Time-varying Volatility and the Power Law Distribution of Stock Returns

July 20, 2017

JEL Classifications: C58, D30, G12

Keywords: Tail distributions, high frequency returns, power laws, time-varying volatility
Abstract

While many studies find that the tail distribution of high frequency stock returns follows a power law, there are only a few explanations for this finding. This study presents evidence that time-varying volatility can account for the power law property of high frequency stock returns. Looking across a number of related models, one finds that a conditional normal model with nonparametric volatility provides the best fit. Specifically, a cross-sectional regression of the power law coefficients obtained from stock returns on the coefficients implied by the nonparametric volatility model yields a slope close to one. Further, for most of the stocks in the sample taken individually, the model-implied coefficient falls within the 95 percent confidence interval for the coefficient estimated from returns data.
1 Introduction

A growing literature has documented that the tail distributions of a broad range of data in the natural and social sciences follow a power law, which implies that the tail probability of a series declines as a power function as it increases in value. This finding has drawn attention both due to the simplicity and scalability of the power law distribution as well as for the ubiquity of this relationship across many fields and data sets. As noted by Stumpf and Porter (2012), however, our understanding of the causes and mechanisms that underpin many of these relationships lag the empirical evidence on them. This study examines whether time-varying volatility can help explain one such finding—the power law property of high-frequency stock returns documented by Plerou, Gopikrishnan, Amaral, Meyer, and Stanley (1999), among others.

Gabaix, Gopikrishnan, Plerou, and Stanley (2003, 2006) provide the leading explanation for the power law property of stock returns. These studies argue that the tail distribution of stock returns arises from the price impact of trades initiated by different market participants, who themselves have a tail distribution of assets. Using data on the tail distributions of trades and assets, they demonstrate how a model of price impact can simultaneously explain the tail properties of trades and returns, as these processes in the model inherit the power law distribution of investor assets.

Building on the insight of Clark (1973) and others that mixture distributions have heavy tails, this study examines whether time-varying volatility can generate the power law property. The key insight underlying the study is that a conditional normal model with time-varying volatility may exhibit a power law distribution in the tails. This result is established formally using a stylized conditional normal model with an inverse-gamma distribution for return variance. As such a model is unlikely to accurately capture the distribution of volatility in the data, the empirical analysis focuses on a conditional normal model with a nonparametric volatility distribution. In addition, as stock returns can inherit any power law properties

\footnote{Gopikrishnan, Plerou, Gabaix, and Stanley (2000) find that returns on stock indices exhibit a power law with a power coefficient around 3; Plerou, Gopikrishnan, Amaral, Meyer, and Stanley (1999) find that the returns on individual stocks follow a power law with coefficients ranging from 2.5 to 4; Axtell (2001) shows that the firm size distribution follows a power law with a coefficient of 1; Rozenfeld, Rybski, Gabaix, and Makse (2011) present evidence that the distribution of populated areas follows a power law with a coefficient of 1; Toda (2012) demonstrates that income follows a double power law; and Toda and Walsh (2015, 2017) examine the implications of the double power law distribution of consumption growth on asset pricing models.}
of the volatility distribution, the study also examines a model with a power law distribution for volatility.

Using more than 925 million distinct observations of stock prices at 1 second intervals, this study investigates whether these model can explain the cross-sectional variation in power law coefficients for the 41 stocks that were included in the Dow Jones Industrial Index at some point from 2003 to 2014. The analysis reveals that the conditional normal model with nonparametric volatility provides the best fit. Specifically, focusing on the cross-section of stocks, one finds that the power law coefficients implied by this model co-move one-for-one with those obtained from return data. More strikingly, examining the stocks in the sample individually one finds evidence that, for the bulk of these stocks, one would not reject the hypothesis that the power law coefficient obtained from the conditional normal model with nonparametric volatility differs from that obtained from return regressions. In comparison, while one obtains a positive relationship between the model-implied power law coefficients and the coefficients from stock returns for the other two models, the fit is clearly weaker.

The empirical analysis begins by estimating power law coefficients using 15 minute returns from 2003 to 2014 for the 41 stocks in the sample. The results indicate that a power law fits the tail distribution of stocks returns. The estimated power law coefficients are centered around 3, consistent with the findings of Plerou, Gopikrishnan, Amaral, Meyer, and Stanley (1999). The cross-sectional distribution of the estimated coefficients, which range from 2.09 to 3.46, provides the key source of variation for the subsequent analysis in the study.

Estimating the conditional normal models with time-varying volatility requires measuring volatility at the high frequencies used to measure returns. This poses a challenge, as the realized volatility method developed by Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002) runs into difficulty with market microstructure noise when applied at such high frequencies.\(^2\) As such, the two-scales realized volatility method developed by Zhang, Mykland, and Aït-Sahalia (2005) is used to obtain return volatility within 15 minute intervals, the primary time interval used in the study. Applying this method to returns at 30, 45, or 60 second intervals offset at intervals of 1 second each, one obtains robust measures of return volatility at 15 minute intervals.

First, the study examines a conditional normal model with an inverse-gamma distribu-

\(^2\)Barndorff-Nielsen and Shephard (2004) extends this method to estimate realized covariances and regression coefficients.
tion for 15-minute return volatility; this model has the parametric property that the tail distribution is given by a power law distribution with a power coefficient equal to twice the shape coefficient of the inverse-gamma distribution for volatility. Comparing the power law coefficient obtained from 15-minute returns with that obtained by estimating the model reveals a positive relationship between these two sets of coefficients. However, one rejects the hypothesis that the two sets of coefficients are equal.

Next, the study examines a conditional normal model with a nonparametric volatility distribution. For each stock, the power law coefficient implied by this model is obtained from a simulated data set constructed using the empirical distribution for return volatility at 15 minute intervals. A comparison of the model-implied power law coefficients and the corresponding estimates from return data reveals a close relationship. The correlation coefficient between these two sets of power law coefficients exceeds 0.95. A regression through the origin of the data coefficients on the model-implied coefficients yields slope coefficients between 1.00 and 1.01. These coefficients are statistically indistinguishable from one, providing support for the hypothesis that the two sets of power law coefficients are identical. Testing the model one stock at a time, one finds that the model-implied coefficient falls within the 95 percent confidence interval of the power law coefficient estimated from returns data for the bulk of the stocks in the sample. The ability of the model to fit the data, both in the cross-section and at the individual stock level, indicates that the conditional normal model with time-varying volatility can generate the power law property of stock returns.

One limitation of the above model is that it abstracts from the clustering of volatility observed in the data. As the model-implied power law coefficient is determined by the unconditional distribution of volatility, in principle, this may not be a concern if there are sufficient observations to accurately capture the unconditional distribution of volatility, regardless of the underlying conditional distribution. That said, in order to address this concern, the nonparametric model is estimated using a block sampling scheme that preserves the observed time-series dependency of volatility. One obtains the same key findings from this setting: the slope coefficients of the regression through the origin of the power law coefficients from returns on the model-implied coefficients equal one and, for most of the stocks in the sample taken individually, the model-implied coefficient falls within the 95 percent confidence interval for the coefficient from the return data.

Last, the study examines where a power law distribution for volatility can explain the
realized power law property of stock returns. A regression of the power law coefficients obtained from stock returns on the corresponding coefficients for volatility yield a slope coefficient of 0.88, significantly less than one. In addition, comparing the power law coefficients one stock at a time one finds that, for most of the sample, the power law coefficient for volatility is significantly higher than that for returns. These findings indicate that while the tail distribution of volatility influences the tail distribution of returns, the fit of the model with a power law distribution for volatility is weaker than that of the conditional normal model with non-parametric volatility.

