxiliary model is based on the conservative BIC criterion, but still it is important to verify that first step is not systematically finding spurious non-Gaussianity. As a check, we estimated the SNP model on $10,000$ replicates of independently and identically distributed Gaussian data sets of size $1,000$, and never once did BIC reject Gaussianity. Thus, it is very unlikely our findings are due to spurious non-Gaussianity.

Second, our estimates of the index β are significantly lower than those found elsewhere, raising the possibility that the procedure itself might be systematically downward biased. To investigate this issue, we did a small scale Monte Carlo by applying the entire procedure to simulated data sets of length equal to our sample size under the model distribution implied by Theorem 1 for two separate parameter settings, $\beta = 1.1$ and $\beta = 1.9$. The following display shows the Monte Carlo means and standard deviations over 10,000 replications:

Estimator	Mean	Std.Dev.
$\beta = 1.10$		
posterior mean	1.1003	0.0162
posterior mode	1.1006	0.0170
$\beta = 1.90$		
posterior mean	1.8698	0.0194
posterior mode	1.8679	0.0221

Evidently, the Monte Carlo values are near spot on center adding credence to our claimed empirical findings.

7 Conclusion

We undertake an extension of the exact Bayesian method of moment estimation methodology proposed by Gallant (2016b,c,a) to simulated method of moments. It can also be viewed as an extension of of the EMM method described in the Handbook of Financial Econometrics (Gallant and Tauchen, 2010). We undertake an extension for the purposes of evaluating the sharp distributional predictions arising from high frequency theory and estimating the jump activity index. Under reasonable regularity conditions, recent results in high frequency predict that at fine time scales financial prices are either locally Gaussian, as commonly presumed, or locally stable with pure jump sample paths governed by the activity index. Since short-dated options derive value from the possibility of taking jump-like moves, we use options and futures data in the empirical evaluation, specifically S&P 500 index options and futures over the period January 3, 2007, to March 22, 2011 of the Andersen et al. (2013, 2015) data set. These options extend over seven levels of moneyness, including deep out of the money puts which present some of the strongest challenges in asset pricing. Our empirical findings strongly suggest that the four most out of the money options are pure jump processes while the three other options much nearer the money and the future are

closer to locally Gaussian. We note that our estimates of the jump index are well below those previously reported in the literature, because, as we have shown, previous estimates are upward biased.

The findings have strong implications for theoretical financial econometrics. Essentially all the central limit theorems for nonparametric volatility estimation exploit in essential ways presumed local Gaussianity. To the extent this presumption fails, as the evidence suggests for some of the options data, the classical central limit theory is invalid. Development of statistical inference theory for the pure jump case is an extremely challenging task deferred to future work.

We emphasize in closing that our findings pertain to the local fine time scale behavior of option price moves. The estimation method is not a technique for applying the stable distribution to data, and it has nothing to do with the large jumps so often associated with the stable distribution. Rather, the technique provides evidence on the characteristic of the small moves. In other words, our findings pertain to the characteristics of the pole of the Lévy density around the origin which governs the intensity of the very small jumps.

