A Conditional Extreme Value Volatility Estimator
Based on High-Frequency Returns

March 2004

Turan G. Bali
Associate Professor of Finance
Department of Economics & Finance
Baruch College, Zicklin School of Business
City University of New York
17 Lexington Avenue, Box 10-225
New York, New York 10010
Phone: (646) 312-3506
Fax : (646) 312-3451
E-mail: Turan_Bali@baruch.cuny.edu

David Weinbaum†
Assistant Professor of Finance
Department of Finance
Johnson Graduate School of Management
Cornell University
375 Sage Hall
Ithaca, NY 14853
Phone: 607-255-1295
Fax : (607) 254-4590
E-mail: dw85@cornell.edu

JEL classification: G12, C13, C22

Key words: extreme value, realized volatility, high-frequency returns, GARCH, implied volatility

* We thank Robert Engle, Stephen Figlewski, Roni Michaely, Salih Neftci, Kenneth Singleton, Bhaskaran Swaminathan, and Liuren Wu for their extremely helpful comments and suggestions. The suggestions of Linda Allen, Charlotte Hansen, Lin Peng, and Robert Schwartz have been very valuable. Turan Bali gratefully acknowledges the financial support from the PSC-CUNY Research Foundation of the City University of New York. All errors remain our responsibility.
† Corresponding author.
A Conditional Extreme Value Volatility Estimator
Based on High-Frequency Returns

ABSTRACT

This paper introduces a conditional extreme value volatility estimator (EVT) based on high-frequency returns. The relative performance of the extreme value volatility estimator is compared with the discrete-time GARCH and implied volatility models for 1-day and 20-day-ahead forecasts of realized volatility. This is also a first attempt towards detecting any time-series variation in extreme value distributions using high-frequency intraday data. The informational content of EVT is examined in the context of forecasting S&P 100 index volatility. Adjusted-$R^2$ values imply superior performance of the implied volatility index (VIX) and EVT in capturing time-series variation in realized volatility. The forecasting ability of various discrete-time GARCH models turns out to be inferior to VIX and EVT. According to the Theil inequality coefficient and the heteroscedasticity-adjusted root mean squared and mean absolute errors, (1) EVT provides more accurate forecasts than the VIX and GARCH volatility models; (2) VIX generally yields a less accurate characterization of realized volatility than EVT and GARCH models. These results have implications for financial risk management, and are thus relevant to both regulators and practitioners.

Key words: extreme value, realized volatility, high-frequency returns, GARCH, implied volatility

JEL classification: G12, C13, C22
I. Introduction

Modeling and estimating the volatility of economic time series has been high on the agenda of financial economists over the last two decades. Engle (1982) put forward the Autoregressive Conditional Heteroskedastic (ARCH) class of models for conditional variances, which proved to be extremely useful for analyzing financial return series. Since then an extensive literature has been developed for modeling the conditional distribution of stock prices, interest rates, and exchange rates.

Although many studies show that the parameters of different ARCH models are highly significant in-sample, there is mixed evidence that they provide good out-of-sample forecasts of equity return volatility. Some studies [e.g., Akgiray (1989), Frances and Van Dijk (1995), Brailsford and Faff (1996), and Figlewski (1997)] examine the out-of-sample predictive ability of ARCH models, and find that a regression of realized volatility on estimated volatility yields a low $R^2$ statistic. Andersen and Bollerslev (1998) prove that regression methods will produce low $R^2$ values when realized volatility is measured by daily squared returns, even for optimal GARCH forecasts, because squared daily returns are noisy estimates of volatility. They show that intraday returns can be used to construct a realized volatility series that eliminates the noise in measurements of realized volatility. They find remarkable improvements in the forecasting performance of ARCH models when they are used to forecast the new realized series, compatible with theoretical analysis.

An alternative approach to GARCH volatility forecasts is to use implied volatilities from options. Day and Lewis (1992) and Lamoureux and Lastrapes (1993) examine implied volatility (IV) as a source of information. Both studies find that IV contributes a statistically significant amount of information about volatility over the (short-term) forecasting horizon covered by the models, but they also find that IV does not fully impound the information that the model is able to extract from historical prices. Day and Lewis conclude that IV performs as well but no better than forecasts from ARCH models, and mixtures of two forecasts outperform both univariate forecasts. Lamoureux and Lastrapes examine forecasting volatility through option expiration and find that IV alone is less accurate than the models that incorporate historical prices in 8 out of 10 cases. Canina and Figlewski (1993) provide contrary evidence, and indicate that implied volatilities are poor forecasts of volatility, and simple historical volatilities outperform implied volatilities. Christensen and Prabhala (1998) show for a much longer period that while implied volatilities are biased forecasts of volatility they perform better than historical volatility models. Fleming (1998) also provides evidence that implied volatilities are more informative than daily returns when forecasting equity volatility.

The importance of intraday returns for measuring realized volatility is demonstrated by Andersen and Bollerslev (1998), Andersen et al. (2001a,b; 2003), Barndoff-Nielsen and Shephard (2001; 2002a,b),...
and Andreou and Ghysels (2002). They suggest the sum of squared high-frequency intraday returns as an approximation to the daily volatility. The estimate of this quadratic variation is referred to as integrated or realized volatility. Blair et al. (2001) explore the incremental volatility information of high-frequency (5-min) stock index returns. They answer some important empirical questions for the S&P 100 index. Based on their in-sample analysis of low-frequency (daily or weekly) data using GARCH models, Blair et al. find no evidence for incremental information in daily index returns beyond that provided by the implied volatility index (VIX). Their extension of the historical information set to include high-frequency returns suggests that there is only minor incremental information in high-frequency returns, and this information is almost subsumed by implied volatilities. Their out-of-sample comparisons of volatility forecasts show that VIX provides more accurate forecasts than either low- or high-frequency index returns, regardless of the definition of realized volatility and the horizon of the forecasts.

Volatility is a central concept in finance, whether in derivative pricing, asset allocation, or risk management. Despite the importance of conditional volatility, the existing literature has not yet reached an agreement on whether implied, GARCH or stochastic volatility estimators provide more accurate forecasts of realized volatility.

This paper introduces a conditional extreme value volatility estimator (EVT) based on high-frequency returns, and compares its performance with implied and GARCH volatility models for 1-day and 20-day-ahead forecasts of realized volatility. The informational content of the extreme value volatility estimator is examined, and several important questions for forecasting stock market volatility are answered. The relative performance of discrete-time GARCH, VIX, and EVT estimators for 1-day and 20-day-ahead forecasts of realized volatility is found to be sensitive to the choice of performance measures. The adjusted-$R^2$ values imply superior performance of the implied volatility index and EVT in capturing time-series variation in realized volatility. The forecasting ability of discrete-time GARCH models with alternative distribution functions turns out to be inferior to VIX and EVT. In fact, there appears to be no incremental information in GARCH estimates that adds to the explanation of realized volatility. Although the adjusted-$R^2$ provides the direction and magnitude of the relationship between the realized and estimated volatilities, it does not measure how far the volatility forecasts are away from the realized standard deviation. When the summary statistics are calculated based on the deviation between forecasts and realizations, the relative performance of alternative volatility models changes. According to the Theil inequality coefficient, the heteroscedasticity-adjusted root mean squared and mean absolute errors, EVT

1 We apply the extreme value theory (hence the acronym EVT) based approach to S&P 100 index returns to compare the empirical performance of EVT with an implied volatility index (VIX). The same approach can easily be applied to any financial asset returns including exchange rate, interest rate, equity, commodity, and futures returns.
provides more accurate forecasts than the VIX and GARCH volatility models, and VIX generally yields less accurate characterization of realized volatility than EVT and GARCH models.

Several important results emerge from our analysis. First, we show that it is possible to construct good volatility forecasts based solely on intraday extreme returns. In fact, the informational content of intraday extreme returns is such that the resulting volatility model fares well when compared to the traditional stalwarts (e.g., implied volatility and GARCH). Perhaps even more surprising is the fact that EVT compares favorably to forecasts derived directly from realized volatility (as in, e.g., Andersen et al. 2003), which aggregate squared intraday returns and therefore, by construction, exploit the information in the entire series of intraday returns. Of particular practical interest, EVT is easily implemented using standard maximum likelihood methods, and is thus a useful tool for measuring and forecasting volatility.

Second, we show that standard extreme value theory, which routinely assumes iid data, can be useful in analyzing asset returns, which are, of course, conditionally heteroskedastic and thus not i.i.d. We accomplish this (i) by focusing on per-period (daily) extremes, as opposed to exceedances over high thresholds, and (ii) by modeling explicitly the dependence structure in the per-period conditional extreme value distributions. As a byproduct of our analysis, we find that there is in fact substantial time series variation in daily extreme value distributions using intraday data.