Taken together, these findings suggest that time-varying volatility, a key property of financial markets emphasized in influential studies such as Engle (1982), Bollerslev (1986) and Schwert (1989), can explain the power law property of stock returns. The results indicate that economic forces such as time-varying economic uncertainty (see Bloom (2009)) or changes in the arrival rate of information (see Mandelbrot and Taylor (1967) and Kyle and Obizhaeva (2016)) may account for the remarkable regularity of the tail distribution of stock returns.

This study is organized as follows. Section 2 describes the statistical models used in the study; Section 3 details the data used to test these models; Section 4 presents the results of the empirical analysis; Section 5 examines whether the model can fit the entire distribution of returns; and Section 6 concludes.

2 Models

This section provides a brief overview of power law distributions and presents the main parametric and nonparametric models analyzed in the study.

2.1 Power law distribution

A random variable \( X \) follows a power law distribution in the tail if, for large \( x \), it has the following probability density function:

\[
P(X = x) \propto x^{-\alpha - 1},
\]
where $\alpha$ equals the power coefficient. A power law distribution has the property that its tail distribution is given by the following:

$$P(X > x) \propto x^{-\alpha}. \quad (1)$$

This relationship is typically tested using a log-log plot of the tail distribution $P(X > x)$ on $x$. A linear regression of $\log(P(X > x))$ on $\log(x)$ provides an estimate of the power law coefficient, $\alpha$. When the primary focus is on whether the tail distribution of a given series follows a power law, this regression is run only on values of $x$ that exceed a given threshold.

### 2.2 Parametric model

Consider the following highly stylized parametric distribution for high frequency log returns:

$$r_t \sim N(0, \sigma_t^2) \quad (2)$$
$$\sigma_t^2 \sim \text{Inv-Gamma}(\alpha, \beta).$$

That is, log equity returns follow a conditional normal distribution, with variance ($\sigma_t^2$) drawn from an inverse-gamma distribution with shape parameter $\alpha$ and scale parameter $\beta$. The conditional normal distribution has a long history in understanding asset returns dating back to Engle (1982).\(^3\) The inverse-gamma distribution for variance provides a parsimonious specification that captures time-variation in return volatility.\(^4\)

The joint density function of the above model can be written as:

$$p(r_t, \sigma_t^2) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{r_t^2}{2\sigma_t^2}\right) \times \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_t^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma_t^2}\right), \quad (3)$$

where $\Gamma(\cdot)$ denotes the gamma function. Following the derivation in Poirier (1995), one can

---

\(^3\)Andersen, Bollerslev, Diebold, and Ebens (2001) provide evidence that daily stock returns are well described by a conditional normal distribution.

\(^4\)The above specification is also used in the Bayesian literature as a conjugate informative prior.
show that the marginal distribution of returns follows a Student’s $t$-distribution:

$$p(r_t) = \frac{\Gamma((2\alpha + 1)/2)}{\Gamma(\alpha) \sqrt{2\pi\beta}} \left(1 + \frac{r_t^2}{2\beta}\right)^{-(2\alpha+1)/2}. \quad (4)$$

The above $t$-distribution has a degree-of-freedom equal to twice the shape parameter of the inverse-gamma distribution, $2\alpha$, a mean of 0, and, for $\alpha > 1$, a variance equal to $\frac{\beta}{\alpha - 1}$. Using the fact that the tail probability of a $t$-distribution is given by a power law with power coefficient equal to the degree-of-freedom of the $t$-distribution, one can write the unconditional tail distribution of returns as:

$$P(r_t > r) \propto r^{-2\alpha}. \quad (5)$$

The above derivation shows that the parsimonious conditional normal distribution shown in Equation (2) implies a power law distribution for returns.\(^5\) Strikingly, the power law coefficient is inherited directly from the unconditional distribution of the variance of returns, providing a simple testable prediction.

### 2.3 Nonparametric volatility distribution

An inverse-gamma distribution is unlikely to accurately capture the distribution of stock return variance over time. As such, this study also examines the following model:

$$r_t \sim N(0, \sigma_t^2) \quad (6)$$
$$\sigma_t^2 \sim f(\cdot),$$

where $f(\cdot)$ denotes a nonparametric distribution for variance. While this model lacks the clear parametric implications of the previous model, the nonparametric specification has the flexibility to match the observed empirical distribution of return volatility. This model is subsequently tested by comparing the power law coefficient obtained from simulating this

\(^5\)Weitzman (2007) shows that a $t$-distribution for consumption growth arising from parameter uncertainty may help explain the equity premium.
model using the empirical distribution for return volatility at high frequencies with the power law coefficient obtained using data on returns.

One limitation of the power law literature in general, and the models (1), (2) and (6) in particular, is that they are not invariant to time-scales, as these distributions are not invariant under addition. In practice, this appears to have little impact, as Plerou, Gopikrishnan, Amaral, Meyer, and Stanley (1999) find that power law coefficients remain mostly constant for returns measured over time intervals ranging from 5 minutes to 1 day.

2.4 Power law distribution for volatility

A power law distribution for volatility can also yield a power law distribution for stock returns. As such, the study also considers the following model:

\[ r_t \sim g(\cdot, \sigma_t) \]
\[ \sigma_t \sim \sigma_t^{-\lambda-1}, \]

where \( g(\cdot) \) denotes a thin-tailed distribution with conditional volatility, \( \sigma \), which itself follows a power law distribution with power coefficient \( \lambda \). As \( g(\cdot) \) is thin-tailed, one can show using a change of variables that the distribution for returns, \( r_t \), inherits the power law coefficient \( \lambda \). Compared to the model in the previous section, this model imposes a parametric form for the distribution of volatility, while allowing for flexibility in the conditional distribution of returns.

This model is tested by estimating the power law coefficients for the volatility distribution of each stock, and comparing this with the corresponding coefficients for individual stock returns.

3 Data

The data used in the study are obtained from the Consolidated Trades segment of the Trades and Quotes (TAQ) database accessed through Wharton Research Data Services. The sample

---

6I thank the referee for emphasizing this point.
comprises all stocks that were included in the Dow Jones Industrial Average at some point from 01/01/2003 to 12/31/2014. Including additions and deletions to the index, this list comprises 41 stocks. These stocks were selected as they are among the most liquid stocks traded in the US equity markets. As such, market microstructure concerns are likely to have a smaller impact on measured returns and volatilities for these stocks.

For each of the stocks in the sample, stock prices are measured at 1 second intervals daily from 01/01/2003 to 12/31/2014. For stocks that had multiple trades within the same second, the stock price is obtained from the median of the reported transaction prices. The sample excludes dates with an early close to trading activity. In order to avoid the effect of information released at the beginning or the end of the trading day, stock prices are measured from 9:45am to 3:45pm. Even taking into account that not all stocks are traded every second, the data set consists of more than 925 million distinct stock price observations.

Following the convention in the realized volatility literature, throughout the study, returns are measured as logarithmic returns, rather than arithmetic returns. Given the short time period over which returns are measured, the logarithmic returns are almost identical to arithmetic returns, and this choice has little effect on the results (see also Kelly (2014)).

The empirical analysis in the study primarily uses stock returns and volatilities measured over 15 minute intervals, the time interval used in Gabaix, Gopikrishnan, Plerou, and Stanley (2006). A 15 minute interval provides a sufficient number of price observations to measure volatility within that interval—see the discussion below—while also providing a large number of return observations that enable a precise estimate of the power law coefficient. For robustness, the analysis also examines results at 10 and 30 minute intervals.

### 3.1 Measurement of volatility

The analysis requires measurement of stock return volatility within 15 minute intervals. This poses a challenge, as the traditional approach in the realized volatility literature measures volatility at daily frequencies using either 1- or 5-minute returns (see Andersen, Bollerslev,

---

7I thank Yesol Huh for providing data on early closing dates.