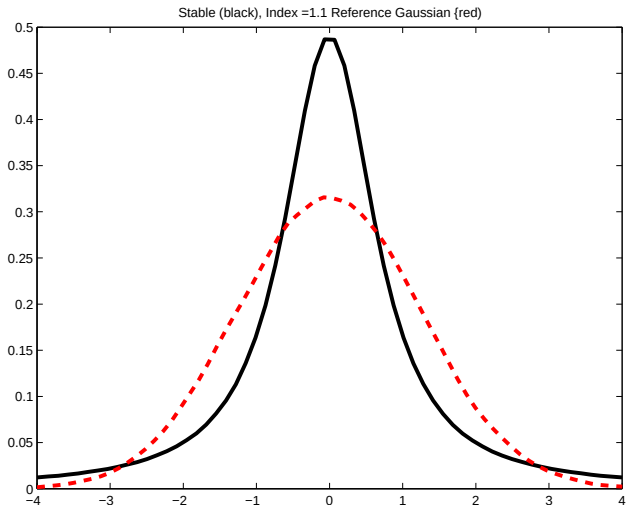

Figure 1: Model implied pdf with $\beta = 1.10$ and reference Gaussian

Table 1: Posterior Distribution of β

Moneyness	Mean	Mode	95% Cred Interval	
IV1	1.0134	1.0073	1.0026	1.0257
IV2	1.1266	1.1231	1.1125	1.1556
IV3	1.3041	1.3090	1.2915	1.3097
IV4	1.5586	1.5617	1.5398	1.5737
IV5	1.7672	1.7706	1.7438	1.7902
IV6	1.8240	1.8210	1.8023	1.8664
IV7	1.8808	1.8808	1.8807	1.8810
Futures	1.8808	1.8808	1.8799	1.8812