Third, our analysis has implications for financial risk management, and is thus relevant to both regulators and practitioners. In recent years, a central issue in risk management has been to determine capital requirements for financial institutions to meet catastrophic market risk. The standard volatility models can be successful in estimating the maximum likely loss of an institution under normal market conditions. However, as pointed out by Longin (2000) and Bali (2001), the traditional volatility measures based on the distribution of all returns cannot produce accurate estimates of market risks during highly volatile periods. Longin and Bali introduce an unconditional extreme value approach to calculating value at risk (VaR). Although their risk measurement technique is based on a sound statistical theory, it does not yield VaR measures which reflect the current volatility background. Given the serial correlation and conditional heteroscedasticity of most financial data, the use of unconditional volatility is a major drawback of any kind of VaR estimator. The conditional methods developed in this paper can be readily applied to extend and improve the unconditional extreme value risk management approach of Longin and Bali.2 This is important because financial institutions calculate and monitor their maximum likely loss

---

2 Funds managers report two types of realized returns. One is average quarterly realized returns, and the other reported by hedge funds often on demand is the maximum drawdown, which is in fact the extreme loss for the life of the hedge fund. Therefore, the conditional distribution of extreme returns will carry additional information about the risk taken by hedge fund managers.
on a daily basis, and a daily VaR can be estimated easily using the conditional extreme value approach proposed in the paper.

The paper is organized as follows. Section II provides alternative volatility models. Section III describes the data on intraday and daily index returns, and daily realized, implied, and GARCH volatilities. Section IV presents the estimation results. Section V discusses the in-sample and out-of-sample performance of implied, GARCH, and extreme value volatility estimators. Section VI provides extensions and robustness of our findings. Section VII concludes the paper.

II. Alternative Volatility Models

The empirical performance of alternative volatility forecasts requires a measure of volatility realizations to assess fit. While the traditional approach of using squared daily returns for this purpose has been criticized in the literature [see, e.g., Andersen and Bollerslev (1998)], one could imagine applications in which this would indeed be the quantity of interest. In this paper, we assess fit with respect to both measures of volatility: (i) realized or integrated volatility (sum of squared high frequency returns), and (ii) the traditional measure based upon squared daily returns.

A. Implied Volatility

To calculate an implied volatility an option valuation model (such as the Black-Scholes (1973) model) is needed as well as inputs for that model (price of the underlying asset, exercise price, risk-free rate, time-to-expiration, dividends) and an observed option price. Many of these variables are subject to measurement errors that may induce biases in a series of implied volatilities.

We use the CBOE implied volatility index (VIX) to mitigate the problems caused by the use of an inappropriate option valuation model [see Harvey and Whaley (1992)] and infrequent trading of the stocks in the index [see Jorion (1995)]. VIX is constructed so that it represents a hypothetical option that is at-the-money and has a constant 22 trading days (30 calendar days) to expiration. VIX is a weighted index of American implied volatilities calculated from eight near-the-money, near-to-expiry, S&P 100 call and put options and it is constructed so as to reduce mismeasurement and smile effects. This makes it a more accurate measurement of implied market volatility.3,4

---

3 For a detailed explanation of the construction of the VIX index, see Whaley (1993) and Fleming et al. (1995).
4 Our empirical analysis uses the original version of VIX. The new VIX, introduced by the CBOE on September 22, 2003, is obtained from S&P 500 index option prices and incorporates information from the volatility skew by using a wider range of strike prices rather than just at-the-money series.
B. Discrete-Time GARCH Models

Following the introduction of ARCH processes by Engle (1982) and their generalization by Bollerslev (1986), there have been numerous refinements of this approach to modeling conditional volatility.\(^5\) This paper uses the linear symmetric GARCH model of Bollerslev (1986) and the absolute value GARCH process of Taylor (1986) and Schwert (1989) to estimate time-varying conditional volatility.\(^6\) Most asset pricing models postulate a positive relationship between stock market expected return and risk. The following GARCH-in-mean process is used with alternative density functions to model the relation between mean returns on a stock portfolio and its conditional volatility:

\[
R_t = \alpha_0 + \alpha_1 \sigma_{t-1} + z_t \sigma_{t-1} = \mu_{t-1} + z_t \sigma_{t-1}
\]  

(1)

GARCH:

\[
\sigma_{t-1}^2 = \beta_0 + \beta_1 z_{t-1}^2 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2
\]

(2)

TS-GARCH:

\[
\sigma_{t-1} = \beta_0 + \beta_1 |z_{t-1}| \sigma_{t-1} + \beta_2 \sigma_{t-1}
\]

(3)

where \(R_t\) is daily return on S&P 100 index for period \(t\) defined in the standard way by the natural logarithm of the ratio of consecutive daily closing index levels. \(\mu_{t-1} = \alpha_0 + \alpha_1 \sigma_{t-1}\) is the conditional mean for period \(t\) based on the information set up to time \(t-1\), \(\Omega_{t-1}\). \(\sigma_{t-1}^2\) and \(\sigma_{t-1}\) are the conditional variance and standard deviation of daily returns based on \(\Omega_{t-1}\). In equations (2)-(3), the current conditional volatility is defined as a function of the last period’s unexpected news (or information shocks), \(z_{t-1}\), and the last period’s volatility, \(\sigma_{t-1}\).

Most empirical work shows that the excess kurtosis values for financial returns are extremely high and statistically significant, implying that the tails of the empirical distribution are much thicker than the tails of the normal distribution. In light of the empirical evidence of fat-tailed errors, Bollerslev (1987) and Nelson (1991) use leptokurtic distributions such as the standardized \(t\) and the generalized error distribution (GED). In this paper we use the fat-tailed conditional GED density given by\(^7\)

\[
GED : f(R_t; \mu_{t-1}, \sigma_{t-1}, \nu) = \frac{\nu \exp[-(1/2)|z_{t}/\lambda|^{\nu}]}{\lambda \Gamma(1/\nu) \lambda^{(\nu+1)/\nu}}
\]

(4)


\(^6\) At an earlier stage of the study, the asymmetric GARCH models of Engle and Ng (1993), Glosten et al. (1993), and Zakoian (1994) were used also. To save space, we do not present the empirical results from the NGARCH, GJR-GARCH, and TGARCH models; they are very similar to those reported in our tables.

\(^7\) At an earlier stage of the study, we used the fat-tailed GED and Student-\(t\) as well as the thin-tailed normal distribution. Since the results turn out to be similar, and the GED TS-GARCH model performs slightly better than alternative GARCH specifications, we report results from the GED TS-GARCH model. The full set of details is available upon request.
where \( z_i = \frac{R_i - \mu_{it-1}}{\sigma_{it-1}}, \lambda = \left[ \frac{2^{(-2/v)} \Gamma(1/v)}{\Gamma(3/v)} \right]^{1/2}, \Gamma(a) = \int_0^\infty x^{a-1}e^x dx \) is the gamma function, and \( v \) is a positive parameter, or degrees of freedom governing the thickness of the tails.\(^8\)

**C. A Conditional Extreme Value Volatility Estimator**

This section introduces a new approach that uses only the extremes in the stock price process to analyze the dynamic behavior of financial return volatility. The key insight is that volatility dynamics can be directly related to the changing parameters of extremal distributions. That is to say, volatility dynamics can be captured to a large extent by a parsimonious extremal distribution theory, albeit with time-varying parameters.\(^9\)

As pointed out by Diebold, Schuermann and Stroughair (1998) and Diebold and Schuermann (1999), a pitfall of using the standard extreme value theory in financial applications is that it is developed almost exclusively for iid series, whereas financial return series are serially correlated and heteroskedastic, and thus not iid. Following Diebold, Schuermann and Stroughair (1998), we focus on per-period extremes, and not on exceedances over thresholds. Given that our objective is to construct conditional daily volatility forecasts, we take the basic period to be one day, and model the extreme intraday returns within the day. Of importance, we do not assume that the daily extremes (e.g., the largest 5-minute returns each day) are iid; rather, we model the dependence in daily extremes explicitly.

To set forth notation, let \( \ln P_t \) denote the time \( t \geq 0 \) log level of the S&P 100 index, with the unit interval corresponding to one day. The discretely observed time series process of index returns with \( q \) observations per day, or a return horizon of \( 1/q \), is then defined by

\[
R_{(q)t} = \ln P_t - \ln P_{t-1/q}
\]

where \( t = \{1/q, 2/q, \ldots\} \). In our empirical analysis, we use 5-min returns \( (q = 79) \), 15-min returns \( (q = 26) \), and 30-min returns \( (q = 13) \) to check the robustness of our results.\(^10\)

---

\(^8\) For \( v = 2 \), the GED yields the normal distribution, while for \( v = 1 \) it yields the Laplace or the double exponential distribution. If \( v < 2 \), the density has thicker tails than the normal, whereas for \( v > 2 \) it has thinner tails.