8Given a stock price process, $Y_t$, logarithmic returns are given by $\log\left(\frac{Y_{t+1}}{Y_t}\right)$, while arithmetic returns are given by $\frac{Y_{t+1}}{Y_t} - 1$.

9The stock price at the end of each 15 minute interval is obtained as the last price at which the stock traded prior to the end of the 15 minute interval.
Diebold, and Labys (2003)). There are too few observations of 1- or 5-minute returns to derive a robust measure of volatility at 15 minute intervals. One cannot address this issue by measuring returns at much higher frequencies, as that increases the market microstructure component of observed returns, thereby adding noise to the realized volatility measures.10

This study tackles this challenge by measuring volatility at 15 minute frequencies using the two-scales realized volatility (hereafter, TSRV) method developed by Zhang, Mykland, and Aït-Sahalia (2005) and Aït-Sahalia, Mykland, and Zhang (2011). They note that measuring realized volatility using returns over 5 minute intervals involves a considerable loss of information and develop a method to measure realized volatility utilizing stock price data at much higher frequencies. This method involves first computing realized volatility using returns at a frequency typically used in the literature, such as a 1 minute frequency. Based on the observation that there are many such overlapping 1-minute return intervals within a period of interest, the above studies show that the average of the realized volatility measure over these overlapping intervals, adjusted for a bias correction term involving realized volatility calculated using returns at the higher frequency, provides an estimate of realized volatility over the desired interval.11 For instance, using a smaller time scale of 1 second, one could obtain realized volatility based on the average of 60 different 1-minute returns, each offset by 1 second. This method enables one to measure realized volatility using stock price information at the much higher frequency by which the returns are offset rather than the lower frequency at which the squared returns are measured. The Appendix provides further details on the measurement of volatility using this method.

The results reported in this study are obtained using the TSRV method applied at frequencies of 30, 45 and 60 seconds offset at 1 second intervals. This enables one to use up to 900 observations of stock prices at 1 second intervals to measure realized volatility within each 15 minute interval. Observations for which the bias correction term causes one of the TSRV measures to be negative are discarded from the sample. Finally, the realized volatility measures in the sample are Winsorized at the 0.01th percentile to remove the effect of extreme outliers.12 This affects only a very small fraction (0.02%) of the observations, and is

10Hansen and Lunde (2006) discusses the effect of market microstructure noise on realized volatility measures as the interval over which returns are measured shrinks.

11The TSRV measure used in the study also incorporates the small sample refinement proposed in Aït-Sahalia, Mykland, and Zhang (2011).

12That is, observations with realized volatility above the 99.99th percentile are set to the value at this percentile while observations below the 0.01th percentile are set to that value.
done primarily to limit the effect of outliers on the parameter estimates of the inverse-gamma model.

### 3.2 Summary statistics

Table 1 reports summary statistics for the 15-minute return and volatility data used in the subsequent analysis. Panel A reports univariate statistics; Panel B reports the correlation coefficients among the TSRV volatility measures obtained using different return frequencies; and Panel C reports the correlation coefficients among the corresponding realized volatility measures. Taken together, the data set used in the analysis includes more than 2 million observations of stock returns and volatilities measured over 15-minute intervals (note that many more stock price observations were used to construct these volatility measures). On a per stock basis, the data set contains about 70,000 15-minute return and volatility data for most of the 41 stocks in the sample. The large sample size helps ensure that small sample concerns do not affect the estimated power law coefficients and that the data set captures the unconditional distributions of return volatility.

The univariate statistics indicate that the mean 15-minute return among the stocks in the sample is slightly positive, reflecting the fact that overall stock returns were positive over the sample period. The corresponding median return equals zero. The statistics also indicate that the TSRV measures provide a robust estimate of realized volatility at 15 minute intervals over the three return frequencies used in the study. Moving from measuring squared returns at a 60 second interval to a 30 second interval leads to only a slight increase in the mean and dispersion of realized volatilities, suggesting that the TSRV approach manages to successfully account for market microstructure noise in the data while incorporating a sizeable quantity of stock price information. The robustness of this approach is further supported by the correlation matrix shown in Panel B, which indicates that the three TSRV measures have a very high correlation with one another.

Panel C reports the corresponding correlation matrix for realized volatility measures at 15-minute intervals obtained using the approach in Andersen, Bollerslev, Diebold, and Labys (2003) applied to squared returns at 30, 45 and 60 second intervals. The results indicate that while the realized volatility measures are closely correlated, the correlations among them are not as strong as those for the TSRV volatility measures. This suggests that the TSRV
approach provides a more robust measure of volatility within 15 minute intervals than the realized volatility method.

4 Results

The empirical analysis tests the models presented in Section 2 by comparing the cross-section of power law coefficients obtained from the tail distribution of returns with the corresponding model-implied coefficients. That is, the tests examine whether models with time-varying volatility can explain the observed variation in the power law coefficients across the stocks in the sample. In comparison, prior studies such as Gabaix, Gopikrishnan, Plerou, and Stanley (2006) have focused on explaining a single power law coefficient obtained by aggregating information from different stocks.

4.1 Power law coefficients from stock returns

The power law coefficients for the tail distribution of stock returns are obtained from regressions of the log probability that the absolute stock return exceeds a given value on the log of the absolute return. These regressions are carried out for absolute stock returns greater than the 95\textsuperscript{th} percentile. The standard error of each regression coefficient is obtained using the method in Gabaix (2009). As the sample consists of about 70,000 returns for most stocks, each regression uses about 3,500 observations on tail probabilities, mitigating small sample concerns.

Figure 1 presents log-log plots of the tail distribution of the absolute value of 15-minute returns for each of the stocks in the sample. It also shows the fitted line obtained from a regression of the log tail probability on the absolute stock return. As the figure shows, a linear regression fits the tail distribution for the absolute returns of the stocks used in the sample well, indicating that the tail distribution of returns follows a power law, consistent with the literature. For some of the stocks in the sample, the fitted line exceeds the implied probability that the absolute return will be greater than a given value for the small number of observations near the very tail of the distribution.\textsuperscript{13}

\textsuperscript{13}This effect is visually enhanced by the log-log plot, which gives equal prominence to values that have tail probabilities between $10^{-4}$ and $10^{-5}$ as those that have a tail probability between $10^{-2}$ and $10^{-3}$, even
Figure 2 gathers the above results and plots the distribution of the estimated power law coefficients for all the stocks in the sample. As the figure indicates, the distribution of the estimated power law coefficients is centered around 3, with a mean value of 2.92 and a median of 3.00. This result matches the finding in Gabaix, Gopikrishnan, Plerou, and Stanley (2003) and others that high frequency stock returns exhibit a power law coefficient of around 3. The estimated power law coefficients exhibit some dispersion, with a cross-sectional standard deviation of 0.32. The subsequent empirical analysis examines whether the models presented in Section 2 can capture this cross-sectional variation. In comparison, prior studies have not focused on examining the cross-section of power law coefficients observed in the data.

4.2 Estimation of parametric model

As noted in Section 2.2, a conditional normal model with an inverse-gamma volatility distribution implies a power law distribution for stock returns. This section reports the results from testing this model by comparing the power law coefficients obtained from the data with the model-implied coefficients, given by 2 times the shape coefficient of the inverse-gamma distribution for stock return variance measured at 15 minute intervals. The shape and scale parameters of the inverse-gamma distributions are obtained using maximum likelihood estimation.

Figure 3 plots the power law coefficient obtained from the regression analysis in Section 4.1 on two times the shape parameters of the inverse-gamma distributions for return variance. Panels A, B and C, respectively, plot this correspondence for the variance distributions obtained using the TSRV method with volatility measured within each 15-minute window using returns at 30, 45 and 60 second intervals offset at 1 second. The figures also report the slope of a regression through the origin of the power law coefficients from the data on the model-implied coefficients.