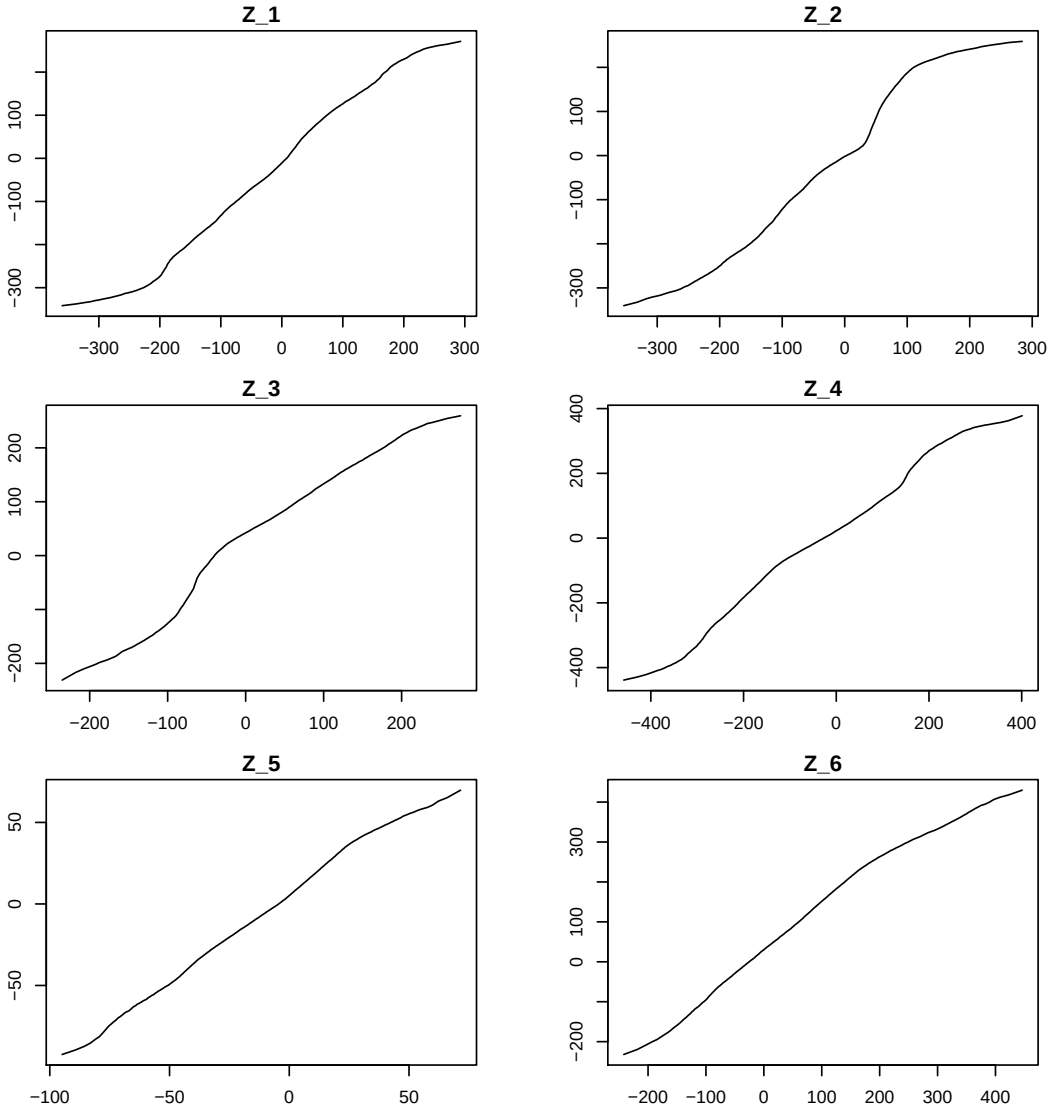

Figure 2: Quantile-quantile plot of z draws vs. normal

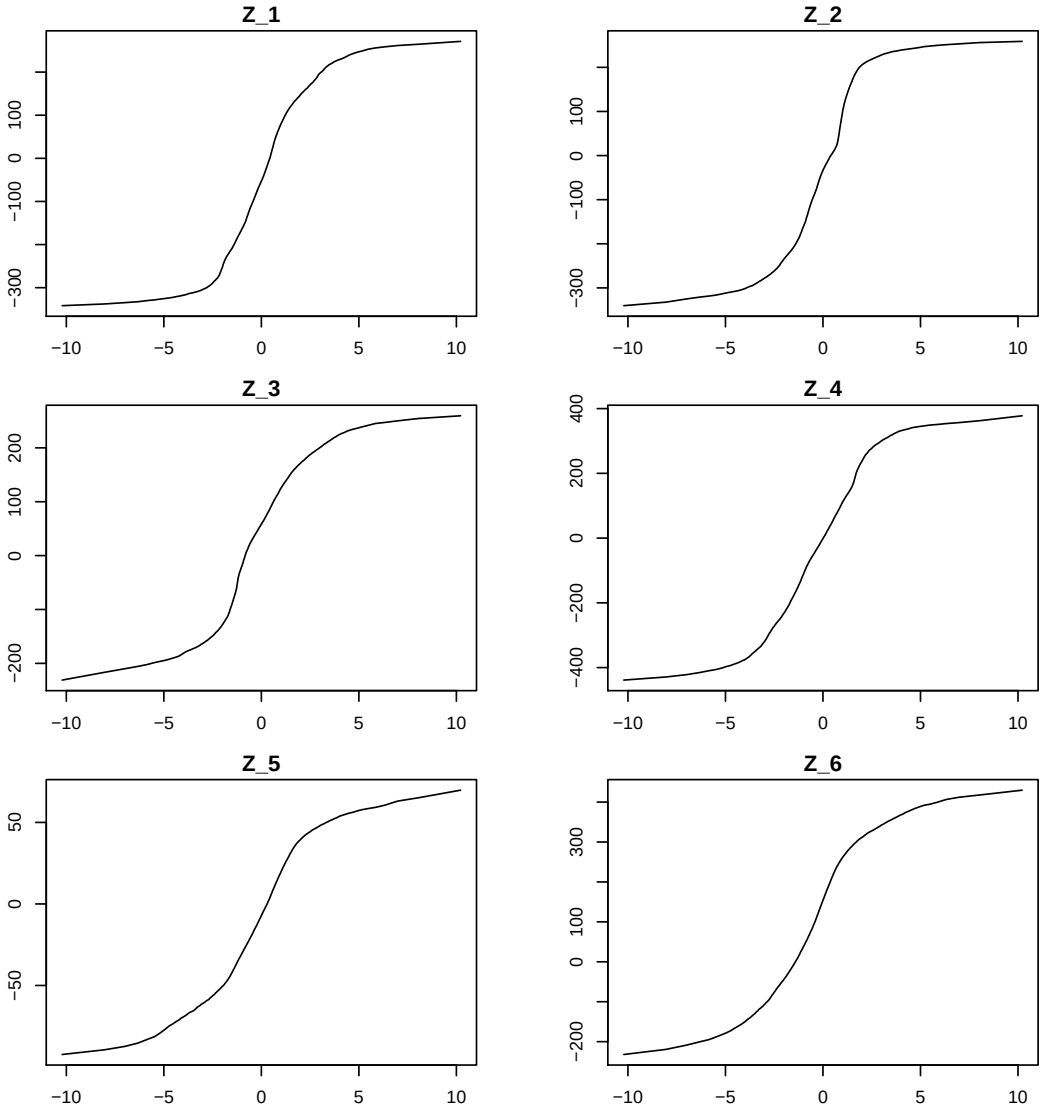

Figure 3: Quantile-quantile plot of z draws vs. 3 d.f. t

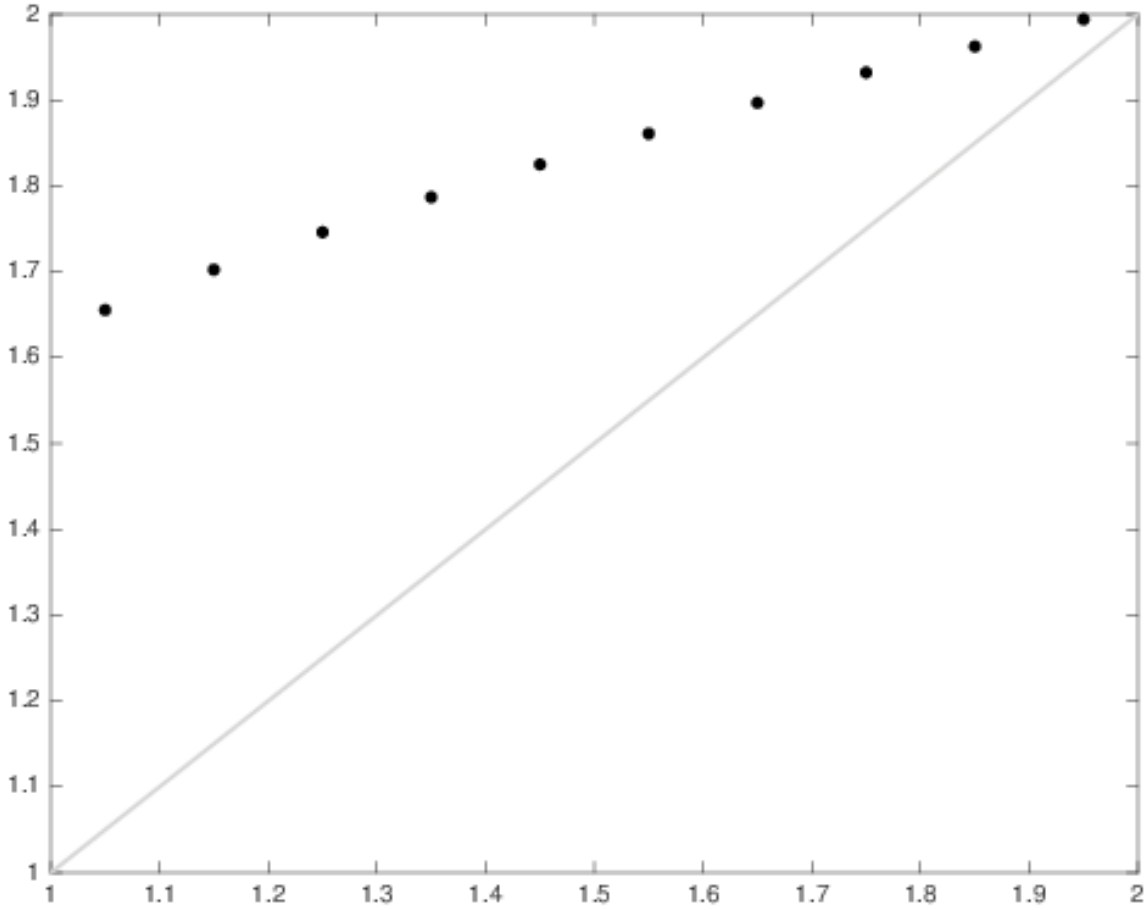

Figure 4: Probability limit (dots) computed via simulation of the power-variation estimation $\hat{\beta}_{pv}$ computed on truncated data versus true β. The x–y coordinated axes are $(\beta, \hat{\beta}_{pv})$.