\(^9\) The extreme value distributions have recently become popular in the finance literature because of their superior performance in estimating VaR, see e.g. Longin (2000), McNeil and Frey (2000), and Bali (2001) for the risk management performance of alternative extreme value distributions.

\(^10\) We present our results from 5-min returns because the daily maxima and minima are obtained from the intraday data, and we prefer to generate extremes from a larger intraday sample. Using the highest frequency provides larger intraday sample, and yields more reliable extreme observations [Andersen et al. (2000) use the volatility signature plots to determine the optimal sampling frequency. Fleming et al. (2003) provide a simple bias correction]. Note that the empirical findings from 15-min and 30-min returns on S&P 100 index turn out to be very similar to those reported in our tables. The parameter estimates and performance measures of extreme value volatility estimators based on 15-min and 30-min returns are available upon request.
Let \( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \) be a sequence of 5-minute returns on intraday\( q = 1/79, 2/79, 3/79, \ldots, 79/79 \) on day \( t \). Extremes are defined as the maxima and minima of the \( q \) random variables \( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \). Let \( M_{q,t} \) represent the highest (maximum) and \( m_{q,t} \) denote the lowest (minimum) 5-minute returns over trading day \( t \):

\[
M_{q,t} = \max\left( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \right)
\]

\[
m_{q,t} = \min\left( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \right)
\]

To find a limit distribution for maxima, the maximum variable, \( M_{q,t} \), is transformed such that the limit distribution of the new variable is a non-degenerate one. Following Fisher-Tippett (1928) theorem, the variate, \( M_{q,t} \), is reduced with a location parameter, \( \mu_{q,t} \), and a scale parameter, \( \sigma_{q,t} \), in such a way that

\[
x_t = \frac{(M_{q,t} - \mu_{q,t})}{\sigma_{q,t}} \xrightarrow{d} H_{\text{max}}(x).
\]

Assuming the existence of a sequence of such coefficients \( \{\mu_{q,t}, \sigma_{q,t} > 0\} \), three types of non-degenerate distributions are obtained for the standardized maximum:

- **Frechet**:
  \[
  H_{\text{max},d}(x) = \exp\left(-x^{-1/\xi}\right)
  \]

- **Weibull**:
  \[
  H_{\text{max},d}(x) = \exp\left(-(-x)^{-1/\xi}\right)
  \]

- **Gumbel**:
  \[
  H_{\text{max},0}(x) = \exp\left[-\exp(-x)\right]
  \]

Jenkinson (1955) proposes a generalized extreme value (GEV) distribution, which includes the three limit distributions in (8)-(10), distinguished by Gnedenko (1943):

\[
H_{\text{max}}(M; \mu, \sigma, \xi) = \exp\left\{ -\left[1 + \xi \left(\frac{M - \mu}{\sigma}\right)^{-1/\xi}\right]^{-1/\xi} \right\}; \quad 1 + \xi \left(\frac{M - \mu}{\sigma}\right) \geq 0
\]

where \( \xi \) is a shape parameter. For \( \xi > 0 \), \( \xi < 0 \), and \( \xi = 0 \) we obtain the Frechet, Weibull and Gumbel families, respectively. The Frechet distribution is fat tailed as its tail is slowly decreasing; the Weibull distribution has no tail – after a certain point there are no extremes; the Gumbel distribution is thin-tailed as its tail is rapidly decreasing. The shape parameter \( \xi \), called the tail index, reflects the fatness of the distribution (i.e., the weight of the tails), whereas the parameters of scale, \( \sigma \), and of location, \( \mu \), determine the mean and standard deviation of extremes along with \( \xi \).

One potential problem with equations (8)-(11) is that the 79 intraday returns are not i.i.d. As explained in Appendix A, when the data are dependent, it is possible that extreme value distributions cannot be described as in (8)-(11). Not surprisingly, no general statement can be made without further assumptions on the exact nature of the dependence structure. However, the theory of extremes for the case of dependence, while not completely developed, has identified a number of empirically relevant
cases for which inferences based on standard extreme value theory remain valid. For example, this is true if the intraday returns are stationary, and follow an $MA(q)$, $AR(p)$, or $ARMA(p,q)$ model (see Appendix A for a summary, and Leadbetter et al. (1983), Resnick (1987) and Castillo (1988) for detailed treatment). Because these conditions describe our data reasonably well, we prefer to stick to the extreme value distributions described in (8)-(11) as opposed to making assumptions on the dependence structure that would lead to other forms of extreme value distributions (as in, e.g., eq. (21) in Appendix A).

To make sure that our results are not driven by a misspecified conditional extreme value distribution, we follow Diebold, Schuermann and Stroughair (1998) and conduct the following robustness check: we first fit a conditional mean-volatility model to the raw intraday returns, standardize the data by the estimated conditional mean and volatility, and then repeat our analysis based on these standardized residuals. The results (presented in Section VI) remain qualitatively the same and highlight the superior performance of EVT.

Our main objective is to estimate the conditional volatility of extremes based on high-frequency intraday data. Appendix B presents the first and second moments of the maxima and minima for the generalized extreme value distribution. Equation (12a) indicates that the time-varying conditional mean of extremes depends on the location ($\mu_i$), scale ($\sigma_i$), and shape ($\xi_i$) parameters of the GEV distribution. Equation (12b) shows that the time-varying conditional standard deviation (or volatility) of extremes is determined by the scale and shape parameters of the GEV:

\begin{align}
  \text{Mean}_{GEV} &= \mu_i + \frac{\sigma_i}{\xi_i} \left( \Gamma(1-\xi_i) - 1 \right) \\
  \text{Volatility}_{GEV} &= \frac{\sigma_i \left\{ \Gamma(1-2\xi_i) - \left[ \Gamma(1-\xi_i) \right]^2 \right\}^{0.5}}{\xi_i} 
\end{align}

To generate a time-varying conditional extreme value volatility estimator, we first specify the location parameter of the GEV distribution as a linear function of the last period’s extreme returns. The first-order serial correlation between the extreme returns is modeled by introducing an autoregressive of order one, AR(1), process in $\mu_i$:

\[ \mu_i = \mu + \phi M_{i-1} \]  

\[ \text{We should note that the time-variation in extreme value distributions is originally considered by Bali and Neftci (2002) for the generalized Pareto distribution (GPD) of Pickands (1975). They analyze the dynamic behavior of the maximal and minimal changes in daily federal funds rates, and find a significant time-series variation in the conditional mean and standard deviation of the excesses over high thresholds.}\]
Using equation (13), one can test whether the last period’s extremes comprise some significant information which can be used to explain the dynamic behavior of the current extremes. Specifically, one can test whether the coefficient \( \phi \), on \( 1 - t_i M \), is statistically significant.

Second, we parameterize the current scale parameter of the GEV distribution, \( \sigma_{i,t} \), as a function of the last period’s unexpected news, \( \varepsilon_{i,t-1} \), and the last period’s scale parameter, \( \sigma_{i,t} \) :

\[
\sigma_{i,t} = \sigma + \lambda_1 |\varepsilon_{i,t-1}| + \lambda_2 \sigma_{i,t-1},
\]

(14)

where \( \varepsilon_i = M_i - \mu_i \) measures the deviation of the predicted maxima \( \mu_i \) from the actual \( M_i \), and can thus be viewed as an unexpected information shock to the stock market during its largest falls and rises. Equation (14) models the scale parameter as a moving average of the past absolute shocks \( \{|\varepsilon_{i,t-1}|, |\varepsilon_{i,t-2}|, \ldots\} \) since \( \lambda_1 > 0, \lambda_2 > 0 \), and \( \lambda_1 + \lambda_2 < 1 \). Using equation (14), one can determine the presence of volatility persistence in the standard deviation of extremes. Specifically, one can test the hypothesis \( \lambda_1 = \lambda_2 = 0 \) to see if the standard deviation of extremes is constant or accommodates volatility clustering.

Finally, we model the current shape parameter of the GEV distribution, \( \xi_{i,t} \), as a function of the last period’s information shocks, \( \varepsilon_{i,t-1} \), and the last period’s shape parameter, \( \xi_{i,t} \) :

\[
\xi_{i,t} = \xi + \gamma_1 |\varepsilon_{i,t-1}| + \gamma_2 \xi_{i,t-1},
\]

(15)

In equation (15), the tail index that measures the fatness of the distribution (or the weight of the tails) is parameterized as a moving average of the past absolute shocks \( \{|\varepsilon_{i,t-1}|, |\varepsilon_{i,t-2}|, \ldots\} \). Based on the estimated values of \( \gamma_1 \) and \( \gamma_2 \), one can test whether the current weights of the tails are affected by the last period’s unexpected news and the last period’s tail-thickness parameter.