The null hypothesis in this (and subsequent related) analysis is whether the power law coefficients implied by the model equal their data counterparts. This section examines the null hypothesis by testing whether the slope coefficient of a regression through the origin of the power law coefficients from the data on the model-implied coefficients equals 1. Notably, though there are far fewer observations of the former.
the constant term in the regression equals zero by construction to reflect the fact that a simple linear model with a non-zero constant term would be a misspecified model for examining whether the two sets of coefficients are identical. Imposing the restriction that the constants term in the regression equals zero implies that the fitted line obtained from the regression shown in Figure 3 (and in subsequent figures) does not line up with a more conventional best-fit line obtained from a linear regression with a non-zero constant.

As the figure indicates, there exists a clear positive relationship between the power law coefficients obtained from stock returns and the power law coefficient implied by the conditional normal model given in Equation (2). The slope of the linear regressions through the origin for the three volatility measures are positive and significant, albeit significantly greater than 1, implying that one rejects the null that the two sets of coefficients are identical. While stocks with high/(low) model-implied power law coefficients also have a high/(low) coefficient in the data, there is a greater dispersion in the model-implied coefficients than in the data. Further, the model-implied coefficients are lower, on average, than the coefficients from the returns data.

The estimated model uses a highly stylized distribution for volatility. Comparing the empirical distribution for variance to that obtained by fitting the inverse-gamma distribution, one finds that the inverse-gamma distribution has greater mass at the center of the distribution and less mass at the tails than in the data. The positive relationship between the power law coefficients implied by the model and those from the return regressions suggests examining whether a conditional normal model with a more general volatility distribution can provide a better fit for the power law property of stock returns.

4.3 Estimation of conditional normal model with nonparametric volatility

This section examines whether the conditional normal model with a nonparametric distribution for volatility shown in Equation (6) can generate the power law coefficients observed in the data. Using a nonparametric specification allows one to match the observed volatility distribution in the data. However, this poses the limitation that one no longer obtains a parametric derivation of the power law coefficient implied by the model. As such, the model-implied power law coefficients are obtained using simulation methods.
Specifically, for each stock, one obtains the model-implied power law coefficient from a simulated data set of returns, \( r_i, i = 1, ..., S \) constructed as follows:

1. Obtain a draw for variance, \( \sigma_i^2 \), by sampling with replacement from the empirical distribution for variance, \( f(\cdot) \).

2. Obtain a draw for returns, \( r_i \), by sampling from a normal distribution with mean zero and variance \( \sigma_i^2 \).

This simulated sample has a size, \( S \), of 100 times the number of 15-minute return observations in the data, enabling a precise estimate of the power law coefficient implied by the model; the standard errors of the power law coefficients obtained from the simulated data are all less than 0.01. As such, the subsequent analysis abstracts from the estimation error for the model-implied power law coefficients.\(^{14}\)

### 4.3.1 Cross-sectional evidence

Figure 4A plots the power law coefficients obtained from regression analyses of 15-minute returns data on the power law coefficients obtained using simulated data from the conditional normal model with nonparametric volatility distribution. The volatility distribution underlying the model is given by the empirical distribution of 15-minute return volatility obtained using the TSRV method with 30-second returns offset at 1-second intervals. Figures 4B and 4C, respectively, plot the corresponding results for the simulated data sets from the conditional normal models with volatility obtained using the TSRV method applied to 45- and 60-second returns offset at 1-second intervals. The figures also present the slope coefficients, associated 95 percent confidence intervals and fitted lines obtained from a regression through the origin of the power law coefficients from the returns data on the model-implied power law coefficients. As noted earlier, imposing the restriction that the constant term of the regression equals zero enables one to test the hypothesis that the two sets of power law coefficients are identical.

The figures indicate a striking concordance between the power law coefficients obtained from 15-minute returns and those obtained from the conditional normal models. For each of

\(^{14}\)A Monte Carlo analysis using 400 different simulation samples confirms that there is little simulation error in the model-implied coefficients and that this simulation error has no impact on the subsequent analysis.
the three 15-minute return volatility measures, stocks with high/(low) model-implied power law coefficients also have a high/(low) high power law coefficient in the return data. Indeed, the correlation coefficient between the two sets of power law coefficients ranges from 0.96 to 0.98 across the three volatility measures. In addition, the dispersion of the model-implied power law coefficients is also quite similar to that of the data coefficients.

Turning to the regression analysis, one finds that the slope of a regression line through the origin of the power law coefficient from 15-minute returns on the corresponding power law coefficient from the conditional normal models ranges from 1.004 to 1.007. For each of the three volatility measures, the 95 percent confidence interval of the estimated slope coefficient encompasses one. The corresponding $p$-values for the tests that the slope coefficients reported in Figures 4A, 4B and 4C equal one are 0.41, 0.27, and 0.06, respectively. This indicates that, for all these volatility measures, one would not reject the hypothesis that the two sets of power law coefficients are identical. This close fit indicates that the model can explain the cross-section of power law coefficients observed in the data.

Note that the above finding is a stronger result than that typically obtained in empirical studies, which focus on examining whether a given coefficient is positive or not. Instead, the analysis tests the hypothesis that the two sets of coefficients are identical by examining whether the slope coefficient equals a specific value, one, and finds that one fails to reject this hypothesis.

What explains the improved performance of the nonparametric model compared to the results obtained using the parametric model? Examining the fit of the inverse-gamma distribution for volatility, one finds that the model implies less frequent realizations for inverse volatility than observed in the data around the mode of the distributions and more frequent realizations for very low values. As such, the inverse-gamma distribution predicts a higher frequency of very high volatility realizations than observed in the data. As high volatility realizations are associated with high absolute returns, the return distributions implied by the parametric model has fatter tails than the corresponding distributions implied by the nonparametric model. The fatter tails in the parametric model leads to lower estimates for the power law coefficients $\alpha$, as can be seen by comparing the results reported in Figure 3 with those reported in Figures 4A, 4B and 4C.

The above confidence intervals and $p$-values are obtained using OLS standard errors applied to a $t$-distribution with 40 degrees of freedom. The analysis does not use het-
eroskedasticity robust standard errors due to the fact that the finite sample properties of the resulting $t$-statistics are not well known. As such, it is necessary to examine whether the above regressions suffer from heteroskedasticity in the error term. Applying the White (1980) and Breusch and Pagan (1979) heteroskedasticity tests to the regressions for the three volatility measures, one finds little significant evidence of heteroskedasticity.\footnote{The $p$-values for the White test equals 0.04, 0.21 and 0.45, respectively, while the $p$-values for the Breusch-Pagan test range from 0.34 to 0.52.} This finding is confirmed by a visual examination of the residuals. Consistent with these findings, the heteroskedasticity robust standard errors for the slope coefficients are within 2 basis points of the OLS standard errors.

4.3.2 Evidence at the individual stock level

Another way of evaluating the model is to investigate whether, for a given stock, the power law coefficient obtained from the simulated data falls within the 95 percent confidence interval of the power law coefficient estimated using 15-minute returns data. This examines whether the model can explain the data at the individual stock level, thus providing a sterner test of the model. Table 2 presents the results of this analysis. For each of the stocks in the sample, it reports the power law coefficient estimated from the returns data, the associated 95 percent confidence interval, and the model-implied power law coefficients obtained using simulated data for each of the three measures of 15-minute return volatility used in the study. * and **, respectively, denote stocks for which one can reject the hypothesis the model-implied power law coefficient equals that obtained from the return regression at the 95 and 99 percent confidence intervals.