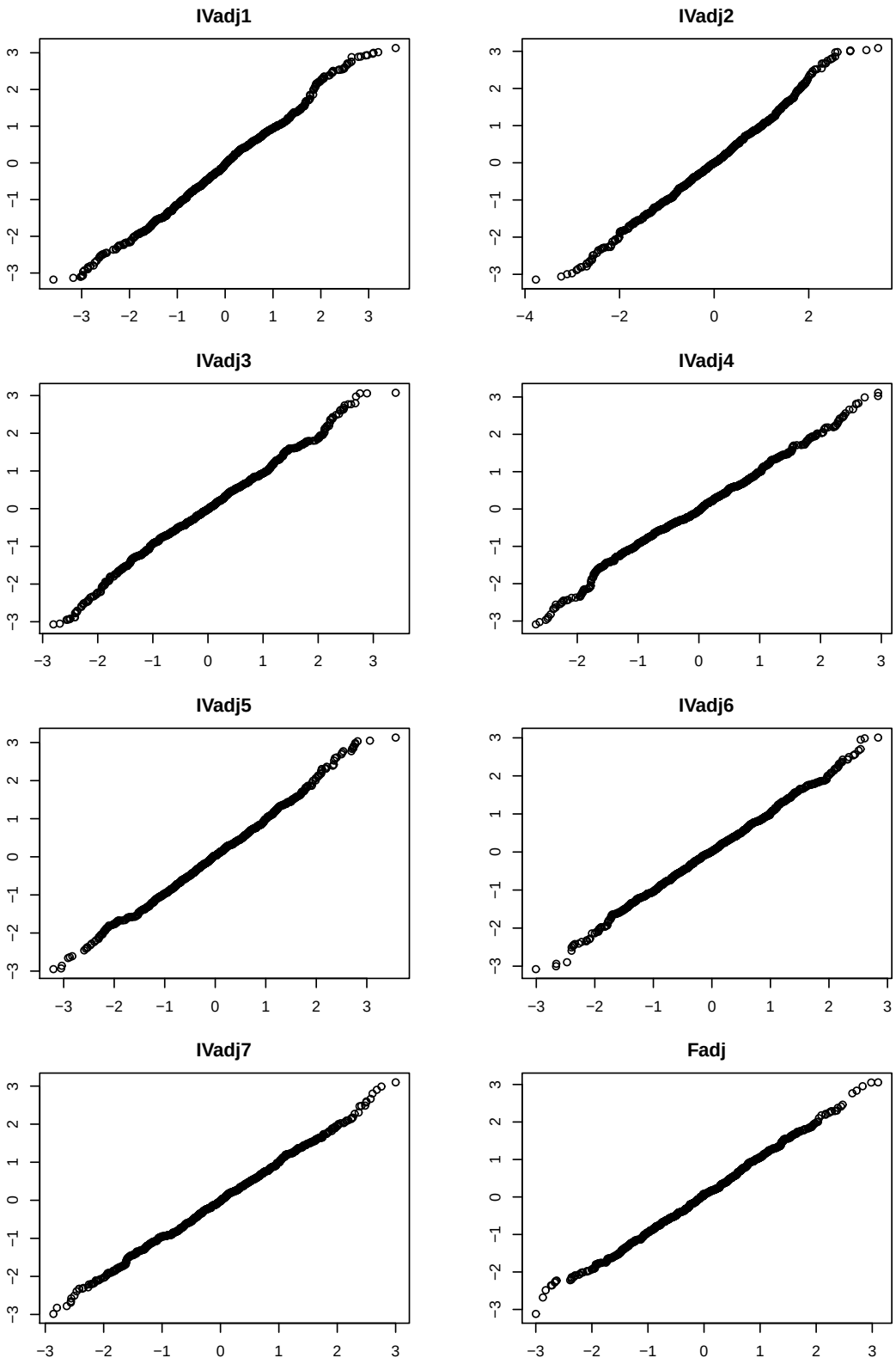

Figure 5: QQ-plots: quantiles of model-generated simulated data versus quantiles of observed data.