Two parametric approaches are commonly used to estimate the extreme value distributions: (1) the maximum likelihood method which yields parameter estimators which are unbiased, asymptotically normal, and of minimum variance, and (2) the regression method which provides a graphical method for determining the type of asymptotic distribution. In this paper the maximum likelihood method is used to estimate the conditional distribution of extreme high-frequency returns.

The GEV distribution in equation (11) has a conditional density function for the extremes,

\[
h(M_i; \mu_i, \sigma_i, \xi_i) = \left( \frac{1}{\sigma_i} \right) \left( 1 + \xi_i \left( \frac{M_i - \mu_i}{\sigma_i} \right) \right)^{-\frac{1+\xi_i}{\xi_i}} \exp \left\{ - \left[ 1 + \xi_i \left( \frac{M_i - \mu_i}{\sigma_i} \right) \right]^{-\frac{1}{\xi_i}} \right\},
\]

(16)

which gives the conditional log-likelihood function:
\[
\log \mathcal{L}_{GEV} = -T \ln \sigma_i - \left( \frac{1 + \xi_i}{\xi_i} \right) \sum_{t=1}^{T} \ln \left( 1 + \xi_i \left( \frac{M_i - \mu_i}{\sigma_i} \right) \right) - \sum_{t=1}^{T} \left( 1 + \xi_i \left( \frac{M_i - \mu_i}{\sigma_i} \right) \right)^{-\frac{1}{\xi_i}}. \tag{17}
\]

Maximizing the conditional log-likelihood function in equation (17) with respect to the parameters \(\mu, \phi, \sigma, \lambda_1, \lambda_2, \xi, \gamma_1,\) and \(\gamma_2\) yields time-varying scale \((\sigma_i)\) and shape \((\xi_i)\) parameters. Then, we substitute estimated \(\sigma_i\) and \(\xi_i\) in equation (12b) to obtain the conditional extreme value volatility estimator.\(^{12}\)

### III. Data

There are three main types of data: daily index returns, daily implied volatilities, and intraday index returns. The first data set consists of daily closing levels of the S&P 100 index. The time period of investigation for daily index returns extends from 1/2/1987 to 8/31/2000, giving a total of 3,426 observations.\(^{13}\) This data set is used to estimate the conditional volatilities based on the discrete time GARCH models with the generalized error, Student-t, and normal distributions. Because of the autoregressive of order one process in the GARCH (or TS-GARCH) models [see equations (1)-(3)], 3,426 daily index returns generate a time-series of 3,425 daily volatilities from 1/5/1987 to 8/31/2000.

The second data set includes the CBOE’s implied volatility index, VIX, from 1/5/1987 to 8/31/2000, yielding a total of 3,425 daily observations. Daily implied volatilities are computed from an annualized implied volatility index as \(\frac{252}{VIX}\).

The third data set contains high-frequency intraday data from 1/2/1987 to 8/31/2000, giving a total of 269,731 5-min returns. These 5-min returns are constructed from index levels recorded every 15 seconds. As discussed earlier, we use 5-min returns from the S&P 100 index to calculate a measure of realized (or integrated) volatility because this is the highest frequency that previous research uses. To construct the daily realized volatility, we square and then sum 5-minute returns for the period from 9:30 EST to 16:00 EST. The original intraday data are also used to obtain the 1-day maxima and minima. To check the robustness of our findings on time-series variation in extreme value distributions of high-frequency returns, we use 5-min, 15-min, and 30-min returns. 3,426 daily maxima and minima from 1/2/1987 to 8/31/2000 are utilized to estimate the conditional extreme value volatility estimator (EVT).

\(^{12}\) Differentiating the log-likelihood function in equation (17) with respect to the location, scale, and shape parameters yields the first-order conditions of the maximization problem. Clearly, no explicit solution exists to these nonlinear equations, and thus numerical procedures or search algorithms are called for. Details and presentation of alternative statistical estimation methods can be found in Leadbetter et al. (1983), Resnick (1987), Castillo (1988), and Embrechts et al. (1997).

\(^{13}\) The October 1987 crash period is included in our analysis. To check the robustness of our results on the in-sample and out-of-sample performance of alternative volatility models, we repeated the analysis, excluding the October 1987 episode. The empirical findings turn out to be very similar to those reported in our tables. They are available upon request.
Table 1 shows descriptive statistics for 5-min and daily returns. The unconditional mean of 5-min returns is about 0.0007% with a standard deviation of 0.10%. The unconditional mean of daily returns is about 0.057% with a standard deviation of 1.13%. The maximum and minimum values are about 2.25% and -2.25% for 5-minute, and about 8.53% and -23.78% for daily returns. The skewness and excess kurtosis statistics are reported for testing the distributional assumption of normality. The skewness statistics for 5-minute and daily returns are close to zero, but significant at the 1% level. The excess kurtosis statistics are considerably high and significant at the 1% level, implying that the distribution of equity returns has much thicker tails than the normal distribution. The fat-tail property is more dominant than skewness in the sample. The Augmented Dickey-Fuller (ADF) statistic indicates strong rejection of the null hypothesis of a unit root for 5-min and daily returns. The autocorrelations are generally small and not systematically positive or negative, but they are all significant at the 1% level.

Table 1 also provides descriptive statistics for daily realized, implied, GED TS-GARCH, and extreme value volatility estimator (EVT). A notable point is that the means, standard deviations, skewness, kurtosis, maximum and minimum values of daily realized and EVT are very similar especially as compared to VIX and GED TS-GARCH. The kurtosis values for VIX and GED TS-GARCH models turn out to be considerably higher than those for the realized volatility. Overall, the skewness and kurtosis statistics of different volatility measures imply that the volatility distribution is skewed to the right and has much thicker tails than the normal distribution.

IV. Estimation Results

Table 2 presents the maximum likelihood estimates of the time-varying location \( \mu_t \), scale \( \sigma_t \), and shape \( \xi_t \) parameters of the GEV distribution. Asymptotic \( t \)-statistics are given in parentheses. The maximized log-likelihood (Log-L) values are reported for each model to test the presence of time-series variation in the location, scale and shape parameters. The GEV parameters are estimated using the 1-day maximal and minimal returns on the S&P 100 index.

According to the asymptotic \( t \)-statistics shown in Table 2, all of the estimated GEV parameters are statistically significant at the 1% or 5% level. For the extreme return process with constant location \( \mu_t = \mu \) and scale \( \sigma_t = \sigma \) parameters, the estimates of \( \xi \) are positive and highly significant. Specifically, the estimated shape parameter for the maxima \( \xi_{\text{max}} = 0.22 \) is slightly greater than that for
the minima ($\xi_{\text{min}} = 0.18$). Since the higher $\xi$, the fatter the distribution of extremes, the maximal returns have slightly thicker tails than the minimal returns.

A notable point in Table 2 is that for all specifications of the scale and shape parameters, the AR(1) coefficient, $\phi$, in the time-varying location parameter ($\mu_i$) is statistically significant at the 1% level for both the maximal and minimal returns. The maximized log-likelihood values of the models with constant $\mu$ and time-varying $\mu_i$ indicate strong rejection of the null hypothesis $\phi = 0$. The results imply the presence of first-order serial correlation in daily extremes, i.e., the last period’s maxima (or minima) comprise statistically significant information, which can be used to explain the dynamic behavior of the current extremes. Equation (12a) shows that the first moment of the GEV distribution largely depends on the time-varying location parameter. Therefore, we can conclude that the conditional mean of the current maxima (or minima) depends on the last period’s extremes.

As presented in Table 2, the parameters ($\lambda_1, \lambda_2$) are estimated to be positive, and highly significant, indicating substantial time-series variation in the scale parameter ($\sigma_i$). The maximized log-likelihood values of the models with constant $\sigma$ and time-varying $\sigma_i$ indicate strong rejection of the null hypothesis $\lambda_1 = \lambda_2 = 0$. A notable point in Table 2 is that the current scale parameter ($\sigma_i$) of the maxima and minima turns out to be more sensitive to the last period’s scale parameter ($\sigma_{i-1}$) than to the last period’s unexpected news ($\epsilon_{i-1}$) since $\lambda_2 > \lambda_1$ in all cases. Another notable point is that when the current scale parameter is defined as a function of $\epsilon_{i-1}$ and $\sigma_{i-1}$ as in eq. (14), the constant shape parameter ($\xi$) is estimated to be somewhat lower than in the model with constant scale parameter $\sigma$. In other words, the tail-thickness of the maximal and minimal returns is reduced when conditional heteroscedasticity in extremes is modeled with the GARCH-type specification of the scale parameter. The observed leptokurtosis in the maximal and minimal returns is reduced when the extremes are normalized by the time-varying scale parameter, which implies that the presence of excess kurtosis in extreme returns could be related to the time-variation in conditional volatility that depends on $\sigma_i$ and $\xi_i$.