The results show that the model-implied power law coefficient falls within the 95 percent confidence interval for the regression estimate from returns data for between 36 to 38 of the 41 stocks in the sample. This indicates that, for most of the stocks in the sample taken individually, one would not reject the hypothesis that the power law coefficient implied by the conditional normal model is different from the corresponding coefficient obtained from the regression analysis using 15-minute returns. The ability of the model to fit the power law coefficients obtained from returns at the individual stock level provides clear evidence that the conditional normal model with time-varying volatility can explain the power law property of stock returns.
4.3.3 Volatility clustering

The conditional normal models used in the study abstract from the time-series dependency of volatility, highlighted in Engle (1982), Bollerslev (1986) and Bollerslev and Wright (2001), among others. I.e., the estimation of the non-parametric model employs random sampling of the empirical volatility distribution. This approach is used as the models’ implications are based on the unconditional volatility distribution and, in principle, one would be able to accurately obtain the model-implied power law coefficient if one had sufficient observations to capture the unconditional distribution for volatility, regardless of the underlying conditional distribution. The empirical distribution includes about 70,000 15-minute volatility observations over 10 years for most of the stocks in the sample, suggesting that one has sufficient observations to capture the unconditional distribution. Nonetheless, this section uses an alternate sampling scheme to examine whether the results differ when one estimates the model while accounting for the time-series dependency of the volatility distribution.

The nonparametric conditional normal model with volatility clustering is estimated using a block sampling method. That is, the model-implied volatility distribution is obtained by sampling from blocks of 120 consecutive observations for 15-minute volatility from the corresponding empirical distribution. This sampling approach, akin to the block bootstrap method, helps preserve the time series correlation of volatility in the data. The model-implied power law coefficient is obtained from a simulated sample of returns obtained from a conditional normal distribution with these volatility draws.

Figure 5 presents scatter plots of the power law coefficients obtained from the return regression against the coefficients implied by the nonparametric model with volatility clustering. Panels A, B, and C, respectively, present results from measuring volatility at 15 minute intervals using the TSRV method applied to 30, 45, and 60 second intervals offset at one second. Across the three volatility measures and the 41 stocks in the sample, the volatility distribution in the simulated data exhibit an auto-correlation of 0.581, close to the mean auto-correlation coefficient in the volatility data of 0.588. This indicates that the sampling method captures the time-series dependency of volatility in the data.

As the figure indicates, one continues to obtain the key result using this specification. That is, the slope coefficient of the regression through the origin of the power law coefficient

---

\footnote{The overall simulation sample size is approximately equal to 100 times the number of 15-minute return observations, similar to the simulation sample size in the estimation without volatility clustering.}
from the return regression on the model-implied coefficient is statistically and economically indistinguishable from one for each of the three volatility measures. The 95 percent confidence intervals for the slope coefficient encompass one, and the $p$-values for testing whether the slope coefficient equals one ranges from 0.11 to 0.45. This indicates that, for all the three volatility measures, one would not reject the hypothesis that the model-implied power law coefficients differ from those obtained from the return regressions.

Furthermore, comparing the results at the individual stock level one finds that for the three volatility measures, the model-implied power law coefficient falls within the 95 percent confidence interval for the coefficient from the return regression for between 36 to 38 stocks in the sample, similar to that obtained from the nonparametric model without volatility clustering.

One limitation of the analysis is that it does not account for the leverage effect of returns on future volatility (see Christie (1982)). This partly reflects the challenge of nonparametric sampling from the unconditional distribution of volatility while incorporating a leverage effect. It also reflects the fact that, empirically, lagged volatility has much stronger predictive power for future volatility than lagged returns, which suggests that incorporating a leverage effect is unlikely to substantially change the above results.

### 4.3.4 Results from different time intervals

The previous results were obtained using a 15 minute time interval to measure stock returns and volatility. This time interval was chosen partly as it has been used elsewhere in the literature, and partly as it enabled sufficient data on prices at a one-second interval to measure return volatility using the TSRV method. This section examines whether the findings in the study are robust to changes in the time interval over which returns and volatility are measured.

Figure 6 compares the power law coefficients from the data and the corresponding coefficients from the nonparametric conditional normal model using returns and volatility measured at 10 minute intervals.\(^{17}\) Panels A, B, and C, respectively, report results using the empirical distribution for return volatility within 10 minute intervals obtained using the TSRV method applied to 30, 45, and 60 second returns offset at 1 second intervals. The

\(^{17}\)These results are obtained by sampling volatility observations one at a time from the unconditional distribution.
panels also report the slope coefficient and associated 95 percent confidence interval from a regression through the origin of the power law coefficients in the data on the coefficients from the nonparametric model.

The results indicate that while the model continues to closely match the dispersion of power law coefficients observed in the data, the fit is not quite as good as that for the 15-minute returns. The slope coefficients for the regressions through the origin of the data coefficients on the model-implied coefficients derived from volatility measured using the TSRV method are statistically different from one, with \( p \)-values for the test that the coefficient equals one less than 0.01. However, in economic magnitudes, the slope coefficients—which range from 1.019 to 1.024—are quite close to one. Comparing the model-implied power law coefficients with the coefficients obtained from the return data at the individual stock level reveals that, for each of the three volatility measures, the model implied coefficient falls within the confidence interval of the return coefficient for between 30 to 33 of the 41 stocks examined in the study.

Figure 7 presents the corresponding results using stock returns and volatility measured at 30 minute intervals. As above, Panels A, B, and C, respectively, report results using the empirical distribution for volatility obtained using the TSRV method applied to 30, 45, and 60 second returns offset at 1 second intervals. The cross-sectional results are quite similar to those obtained using 10 minute data. The slope coefficients of the regressions through the origin are statistically different from one for each of the three volatility measures, with \( p \)-values for the test that the slope coefficients equal one less than 0.05. While statistically different from one, the slope coefficients—which range from 0.982 to 0.985—are quite close to one in economic terms.

Evaluating the model at the individual stock level, one finds that the fit is similar to that obtained using 15-minute returns (and better than that for the 10-minute returns). Specifically, the model-implied power law coefficient falls within the 95 percent confidence interval for the regression estimate for 37 or 38 of the 41 stocks in the sample. Taken together, these findings indicate that the conditional normal model with a nonparametric distribution for volatility can fit the power law coefficients observed in the data over a range of time intervals.
4.4 Power law distribution for volatility

This section examines whether a power law distribution for volatility can explain the observed power law for returns (see Equation (7)). Indeed, Liu, Gopikrishnan, Cizeau, Meyer, Peng, and Stanley (1999) and Gopikrishnan, Plerou, Amaral, Meyer, and Stanley (1999) find that the distribution of stock return volatility itself follows a power law. This model is examined by estimating the power law coefficients for the distribution of volatility within 15 minute intervals, and comparing them to the coefficients obtained from stock returns. Figure 8 presents the results of this analysis, with panels A, B, and C, respectively, reporting the results obtaining using volatility at 15 minute intervals measured using the TSRV method applied to 30, 45, and 60 second intervals offset at one second.

The comparison indicates that a positive relationship between the power law coefficients for volatility and stock returns, implying that the tail distribution for volatility influences the tail distribution for returns. However, for each of the three volatility measures, the slope of a line through the origin from a regression of the power law coefficient on returns on the corresponding power law coefficients on volatility equals 0.88, significant less than one. This indicates that one would reject the hypothesis that the two sets of coefficients are identical. Further, for the majority of the stocks in the sample, the power law coefficient for the volatility distribution is significantly different from the corresponding coefficient from the return regression.