References

Ait-Sahalia, Y. and J. Jacod (2014). *High Frequency Financial Econometrics*. Princeton.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T., O. Bondarenko, V. Todorov, and G. Tauchen (2015). The fine structure of equity-index option dynamics. *Journal of Econometrics 187*, 532–546.

Andersen, T. G., O. Bondarenko, and M. Gonzalez-Perez (2013). Uncovering Novel Features of Equity-Index Return Dynamics via Corridor Implied Volatility. Working paper, Northwstern University and University of Illinois.

Barndorff-Nielsen, O. and N. Shephard (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics 2*, 1–37.

Barndorff-Nielsen, O. and N. Shephard (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics 4*, 1–30.

Berger, J. O. and R. L. Wolpert (1988). *Chapter 1: Introduction*, Volume Volume 6 of *Lecture Notes–Monograph Series*, pp. 1–4. Hayward, CA: Institute of Mathematical Statistics.

Chernozhukov, V. and H. Hong (2003). An mcmc approach to classical estimation. *Journal of Econometrics 115*(2), 293 – 346.

Coppejans, M. and A. R. Gallant (2002). Cross-validated snp density estimates. *Journal of Econometrics 110*, 27–65.

Fenton, V. M. and A. R. Gallant (1996). Qualitative and asymptotic performance of snp density estimators. *Journal of Econometrics 74*, 77–118.

Gallant, A. R. (2015). Addendum to reply to reflections. Working paper, Department of Economics, Penn State University. http://www.aronaldg.org/papers/addendum.pdf.

Gallant, A. R. (2016a). Complementary bayesian method of moments strategies. Working paper, Department of Economics, Penn State University. http://www.aronaldg.org/papers/cb.pdf.

Gallant, A. R. (2016b). Reflections on the probability space induced by moment conditions with implications for bayesian inference. *Journal of Financial Econometrics 14*, 229–247.

Gallant, A. R. (2016c). Reply to comments on reflections. *Journal of Financial Econometrics 14*, 284–294.

Gallant, A. R. and J. R. Long (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika 84*, 125–141.

Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55*, 363–390.

Gallant, A. R. and G. Tauchen (1996). Which moments to match?. *Econometric Theory 12*(4), 657 – 681.

Gallant, A. R. and G. Tauchen (2010). Simulated score methods and indirect inference for continuous-time models. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*, Volume 1. North-Holland.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). Boca Raton, FL: Chapman and Hall.

Garcia, R., E. Renault, and D. Veredas (2011). Estimation of stable distributions by indirect inference. *Journal of Econometrics 161*(2), 325337.

Gennady, S. and M. S. Taqqu (2000). *Stable Non-Gaussian Random Processes*. Taylor Francis.

Jacod, J. and P. Protter (2012). *Discretization of Processes*. Springer.

Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.

Mykland, P. and L. Zhang (2009). Inference for Continuous Semimartingales Observed at High Frequency. *Econometrica 77*, 1403–1445.

Ng, S. and J.-J. Forneron (2016). The abc of simulation estimation with auxiliary statistics. Working paper, Department of Economics, Columbia University. http://www.columbia.edu/∼sn2294/papers/abc.pdf.

Podolskij, M. and M. Vetter (2009). Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps. *Bernoulli 15*, 634–658.

Todorov, V. (2015). *The Annals of Statistics 43*, 1831–1864.

Todorov, V. (2016). Testing for time-varying jump activity for pure jump semimartingales. *The Annals of Statistics* , .

Todorov, V. and G. Tauchen (2014). Limit Theorems for the Empirical Distribution Function of Scaled Increments of Ito Semimartingales at High Frequencies. *Annals of Applied Probability 24*, 1850–1888.