\textsuperscript{14} In all cases, the tail index $\xi$ is estimated to be positive and statistically different from zero. Although not presented in the paper, the likelihood ratio (LR) test between the Frechet case and Gumbel case leads to a firm rejection of the Gumbel distribution (and a fortiori a rejection of the Weibull distribution).

\textsuperscript{15} The observed leptokurtosis in the maximal and minimal returns is reduced when the extremes are normalized by the time-varying scale parameter, which implies that the presence of excess kurtosis in extreme returns could be related to the time-variation in conditional volatility that depends on $\sigma_i$ and $\xi_i$. 
Table 2 provides evidence that the current shape parameter \( (\xi_t) \) is highly affected by the last period’s unexpected news \( (\epsilon_{t-1}) \) and the last period’s shape parameter \( (\xi_{t-1}) \). The parameters \( (\gamma_1, \gamma_2) \) are estimated to be positive and highly significant, implying considerable time-variation in the tail index. The maximized log-likelihood values of the models with constant \( \xi \) and time-varying \( \xi_t \) indicate strong rejection of the null hypothesis \( \gamma_1 = \gamma_2 = 0 \). Similar to our findings for \( \sigma_t \), the current tail-thickness parameter of the maxima and minima turns out to be more sensitive to the last period’s tail index \( (\xi_{t-1}) \) than to the last period’s unexpected news \( (\epsilon_{t-1}) \) since \( \gamma_2 > \gamma_1 \) without any exception. Another notable point in Table 2 is that when \( \xi_t \) is specified as a function of \( \epsilon_{t-1} \) and \( \xi_{t-1} \) as in eq. (15), the degree of first-order serial correlation in extremes is slightly reduced. Specifically, the magnitude of the AR(1) coefficient, \( \phi \), in \( \mu_t = \mu + \phi M_{t-1} \) turns out to be somewhat lower than in the model with constant \( \sigma \) and \( \xi \). Table 2 also shows that unexpected information shocks persist longer in the tails of the minima than in the tails of the maxima because the sum of the parameters \( (\gamma_1 + \gamma_2) \) in \( \xi_t \) is greater for the minimal returns than for the maximal returns. For all specifications considered in the paper, the sum \( (\gamma_1 + \gamma_2) \) is in the range of 0.98 to 0.99 for the minima, and 0.89 to 0.92 for the maxima.

These results point to the presence of volatility clustering and persistence (i.e., ARCH effects) in the standard deviation of high-frequency extreme returns since the conditional volatility of extremes is determined by \( \sigma_t \) and \( \xi_t \) that are characterized by substantial persistence of past information shocks.\(^{16}\)

### V.A In-Sample Performance of Alternative Volatility Models

The in-sample performance of implied, GARCH, and extreme value volatility estimators is first measured by testing their predictive power for the square-root of the sum of squared 5-minute returns on the S&P 100 index.\(^{17}\) We compute the proportion of the total variation in daily realized volatility that can be explained by the estimated conditional variances (or standard deviations). We refer to this as the coefficient of determination, namely the adjusted-\( R^2 \). The predictive power of alternative models is evaluated by estimating a series of OLS regressions of the form:

\[
\text{Realized}_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t-1} + \omega_3 GARCH_{t-1} + \omega_4 + u_t , \tag{18}
\]

\(^{16}\) To save space, we do not present the maximum likelihood estimates of the symmetric and asymmetric GARCH models with GED, Student-t, and Normal density. They are available upon request.

\(^{17}\) The conditional standard deviations of the maximal and minimal returns based on the time-varying location, scale, and shape parameters turn out to be very similar. However, the left tail of the empirical distribution (i.e., the minimal returns) is more extensively studied by the former researchers (especially for financial risk management calculations). Therefore, we prefer to present the relative performance of EVT based on the minimal returns.
where $\text{Realized}_t = \sqrt{\sum_{q=1}^{79} R_{tq/79}^2}$ is the square-root of the sum of squared 5-min returns on S&P 100 index at time $t$. $VIX_t$ is the daily implied volatility index at time $t$, $EVT_{t-1}$ is the daily extreme value volatility estimator at time $t$ given the information set until time $t-1$, $GARCH_{t-1}$ is the daily volatility of index returns at time $t$ given the information set until time $t-1$, and $u_t$ is the forecast error.18

In addition to comparing the adjusted-$R^2$ values of the regressions for each volatility model, we test the statistical significance of $\omega_1$, $\omega_2$, and $\omega_3$. First, we estimate the univariate regressions of the form, 

$$\text{Realized}_t = \omega_0 + \omega_1 VIX_t + u_t$$

$$\text{Realized}_t = \omega_0 + \omega_2 EVT_{t-1} + u_t$$

$$\text{Realized}_t = \omega_0 + \omega_3 GARCH_{t-1} + u_t$$

and then determine the statistical significance of $\omega_1$, $\omega_2$, and $\omega_3$ using the standard $t$-test. Standard errors of the regression coefficients are calculated using the procedure of Newey and West (1987). As shown in Table 3, VIX performs better than GARCH and EVT estimators based on the determination coefficients. The adjusted-$R^2$ values for in-sample forecasts are about 48.47% for VIX, 41.74% for EVT, and 40.17% for the GARCH model. In univariate regressions, the coefficients ($\omega_1$, $\omega_2$, $\omega_3$) on $VIX_t$, $EVT_{t-1}$, and $GARCH_{t-1}$ are found to be significant at the 1% level.

Second, the two-variable regressions of the form: 

$$\text{Realized}_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t-1} + u_t$$

and 

$$\text{Realized}_t = \omega_0 + \omega_1 VIX_t + \omega_3 GARCH_{t-1} + u_t$$

are estimated to determine the marginal contribution of $EVT_{t-1}$ and $GARCH_{t-1}$. Table 3 shows the adjusted-$R^2$ increased only by 0.37% (from 48.47% to 48.79%) after including $GARCH_{t-1}$ to the univariate regression equation $\text{Realized}_t = \omega_0 + \omega_1 VIX_t + u_t$. In addition, the coefficient $\omega_3$ on $GARCH_{t-1}$ is found to be statistically insignificant based on the Newey-West $t$-statistic = 1.29, and the Wald statistic = 1.67 with a $p$-value of 0.1957. Table 3 indicates statistically significant contribution of EVT: adjusted-$R^2$ increased by 3.05% (from 48.47% to 51.52%), and the coefficient $\omega_2$ on $EVT_{t-1}$ is found to be highly significant based on the Newey-West $t$-statistic = 7.01, and the Wald statistic = 49.15 with a $p$-value of zero. The results indicate a strong rejection of the null hypothesis $\omega_2 = 0$, and a failure to reject the null $\omega_3 = 0$ even at the 10% level.

Finally, we estimate the regression equation (16) and test the statistical significance of $\omega_2$ and $\omega_3$ using the standard $t$ and Wald tests. Table 3 implies statistically significant contribution of $EVT_{t-1}$, and almost no contribution of $GARCH_{t-1}$ to the explanation of realized volatility. A notable point is that the adjusted-$R^2 = 51.52\%$ remains almost the same after including $GARCH_{t-1}$ to the bivariate regression

18 At an earlier stage of the study we find that the GED TS-GARCH model outperforms alternative GARCH models with Student-t or Normal density in forecasting realized volatility. Therefore, in eq. (18), we prefer to use the GED TS-GARCH model for $GARCH_{t-1}$ in evaluating the relative performance of implied, GARCH, and extreme value volatility estimators.
equation \( \text{Realized}_t = \omega_0 + \omega_1 \text{VIX}_t + \omega_2 \text{EVT}_t |_{t-1} + \epsilon_t \), and the coefficient \( \omega_3 \) on \( \text{GARCH}_t |_{t-1} \) is found to be statistically insignificant based on the Newey-West \( t \)-statistic = -0.33 and the Wald statistic = 0.11 with a p-value of 0.7430. Table 3 implies strong rejection of the null hypothesis \( \omega_2 = 0 \) because the coefficient \( \omega_2 \) on \( \text{EVT}_t |_{t-1} \) remains highly significant based on the Newey-West \( t \)-statistic = 5.97 and the Wald statistic = 35.64 with a p-value of zero.