Taken together, the above findings indicate that the three models examined in the study all generate a positive relationship between the model-implied power law coefficients and those obtained in the data. However, the conditional normal model provides the best fit across the three models, with a cross-sectional slope coefficient of the data coefficients on the model-implied coefficients not significantly different from one. For the other two models, one rejects the hypothesis that the model-implied coefficients are identical to those observed in the returns data.

5 Distribution of normalized returns

One possible concern is that, by focusing exclusively on understanding the power law properties of returns, the above analysis lacks power to reject the model. Examining whether the
model can fit the entire distribution of stock returns provides a more powerful test. As the
previous analysis found that the conditional normal model with non-parametric volatility
provides the best fit among the models considered, this section examines whether this model
can fit the entire distribution of stock returns. The analysis focuses on the observation that
the conditional normal model implies that returns normalized by volatility should follow a
normal distribution with variance equal to 1. I.e.,

$$\frac{r_t}{\sigma_t} \sim N(0, 1).$$

Table 3 presents the mean, standard deviation, skewness and kurtosis of the normalized
returns for all the stocks in the sample, pooled together. The normalized returns have
a slightly positive mean, consistent with the mean return reported in Table 1. Notably,
one finds that the standard deviation of the normalized returns is less than one, indicating
that normalized returns are less volatile than implied by the model. Indeed, a Kolmogorov-
Smyrnoff test for normality rejects the hypothesis that the normalized returns follow a normal
distribution for almost all the stocks in the sample, taken individually. Finally, while the
normalized returns exhibit very little skewness, their kurtosis for two of the three volatility
measures differs somewhat from the value of 3 that would be obtained from a standard
normal.

One can further understand the distribution of normalized returns by examining the
joint probability density of volatility and normalized returns for the pooled sample, shown
in Figure 9. The conditional normal model implies that the distribution of normalized
returns would be independent of volatility. This prediction is clearly rejected in the data.
Instead, one finds that normalized returns exhibit high/(low) volatility when the underlying
volatility itself is low/(high). This finding indicates that the conditional normal model is
unable to fully capture the joint distribution of returns and volatility.\footnote{Measurement error in volatility can also contribute to the observed joint distribution for volatility and normalized returns. Observations for which measured volatility would be higher/(lower) than the true value would have a lower/(higher) volatility for measured normalized returns due to the scaling by the mismeasured volatility.}

The inability of the conditional normal model with nonparametric volatility to fit the
entire distribution of returns does not invalidate the key findings of Section 4. Namely, that

\footnote{This figure is constructed in a similar manner to Ioannides and Overman (2003) and Eeckhout (2004), who present analogous joint probability densities for city size and city growth.}
models with time-varying volatility can help explain the cross-sectional distribution of power law coefficients, and that, among these models, the conditional normal with nonparametric volatility provides the best fit.

6 Conclusion

This study examines whether models with time-varying volatility can help explain the observed tail distribution of stock returns. Examining a simple parametric model with an inverse-gamma distribution for variance, one finds a positive relationship between the model-implied power law coefficients and the power law coefficients obtained from the data; however, the model-implied coefficients exhibit greater dispersion than the data coefficients. The investigation of a conditional normal model with a nonparametric distribution for volatility reveals a close relationship between the model-implied power law coefficients and the corresponding data values. A cross-sectional analysis using a regression through the origin yields evidence that the two sets of power law coefficients are identical. Furthermore, comparing the model-implied power law coefficients with the corresponding coefficients from the data at the individual stock level, one fails to reject the hypothesis that the model-implied coefficient equals the data coefficient for most of the stocks in the sample, providing evidence in favor of the model at the individual stock level. These results are generally robust to different time intervals for measuring volatility, different intervals for estimating power laws, and to a sampling approach that accounts for volatility clustering. Finally, one finds while the model-implied coefficients from a power law distribution for volatility also exhibit a positive relationship with those obtained from data on returns, the fit is not as good as with the conditional normal model with nonparametric volatility.

These findings indicate that a conditional normal model with time-varying volatility can explain the power law property of stock returns. Economically, this suggests that either changes in uncertainty over time or changes in the arrival rate of information may explain the tail distribution of stock returns. Econometrically, the findings indicate that mixture distributions may help provide insights into why many financial and economic data series exhibit power law properties. Further research into the channels underpinning the observed power law distributions of other series, such as the distribution of firms and city sizes, may prove fruitful.
References


Appendix

Measurement of volatility using the TSRV method

This section provides a brief summary of the TSRV method developed in Zhang, Mykland, and Aït-Sahalia (2005) and Aït-Sahalia, Mykland, and Zhang (2011) that is used in the study to measure realized volatility at 15-minute intervals. For further details, see the above studies.

Let \( \tilde{r}_t \) denote the true return process of a security and \( \sigma_{t,15} \) denote the volatility of the true return process over a 15-minute window. Let \( r_{t,15}^1 \) denote the observed returns of the security over 1-second intervals within the 15-minute window. In the absence of market microstructure noise, one could estimate the volatility of the true return process based on the realized volatility of observed returns.

\[
RV_{t,15}^{\text{all}} = \frac{1}{n} \sum (r_{t,15}^1)^2, \tag{A.1}
\]

where the number of 1-second returns within a 15-minute window, \( n \), equals 900.

The above estimate is biased in the presence of market microstructure noise, leading one to instead estimate realized volatility by measuring returns over a sparse interval, such as a 60-second or a 300-second window. Letting \( r_{t,15}^{60} \) denote returns over 60 second windows, the corresponding realized volatility estimate is given by

\[
RV_{t,15}^{\text{sparse}} = \frac{1}{\bar{n}} \sum (r_{t,15}^{60})^2, \tag{A.2}
\]

where \( \bar{n} \) denotes the number of such returns. While this estimate is much less influenced by market microstructure noise, it involves a loss of information.

The method of Aït-Sahalia, Mykland, and Zhang (2011) incorporates additional information into the estimate of realized volatility by measuring returns at 60-second windows offset at \( k \) seconds. Denote the corresponding 60-second returns offset at \( k \) seconds by \( r_{t,15}^{60,k} \). The realized volatility measure obtained from these returns is given by:

\[
RV_{t,15}^{\text{sparse},k} = \frac{1}{\bar{n}} \sum (r_{t,15}^{60,k})^2. \tag{A.3}
\]
One can construct $K$ such measures for $k \in 1, 2, \ldots, K$.

Aït-Sahalia, Mykland, and Zhang (2011) show that one could obtain a consistent estimate of the volatility of the true return process by taking the average of the above realized volatility values and adjusting for a bias-correction term involving the realized volatility measure obtained using 1-second returns. Their TSRV estimator is given by

$$TSRV_{t,15} = \frac{1}{K} \sum_{k=1}^{K} RV_{t,15}^{\text{sparse},k} \bar{n} \frac{1}{n} RV_{t,15}^{\text{all}}.$$  

One limitation of the above estimator is that it convergences only at $n^{1/6}$. As such, this study measures realized volatility using the following small-sample refinement proposed by Aït-Sahalia, Mykland, and Zhang (2011):

$$TSRV_{t,15}^{\text{adj}} = \left(1 - \frac{\bar{n}}{n}\right) TSRV_{t,15}.$$  

The analysis in the study employs the above method to measure volatility at 10, 15 and 30 minute intervals. The time period at which returns are offset, $k$, increments at one second. For each time interval, three estimates of TSRV are obtained using squared returns measured over 30, 45 and 60 second intervals (the corresponding $K$ values equal 30, 45 and 60, respectively).
Calculations based on Consolidated Trades data from the TAQ database. The table reports summary statistics for intraday stock returns and return volatility measured at 15 minute intervals. Panel A reports univariate statistics; Panel B reports the correlation coefficients among return volatility measured using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005); and Panel C reports the correlation coefficients among return volatility measured using the realized volatility approach. The sample consists of all 41 stocks that were included in the Dow Jones Industrial Average at some point from 01/01/2003 to 12/31/2014. The volatility measures for 15-minute returns reported in Panels A and B are obtained using the TSRV method applied to stock returns at 30, 45, and 60 second intervals, respectively, offset at 1 second intervals. The volatility measures for 15-minute returns reported in Panel C are obtained using the realized volatility method applied to stock returns at 30, 45, and 60 second intervals.