Further useful insights can be gained from Figures 1, 2, and 3, which present the time-series plots of realized volatility, VIX, and the estimated conditional volatilities of EVT and GED TS-GARCH models. The striking observation in Figure 1 is that even though VIX provides the most accurate in-sample forecasts based on the adjusted-\( R^2 \) values, it is still a biased predictor of realized volatility. As shown in Figure 1, \( \text{VIX}_t \) overestimates realized volatility, \( \text{Realized}_t \), in most periods from 1/5/1987 to 8/31/2000. The average value of \( \text{VIX}_t / \text{Realized}_t \) is about 1.87 for the sample period. In contrast to implied volatility in Figure 1, the extreme value volatility estimator in Figure 2 and the GED TS-GARCH model in Figure 3 track realized volatility much better.

Figures 1-3 imply that if we remain within the statistical tradition of reporting summary statistics based directly on the deviation between forecasts and realizations, the relative performance of alternative volatility models may change. Therefore, as an alternative to \( R^2 \) measures, we compute the Theil Inequality Coefficient (TIC), which always lies between zero and one, where zero indicates a perfect fit:

\[
TIC = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (\text{vol}_{\text{realized},t} - \text{vol}_{\text{estimated},t})^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (\text{vol}_{\text{realized},t})^2} \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\text{vol}_{\text{estimated},t})^2}}
\]

(19)

where \( \text{vol}_{\text{realized},t} \) is the realized volatility of daily returns at time \( t \), \( \text{vol}_{\text{estimated},t} \) is the estimated conditional variance (or standard deviations) at time \( t \) given the information set until time \( t-1 \) based on the implied, EVT, or GARCH models, i.e., \( \text{vol}_{\text{estimated},t} = \text{VIX}_t, \text{EVT}_t |_{t-1}, \) or \( \text{GARCH}_t |_{t-1} \).

The results in Table 3 highlight the superior in-sample performance of EVT against the implied (VIX) and GARCH volatility models. Specifically, the TIC values for in-sample forecasts of realized volatility are found to be 19.21% for EVT, 28.20% for VIX, and 20.67% for the GED TS-GARCH model. As expected from Figures 1-3, VIX turns out to be inferior to EVT and GARCH based on the deviation between forecasts and realizations.

As an alternative to TIC, we use the heteroscedasticity-adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) to measure how far the estimated volatilities are away from the
realized volatility. In other words, we compute the deviation between $\text{vol}_{\text{realized},t}$ and $\text{vol}_{\text{estimated},t}$ using the mean absolute percentage errors ($HMAE$) and the root mean squared percentage errors ($HRMSE$):\(^{19}\)

$$HMAE = \frac{1}{n} \sum_{t=1}^{n} \left| 1 - \frac{\text{vol}_{\text{realized},t}}{\text{vol}_{\text{estimated},t}} \right|$$  \hspace{1cm} (20)$$

$$HRMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left( 1 - \frac{\text{vol}_{\text{realized},t}}{\text{vol}_{\text{estimated},t}} \right)^2}$$  \hspace{1cm} (21)

These two volatility forecast evaluation criteria follow the statistical tradition of reporting statistics based directly on the deviation between forecasts and realizations, while adjusting for heteroscedasticity in the forecast error. Table 3 provides the $HMAE$ and $HRMSE$ values for in-sample forecasts. The results are very similar to our earlier findings from TIC: the extreme value volatility estimator performs better than the implied and GARCH volatility models, and the GED TS-GARCH model provides more accurate forecasts of realized volatility than VIX. More specifically, the $HMAE$ values for in-sample forecasts of realized volatility are 28.36% for EVT, 42.52% for VIX, and 30.67% for the GARCH model. The $HRMSE$ measures are 30.90% for EVT, 45.35% for VIX, and 35.14% for the GARCH model.

V.B Out-of-Sample Performance of Alternative Volatility Models

The out-of-sample performance of alternative volatility models is evaluated based on their 1-day and 20-day-ahead forecasts of realized variance and standard deviation. The in-sample period is from 1/2/1987 to 12/31/1990 providing 1,008 daily observations, followed by the out-of-sample period from 1/2/1991 to 8/31/2000, yielding 2,417 1-day-ahead and 2,398 20-day-ahead forecasts. To compare the out-of-sample forecasts of implied, GARCH, and extreme value volatility estimators, discrete time GARCH-in-mean models are estimated for daily index returns from 1/2/1987 to 12/31/1990. The most general specification of alternative models is as follows:

$$R_t = \alpha_0 + \alpha_t \sigma_{\text{garch},t-1} + z_t \sigma_{\text{ged},t-1}$$  \hspace{1cm} (22)$$

Conditional Variance:

$$\sigma_{\text{garch},t-1}^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2 + \delta VIX_{t-1}^2 + \gamma EVT_{t-1}^2$$  \hspace{1cm} (23)$$

Conditional Std. Dev.:

$$\sigma_{\text{ged},t-1} = \beta_0 + \beta_1 | z_{t-1} | \sigma_{t-1} + \beta_2 \sigma_{t-1} + \delta VIX_{t-1} + \gamma EVT_{t-1}$$  \hspace{1cm} (24)$$

where $\sigma_{\text{garch},t-1}^2$ and $\sigma_{\text{ged},t-1}$ are the conditional variance and standard deviation of daily index returns based on the information set until time $t-1$. In equations (22)-(24), $z_t$ is drawn from the GED density. Note that

\(^{19}\) The $HMAE$ and $HRMSE$ are used by Andersen, Bollerslev, and Lange (1999), and Andreou and Ghysels (2002).
imposing restrictions on certain parameters in equations (22)-(24) three different conditional variance (or standard deviation) models are obtained using different daily information sets: (1) The GARCH(1,1) model of Bollerslev (1986) or TS-GARCH model of Taylor (1986) and Schwert (1989) with $\delta = \gamma = 0$; (2) A volatility model that uses the latest information in the VIX series alone, $\beta_1 = \beta_2 = \gamma = 0$; (3) A volatility model that uses information in daily extreme returns alone, $\beta_1 = \beta_2 = \delta = 0$.

Time-series forecasts are obtained by estimating rolling VIX, EVT, GARCH, and TS-GARCH volatility models. Each conditional variance and standard deviation model is estimated initially over the 1008 trading days of the in-sample period from 1/2/1987 to 12/31/1990, and forecasts of realized variance and standard deviation are made for the next day, say $t+1$, using the in-sample parameter estimates. The model and data are then rolled forward one day, deleting the observation(s) at time $t-1007$ and adding on the observation(s) at time $t+1$, reestimated and a forecast is made for time $t+2$. This rolling method is repeated until the end of the out-of-sample forecast period. The 1-day-ahead forecasts provide predictions for 12/2/1991 to 8/31/2000 providing time-series of length 2,417. On each day, forecasts are also made for 20-day volatility.

Forecasts are produced at time $t$ for the realized variance and standard deviation, defined by the sum of squared 5-minute returns, $Variance_t = \sum_{q=1}^{79} R_{(q/79),t}^2$ and $StdDev_t = \sqrt{\sum_{q=1}^{79} R_{(q/79),t}^2}$, where $R_{(q/79),t}$ is the 5-minute index return observed in day $t$. For 20-day-ahead forecasts, realized variance and standard deviations are defined as $\sum_{j=1}^{20} Variance_{t+j}$ and $\sqrt{\sum_{j=1}^{20} Variance_{t+j}}$, respectively.

We compute the proportion of the total variation in daily realized variance (standard deviation) that can be explained by the estimated conditional variances (standard deviations). The adjusted-$R^2$ values in this section provide information about how well each model is able to forecast the future volatility of stock market returns. The predictive power of alternative volatility models is evaluated by estimating a series of OLS regressions of the form:

$$\text{Realized}_{t+j} = \omega_0 + \omega_1 \text{VIX}_t + \omega_2 \text{EVT}_t + \omega_3 \text{GARCH}_t + u_t$$

(25)

where $j = 1$ or 20 days, $\text{Realized}_{t+j}$ represents realized standard deviation.