### Panel A: Univariate statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock return</td>
<td>5.63e-06</td>
<td>0.00</td>
<td>0.0030</td>
</tr>
<tr>
<td>Return volatility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 seconds</td>
<td>0.0025</td>
<td>0.0019</td>
<td>0.0021</td>
</tr>
<tr>
<td>45 seconds</td>
<td>0.0024</td>
<td>0.0019</td>
<td>0.0021</td>
</tr>
<tr>
<td>60 seconds</td>
<td>0.0023</td>
<td>0.0018</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

Number of observations: 2796632

### Panel B: Correlation of volatility measures obtained using TSRV method

<table>
<thead>
<tr>
<th>Time interval for TSRV measure</th>
<th>30 seconds</th>
<th>45 seconds</th>
<th>60 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 seconds</td>
<td>1.000</td>
<td>0.994</td>
<td>0.984</td>
</tr>
<tr>
<td>45 seconds</td>
<td>0.994</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>60 seconds</td>
<td>0.984</td>
<td>0.997</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Panel C: Correlation of realized volatility measures

<table>
<thead>
<tr>
<th>Time interval for returns</th>
<th>30 seconds</th>
<th>45 seconds</th>
<th>60 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 seconds</td>
<td>1.000</td>
<td>0.934</td>
<td>0.947</td>
</tr>
<tr>
<td>45 seconds</td>
<td>0.934</td>
<td>1.000</td>
<td>0.932</td>
</tr>
<tr>
<td>60 seconds</td>
<td>0.947</td>
<td>0.932</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Calculations based on Consolidated Trades data from the TAQ database. The table compares the power law coefficients obtained from the returns data with those obtained from the conditional normal model with a nonparametric volatility distribution for each of the 41 stocks that were included in the Dow Jones Industrial Average at some point from 01/01/2003 to 12/31/2014. The second column reports the power law coefficient for stock returns obtained from the 15-minute return data, the third column reports the corresponding 95 percent confidence interval, and the fourth to sixth columns report the coefficients obtained from the conditional normal model with volatility within 15 minutes measured using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005). The latter results are obtained by applying the TSRV method to stock returns at 30, 45, and 60 second intervals offset at 1 second intervals, respectively. * and **, respectively, denote observations for which the model implied power law coefficient falls outside the 95 and 99 percent confidence intervals for the power law coefficient from the return data.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Estimated power law coefficient</th>
<th>Confidence interval</th>
<th>Power law coefficient from model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>30 seconds</td>
</tr>
<tr>
<td>AA</td>
<td>2.93</td>
<td>[2.79, 3.06]</td>
<td>2.95</td>
</tr>
<tr>
<td>AIG</td>
<td>2.29</td>
<td>[2.14, 2.45]</td>
<td>2.37</td>
</tr>
<tr>
<td>AXP</td>
<td>2.60</td>
<td>[2.48, 2.72]</td>
<td>2.62</td>
</tr>
<tr>
<td>BA</td>
<td>3.17</td>
<td>[3.03, 3.32]</td>
<td>3.06</td>
</tr>
<tr>
<td>BAC</td>
<td>2.28</td>
<td>[2.17, 2.38]</td>
<td>2.35</td>
</tr>
<tr>
<td>C</td>
<td>2.21</td>
<td>[2.11, 2.31]</td>
<td>2.36**</td>
</tr>
<tr>
<td>CAT</td>
<td>3.01</td>
<td>[2.87, 3.14]</td>
<td>2.92</td>
</tr>
<tr>
<td>CSCO</td>
<td>3.26</td>
<td>[3.11, 3.41]</td>
<td>3.28</td>
</tr>
<tr>
<td>CVX</td>
<td>3.00</td>
<td>[2.87, 3.14]</td>
<td>2.92</td>
</tr>
<tr>
<td>DD</td>
<td>3.10</td>
<td>[2.96, 3.24]</td>
<td>3.01</td>
</tr>
<tr>
<td>DIS</td>
<td>3.08</td>
<td>[2.93, 3.22]</td>
<td>3.02</td>
</tr>
<tr>
<td>EK</td>
<td>3.05</td>
<td>[2.88, 3.23]</td>
<td>3.26*</td>
</tr>
<tr>
<td>GE</td>
<td>2.50</td>
<td>[2.38, 2.61]</td>
<td>2.56</td>
</tr>
<tr>
<td>GM</td>
<td>2.09</td>
<td>[1.96, 2.22]</td>
<td>2.34**</td>
</tr>
<tr>
<td>GS</td>
<td>2.50</td>
<td>[2.39, 2.62]</td>
<td>2.51</td>
</tr>
<tr>
<td>HD</td>
<td>3.07</td>
<td>[2.92, 3.21]</td>
<td>3.00</td>
</tr>
<tr>
<td>HON</td>
<td>3.10</td>
<td>[2.96, 3.25]</td>
<td>3.01</td>
</tr>
<tr>
<td>HPQ</td>
<td>3.26</td>
<td>[3.11, 3.41]</td>
<td>3.22</td>
</tr>
<tr>
<td>IBM</td>
<td>2.93</td>
<td>[2.79, 3.06]</td>
<td>2.82</td>
</tr>
</tbody>
</table>

Contd.
Table 2 (contd.): Power law coefficients from data and nonparametric model

<table>
<thead>
<tr>
<th>Stock</th>
<th>Estimated power law coefficient</th>
<th>Confidence interval</th>
<th>Power law coefficient from model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>30 seconds</td>
</tr>
<tr>
<td>IP</td>
<td>2.92</td>
<td>[2.79, 3.05]</td>
<td>2.83</td>
</tr>
<tr>
<td>JNJ</td>
<td>3.00</td>
<td>[2.86, 3.14]</td>
<td>3.04</td>
</tr>
<tr>
<td>JPM</td>
<td>2.59</td>
<td>[2.47, 2.71]</td>
<td>2.59</td>
</tr>
<tr>
<td>KFT</td>
<td>3.10</td>
<td>[2.94, 3.26]</td>
<td>3.08</td>
</tr>
<tr>
<td>MCD</td>
<td>3.18</td>
<td>[3.03, 3.33]</td>
<td>3.04</td>
</tr>
<tr>
<td>MMM</td>
<td>3.06</td>
<td>[2.92, 3.21]</td>
<td>3.00</td>
</tr>
<tr>
<td>MO</td>
<td>2.76</td>
<td>[2.64, 2.89]</td>
<td>2.95**</td>
</tr>
<tr>
<td>MRK</td>
<td>2.93</td>
<td>[2.79, 3.06]</td>
<td>2.94</td>
</tr>
<tr>
<td>MSFT</td>
<td>3.16</td>
<td>[3.01, 3.31]</td>
<td>3.26</td>
</tr>
<tr>
<td>NKE</td>
<td>3.14</td>
<td>[3.00, 3.29]</td>
<td>3.05</td>
</tr>
<tr>
<td>PFE</td>
<td>3.20</td>
<td>[3.05, 3.35]</td>
<td>3.29</td>
</tr>
<tr>
<td>PG</td>
<td>3.08</td>
<td>[2.94, 3.22]</td>
<td>3.02</td>
</tr>
<tr>
<td>T</td>
<td>2.89</td>
<td>[2.75, 3.02]</td>
<td>2.92</td>
</tr>
<tr>
<td>TRV</td>
<td>2.63</td>
<td>[2.48, 2.78]</td>
<td>2.57</td>
</tr>
<tr>
<td>UNH</td>
<td>2.78</td>
<td>[2.65, 2.91]</td>
<td>2.76</td>
</tr>
<tr>
<td>UTX</td>
<td>3.15</td>
<td>[3.00, 3.29]</td>
<td>2.98*</td>
</tr>
<tr>
<td>V</td>
<td>2.83</td>
<td>[2.65, 3.00]</td>
<td>2.86</td>
</tr>
<tr>
<td>VZ</td>
<td>2.97</td>
<td>[2.84, 3.11]</td>
<td>2.96</td>
</tr>
<tr>
<td>WMT</td>
<td>3.21</td>
<td>[3.06, 3.35]</td>
<td>3.09</td>
</tr>
<tr>
<td>XOM</td>
<td>3.00</td>
<td>[2.86, 3.14]</td>
<td>2.96</td>
</tr>
</tbody>
</table>
Table 3: Summary statistics for normalized returns

Calculations based on Consolidated Trades data from the TAQ database. The table reports summary statistics for the normalized returns over 15-minute intervals for all the stocks in the sample, pooled together. Normalized returns are constructed by scaling stock returns with the corresponding volatility over that interval. The sample comprises the 41 stocks that were included in the Dow Jones Industrial Average at some point from 01/01/2003 to 12/31/2014. The volatility measures for 15-minute returns reported in Panels A and B are obtained using the TSRV method applied to stock returns at 30, 45, and 60 second intervals, respectively, offset at 1 second intervals.

<table>
<thead>
<tr>
<th>Time interval for TSRV measure</th>
<th>30 seconds</th>
<th>45 seconds</th>
<th>60 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.957</td>
<td>0.940</td>
<td>0.916</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.004</td>
<td>0.005</td>
<td>0.007</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.786</td>
<td>2.898</td>
<td>3.055</td>
</tr>
</tbody>
</table>
Calculations based on Consolidated Trades data from the TAQ database. The figure plots the tail distribution of the absolute value of 15-minute stock returns from 01/01/2003 to 12/31/2014 for each of the stocks in the sample. The x-axis plots the absolute return and the y-axis plots the probability that the absolute return exceeds the given value. Both axes are on log scales. The figures also report the slope and fitted line obtained from a regression of the log tail probability on the log absolute return estimated on returns exceeding the 95th percentile of the empirical distribution for the absolute value of stock returns.
Figure 1 (contd.): Tail distribution of stock returns
Calculations based on Consolidated Trades data from the TAQ database. The figure reports the cross-sectional distribution of the power law coefficients for stock returns at 15 minute intervals for the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014.
Figure 3: Comparison of power law coefficients: Data and parametric model

Panel A: Volatility measured using 30-second returns

Panel B: Volatility measured using 45-second returns

Panel C: Volatility measured using 60-second returns

Calculations based on Consolidated Trades data from the TAQ database. The y-axes of the scatter plots present the power law coefficient for stock returns over 15 minute intervals; the x-axes present twice the shape coefficient of the inverse-gamma distribution for stock return volatility within 15 minute intervals. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Panels A, B, and C, respectively, measure volatility using the TSRV method of Zhang, Mykland, and Ait-Sahalia (2005) applied to returns over 30, 45, and 60 seconds offset at 1 second. The dashed lines represent OLS estimates of a regression through the origin.
Calculations based on Consolidated Trades data from the TAQ database. The y-axis of the scatter plot present the power law coefficient for stock returns over 15 minute intervals; the x-axis present the power law coefficient obtained from simulating the conditional normal model with a nonparametric volatility distribution for stock return volatility within 15 minute intervals. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Volatility is measured using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 30 second intervals offset at 1 second. The dashed line represents the OLS fit of a regression through the origin of the power law coefficient from the data on the power law coefficient from the simulated nonparametric model. The square brackets report the 95 percent confidence interval for the slope coefficient of the regression.
Calculations based on Consolidated Trades data from the TAQ database. The y-axis of the scatter plot presents the power law coefficient for stock returns over 15 minute intervals; the x-axis presents the power law coefficient obtained from simulating the conditional normal model with a nonparametric volatility distribution for stock return volatility within 15 minute intervals. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Volatility is measured using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 45 second intervals offset at 1 second. The dashed line represents the OLS fit of a regression through the origin of the power law coefficient from the data on the power law coefficient from the simulated nonparametric model. The square brackets report the 95 percent confidence interval for the slope coefficient of the regression.
Calculations based on Consolidated Trades data from the TAQ database. The y-axis of the scatter plot presents the power law coefficient for stock returns over 15 minute intervals; the x-axis presents the power law coefficient obtained from simulating the conditional normal model with a nonparametric volatility distribution for stock return volatility within 15 minute intervals. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Volatility is measured using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 60 second intervals offset at 1 second. The dashed line represents the OLS fit of a regression through the origin of the power law coefficient from the data on the power law coefficient from the simulated nonparametric model. The square brackets report the 95 percent confidence interval for the slope coefficient of the regression.
Calculations based on Consolidated Trades data from the TAQ database. The y-axes of the scatter plots present the power law coefficient for 15-minute stock returns; the x-axes present the power law coefficient obtained from simulating a nonparametric conditional normal model with clustering. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Panels A, B, and C, respectively, measure volatility using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 30, 45, and 60 seconds offset at 1 second. The dashed line represents the OLS fit of a regression through the origin.
Calculations based on Consolidated Trades data from the TAQ database. The y-axes of the scatter plots present the power law coefficient for 10-minute stock returns; the x-axes present the power law coefficient obtained from simulating the conditional normal model with a nonparametric volatility distribution. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Panels A, B, and C, respectively, measure volatility using the TSRV method of Zhang, Mykland, and Ait-Sahalia (2005) applied to returns over 30, 45, and 60 seconds offset at 1 second. The dashed line represents the OLS fit of a regression through the origin.
Figure 7: Comparison of data and nonparametric model with 30 minute returns

Panel A: Volatility measured using 30-second returns

Calculations based on Consolidated Trades data from the TAQ database. The y-axes of the scatter plots present the power law coefficient for 30-minute stock returns; the x-axes present the power law coefficient obtained from simulating the conditional normal model with a nonparametric volatility distribution. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Panels A, B, and C, respectively, measure volatility using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 30, 45, and 60 seconds offset at 1 second. The dashed line represents the OLS fit of a regression through the origin.
Figure 8: Comparison of data and model with power law distribution for volatility

Panel A: Volatility measured using 30-second returns

![Graph showing power law coefficients for volatility and returns.]

Panel B: Volatility measured using 45-second returns

![Graph showing power law coefficients for volatility and returns.]

Panel C: Volatility measured using 60-second returns

![Graph showing power law coefficients for volatility and returns.]

Calculations based on Consolidated Trades data from the TAQ database. The y-axes of the scatter plots present the power law coefficient for 15-minute stock returns; the x-axes present the power law coefficient for volatility. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Panels A, B, and C, respectively, measure volatility using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 30, 45, and 60 seconds offset at 1 second. The dashed line represents the OLS fit of a regression through the origin.
Calculations based on Consolidated Trades data from the TAQ database. The figure plots the empirical joint density of volatility and normalized returns over 15 minute intervals. The sample comprises the 41 stocks that were, at some point, included in the Dow Jones Industrial Average from 01/01/2003 to 12/31/2014. Volatility with a 15-minute interval is measured using the TSRV method of Zhang, Mykland, and Aït-Sahalia (2005) applied to returns over 60 seconds offset at 1 second.