As presented in Table 4, the implied volatility index (VIX) performs better than the GARCH and extreme value volatility (EVT) estimators based on the determination coefficients. More specifically, the adjusted-$R^2$ values for 1-day-ahead forecasts are about 56.69% for VIX, 49.87% for EVT, and 40.41%.
for the GARCH model. The two-variable regressions, \( \text{Realized}_{t+1} = \omega_0 + \omega_1 \text{VIX}_t + \omega_2 \text{EVT}_t + u_t \) and \( \text{Realized}_{t+1} = \omega_0 + \omega_1 \text{VIX}_t + \omega_2 \text{GARCH}_t + u_t \), are estimated to determine the marginal contribution of \( \text{EVT}_t \) and \( \text{GARCH}_t \). Table 4 shows the adjusted-\( R^2 \) increased only by 0.06\% (from 56.69\% to 56.75\%) after including \( \text{GARCH}_t \) to the univariate regression equation \( \text{Realized}_{t+1} = \omega_0 + \omega_1 \text{VIX}_t + u_t \). In addition, the coefficient \( \omega_3 \) on \( \text{GARCH}_t \) is found to be statistically insignificant. Table 5 indicates statistically significant contribution of EVT: adjusted-\( R^2 \) increased from 56.69\% to 58.65\%, and the coefficient \( \omega_2 \) on \( \text{EVT}_t \) is found to be highly significant. We also estimate the regression equation (25) and test the statistical significance of \( \omega_2 \) and \( \omega_3 \) using the standard \( t \) and Wald tests. Table 4 implies statistically significant contribution of \( \text{EVT}_t \) and almost no contribution of \( \text{GARCH}_t \) to the explanation of realized volatility. A notable point is that the adjusted-\( R^2 = 58.65\% \) remain almost the same after including \( \text{GARCH}_t \) to the bivariate regression \( \text{Realized}_{t+1} = \omega_0 + \omega_1 \text{VIX}_t + \omega_2 \text{EVT}_t + u_t \), and the coefficient \( \omega_3 \) on \( \text{GARCH}_t \) is found to be statistically insignificant.

Christensen and Prabhala (1998) indicate that the previous studies by Day and Lewis (1992) and Lamoureux and Lastrapes (1993) are characterized by a “maturity mismatch” problem, in that Lamoureux and Lastrapes examine one-day-ahead and Day and Lewis examine one-week-ahead predictive power of implied volatilities computed from options that have a much longer remaining life. In addition to one-day-ahead forecasts, following Blair et al. (2001), we test the relative of performance of alternative models for the longer forecast horizon of 20 trading days, that closely matches the life of the hypothetical option (22 trading days) that defines VIX. As presented in Table 4, the determination coefficients and the related statistics for 20-day-ahead forecasts imply superior performance of implied volatility index (VIX) and the extreme value volatility estimators (EVT) in capturing time-series variation in realized (or integrated) volatility. The predictive power of discrete-time GARCH models with GED, Student-\( t \), and normal distributions turns out to be inferior to VIX and EVT. In fact, there is only minor incremental information in GARCH models for 20-day-ahead forecasts, and this information is subsumed by implied and extreme value volatility estimators.

As an alternative to \( R^2 \) measures, we compute the Theil Inequality Coefficient (TIC) for 1-day and 20-day-ahead forecasts of alternative models. The results in Table 4 highlight the superior performance of extreme value volatility estimators (EVT) against the implied (VIX) and GARCH volatility models. Specifically, the TIC values for 1-day-ahead forecasts of realized standard deviation are found to be 17.49\% for EVT, 24.83\% for VIX, and 19.54\% for the GARCH model. VIX turns out to be inferior to EVT and GARCH based on the deviation between forecasts and realizations.
We also use the heteroscedasticity-adjusted mean absolute error \((HMAE)\) and root mean squared error \((HRMSE)\) to measure how far the estimated volatilities are away from the realized variance and standard deviation. Table 4 provides the \(HMAE\) and \(HRMSE\) values for 1-day and 20-day-ahead forecasts. The results are very similar to our earlier findings from \(TIC\): the extreme value volatility estimators (EVT) perform better than the implied and GARCH volatility models, and the GED TS-GARCH model provides more accurate forecasts of realized volatility than VIX. More specifically, the \(HMAE\) values for 1-day-ahead forecasts of realized standard deviation are 27.31\% for EVT, 39.64\% for VIX, and 29.96\% for the GARCH model. The \(HRMSE\) measures are 31.41\% for EVT, 42.54\% for VIX, and 34.42\% for the GARCH model.

The \(TIC\), \(HMAE\), and \(HRMSE\) measures for 20-day-ahead forecasts indicate the same ranking of models obtained from 1-day-ahead forecasts. The results in Table 4 imply superior performance of EVT and GED TS-GARCH against VIX in capturing time-series variation in realized volatility.

### VI. Extensions and Robustness

We extend the set of alternative volatility models to include forecasts from modeling integrated volatility directly, and consider the robustness of our results to (i) a different measure of realized volatility, and (ii) possible biases induced by the dependence structure of intraday returns.

#### A. Modeling and Forecasting Realized Volatility Directly

This section models and forecasts realized volatility directly. We use an ARMA\((p,q)\) process for realized volatility and compare the empirical performance of the ARMA\((p,q)\) realized volatility model with the implied, extreme value, and GED TS-GARCH volatility estimators. Specifically, the following ARMA specifications are used to forecast daily realized volatility of S&P100 index returns:

\[
ARMA(5,5): \text{Realized}_t = \alpha_0 + \sum_{i=1}^{5} \alpha_i \text{Realized}_{t-i} + \sum_{i=1}^{5} \beta_i \varepsilon_{t-i} + \varepsilon_t
\]

Following Andersen et al. (2003), we use the lag polynomials up to 5 days (or one week). As will be discussed here, the results are robust across different lag specifications.

The adjusted-\(R^2\) values and the \(TIC\), \(HMAE\), and \(HRMSE\) measures in Table 5 indicate that the in-sample performance of the ARMA model is very similar to that of EVT and VIX, but it performs much better than the GED TS-GARCH model. A notable point in Table 5 is that the 1-day-ahead and especially 20-day-ahead forecasts indicate that the ARMA model cannot outperform the conditional extreme value volatility estimator. The adjusted-\(R^2\) values for the ARMA model are in the range of 45\% for 1-day-ahead and 24\% for 20-day-ahead forecasts, whereas the corresponding values for EVT are
about 50% for 1-day-ahead and 37% for 20-day-ahead forecasts. TIC and HMAE measures presented in Panels B-C of Table 5 and Table 4 indicate that there is no clear evidence whether EVT or ARMA model performs better in predicting the 1-day-ahead and 20-day-ahead realized volatility. However, the HRMSE measures for EVT are smaller than those for the ARMA model.

These results suggest that the information content of extreme intraday returns is such that EVT performs well when compared to the forecasts in (26), which are based on aggregated squared five-minute returns, and thus exploit the information in the entire series of intraday returns.

B. An Alternative Assessment of Fit

The traditional measure of volatility, based on squared daily returns, has been criticized in the recent literature (see, e.g., Andersen and Bollerslev (1998) among others) which advocates the use of squared intraday returns as an essentially error-free and model-free measure of volatility. While there certainly are sound reasons to focus on volatility derived from intraday returns, this measure of volatility may have drawbacks also. For example, 5-minute returns may be contaminated by microstructure noise. To check whether our results are driven by microstructure noise, we repeated the analysis in Section V with 15-minute and 30-minute returns. The results (available upon request) turn out to be very similar to those reported in our tables for 5-minute returns.

Also, in certain applications, it may be of interest to forecast volatility based upon squared (or absolute) daily returns, and not integrated volatility. To this end, we also generate a traditional realized volatility measure based on the residuals of the autoregressive of order five process:

\[ R_{t+1} = \alpha_0 + \alpha_1 R_t + \alpha_2 R_{t-1} + \alpha_3 R_{t-2} + \alpha_4 R_{t-3} + \alpha_5 R_{t-4} + \alpha_6 R_{t-5} + \epsilon_{t+1} \]  

(27)

where \( R_{t+1} \) is daily index return, \( \epsilon_{t+1} \) is the daily conditional mean.21 We run an OLS regression to generate the daily realized volatility.

Table 6 presents the in-sample and out-of-sample performance of alternative volatility models in forecasting the traditional measure of realized volatility. The adjusted-\( R^2 \) values and the TIC, HMAE, and HRMSE measures indicate the superior in-sample and out-of-sample performance of EVT and VIX over the GED TS-GARCH and ARMA(p,q) models. In fact, EVT performs slightly better than VIX. The 1-day-ahead and 20-day-ahead forecasts also indicate that ARMA(p,q) model cannot outperform its alternatives. Table 6 also confirms the findings of Andersen et al. (2003), i.e., the traditional measure of realized volatility cannot be easily modeled or cannot be forecasted directly because the absolute (or squared) residuals are very noisy estimators of day-by-day fluctuations in index returns. Therefore, ARMA(p,q) model performs worse than EVT, VIX, and GED TS-GARCH models in predicting the

---

21 The optimal lag length in eq. (27) is determined using the Akaike Information Criterion and Schwarz Bayesian Criterion.
traditional measure of realized volatility. However, as discussed earlier, when daily realized volatility is measured by high-frequency data, the performance of ARMA(p,q) model is similar to that of EVT and VIX, and it outperforms the GED TS-GARCH model.

C. Dependence Structure of Intraday Returns

As pointed out in Section II, intraday returns are not i.i.d. and, consequently, the limit distribution of extremes need not belong to the domain of attraction of the three standard extreme value distributions (Frechet, Weibull or Gumbel). We verify that our results are not driven by a misspecified conditional extreme value distribution by following Diebold, Schuermann and Stroughair (1998) in conducting the following robustness check: we first fit a conditional mean-volatility model to the raw intraday returns, standardize the data by the estimated conditional mean and volatility, and then repeat our analysis based on these standardized residuals.

We find that an AR(1) GJR-GARCH(1,1) model with GED-distributed errors produces standardized intraday returns that come closest to being i.i.d. (as measured by the correlograms of standardized 5-minute returns and squared standardized 5-minute returns). We use these standardized 5-minute returns to obtain the maxima and the minima for each day, and again estimate the time-varying location, scale and shape parameters of the GEV distribution by maximum likelihood. We then study the relative performance of EVT for 1-day and 20-day ahead forecasts of ‘realized volatility’ (here, the sum of squared standardized five-minute returns). Note that we cannot compare the relative performance of VIX with EVT and GED TS-GARCH models in this context because VIX is the implied volatility of S&P 100 index returns, not the volatility of standardized returns. Since we do not expect VIX to predict the realized volatility of standardized returns, we only compare the in-sample and out-of-sample performance of EVT and GED TS-GARCH models.

Table 7 presents the adjusted $R^2$, TIC, HMAE and HRMSE measures. The results indicate that EVT performs much better than the GED TS-GARCH model in forecasting the volatility of standardized returns.

VII. Conclusions

This paper models the conditional distribution of extreme high-frequency index returns, and introduces a conditional extreme value volatility estimator based on the GEV distribution. This is also a first attempt towards detecting any time-series variation in extreme value distributions using high-frequency intraday data. The relative performance of the extreme value volatility estimator is compared with the discrete-time GARCH and implied volatility models for 1-day and 20-day-ahead forecasts of volatility. Overall,
the determination coefficients and the related test statistics imply superior performance of VIX and EVT in capturing time-series variation in volatility. The forecasting ability of discrete-time GARCH models with GED, Student-\(t\), and normal distributions turns out to be inferior to VIX and EVT.

Although the adjusted-\(R^2\) values provide the direction and magnitude of the relationship between the realized and estimated volatilities, they do not measure how far the volatility forecasts are away from the realized volatility. When the summary statistics are calculated based on the deviation between forecasts and realizations, the relative performance of alternative volatility models changes. The results from the Theil Inequality Coefficient (TIC), heteroscedasticity-adjusted mean absolute (HMAE), and root mean squared errors (HRMSE) all highlight the superior performance of the extreme value volatility estimator against the implied and GARCH volatility models.

Numerous interesting questions for future research remain. Of particular interest to regulators and practitioners, the analysis in this paper has important implications for financial risk management. Previous research has identified the unconditional volatility of extreme returns as an important component of VaR calculations. The conditional extreme value approach advocated here can be readily applied to produce dynamic VaR estimates. In other words, one can propose a conditional extreme value approach to estimating daily VaR based on the time-varying location, scale, and shape parameters of the GEV distribution. This is important, because financial institutions calculate and monitor their VaR on a daily basis, but current extreme-value based methods are unconditional and thus cannot produce VaR estimates that explicitly depend upon current economic conditions.
Appendix A. Asymptotic Distribution of Sequences of Dependent Random Variables

Leadbetter et al. (1983), Resnick (1987), and Castillo (1988) derive inequalities and formulas by which is obtained the exact and approximate cdf of all order statistics for finite sample size \( n \). One interesting conclusion is that the value of the cdf of all order statistics at a given point, \( x \), depends only on the \( n \) values. Castillo shows that complete knowledge of the joint cdf of the random variables \((X_1, X_2, \ldots, X_n)\) is not needed in order to derive the statistical behavior of its order statistics, but of only \( n \) functions \([S_i, n(x), i=1, 2, \ldots, n]^{22}\).

**Definition 1 (Exchangeable Variables):** The random variables \( X_1, X_2, \ldots, X_n \) are said to be exchangeable if the distribution of the vector \((X_{i_1}, X_{i_2}, \ldots, X_{i_n})\) is the same for all permutations of the subscripts \((i_1, i_2, \ldots, i_n)\).

Financial time-series (including the daily S&P 100 index returns) are exchangeable because the entire distribution (or all distributional characteristics) is the same for all permutations of the return observations.

**Dependence Conditions:** We include some dependence conditions which play a central role for a dependent sequence to have a similar limit behavior to that of independent sequences.

**Definition 2 (Strong Mixing Sequence):** A sequence \( \{X_n\} \) of random variables is said to satisfy the strong mixing condition if

\[
\mathfrak{I}(j) = |P(A \cap B) - P(A)P(B)|
\]

where \( A \) is any event generated by \((X_1, X_2, \ldots, X_n)\) and \( B \) is any event generated by \((X_{n+j}, X_{n+j+1}, \ldots)\) goes to zero when \( j \to \infty \), and this holds for any value of \( n \).

Equation (1) is difficult to check in application. However, in some particular cases it can be substituted with a simpler one. Chernick (1981a) gives the following theorem and corollary for Markov process of \( p \), which we define previously.

**Definition 3 (Markov Sequence of Order \( p \)):** Let \( \{X_n\} \) be sequence of random variables. This sequence is said to be a Markov sequence of order \( p \), if \((\ldots, X_{m-1}, X_m)\) is independent of \((X_{m+r}, X_{m+r+1}, \ldots)\) given \((X_{m+1}, X_{m+2}, \ldots, X_{m+p})\) for any \( r > p \) and \( m > 0 \).

This definition in practical terms means that the past \((\ldots, X_{m-1}, X_m)\) and the future \((X_{m+r}, X_{m+r+1}, \ldots)\) are independent, given the present \((X_{m+1}, X_{m+2}, \ldots, X_{m+p})\).

---

22 The value of the cdf of all order statistics at a given point, \( x \), depends only on the \( n \) values

\[
P[m_a(x) = 1], P[m_a(x) = 2], \ldots, P[m_a(x) = n]
\]

or, as an equivalence alternative, on the \( n \) values \( S_1, a(x), S_2, a(x), \ldots, S_n, a(x) \), and this is true for the most general case.
**Definition 4 (Condition D):** Let $X_1, X_2, \ldots$ be sequence of random variables, and let $F_{h_1,h_2,\ldots,h_p}(x_1,x_2,\ldots,x_r)$ be a joint cdf of members $X_{i_1}, X_{i_2}, \ldots, X_{i_q}$ in the sequence. The condition $D$ is said to hold if for any set of integers $i_1 < i_2 < \cdots < i_p$ and $j_1 < j_2 < \cdots < j_q$ such that $j_1 - i_p \geq s$, and any real number $u$ we have

$$| F_{h_1,\ldots,h_i \ldots,h_p}(u,\ldots,u) - F_{h_1,\ldots,h_p}(u,\ldots,u)F_{h_i,\ldots,h_q}(u,\ldots,u) | \leq g(s)$$

(2)

with

$$\lim_{s \to \infty} g(s) = 0$$

(3)

Note that the strong mixing condition implies condition $D$. It is enough to take $g(s) = 3(s)$ and

$$A = \{X_{i_1} < u,\ldots,X_{i_p} \leq u\}$$

$$B = \{X_{j_1} < u,\ldots,X_{j_q} \leq u\}$$

(4)

**Definition 5 (Condition $D(u_n)$):** The condition $D(u_n)$ is said to hold if for any integers satisfying the conditions above, we have

$$| F_{h_1,\ldots,h_i \ldots,h_p}(u_n) - F_{h_1,\ldots,h_p}(u_n)F_{h_i,\ldots,h_q}(u_n) | \leq \alpha_{n,s}$$

(5)

where $\alpha_{n,s}$ is non-increasing in $s$ and

$$\lim_{n \to \infty} \alpha_{n,[n\delta]} = 0 \quad \text{for each } \delta > 0$$

(6)

Note that condition $D$ implies condition $D(u_n)$ for any sequence $\{u_n\}$.

**Definition 6 (Condition $D'(u_n)$):** The condition $D'(u_n)$ is said to hold for the stationary sequence $\{X_n\}$ of random variables and the sequence $\{u_n\}$ of constants if

$$\lim \sup_{k \to \infty} n \sum_{j=2}^{n} P[X_1 > u_n, X_j > u_n] = 0$$

(7)

**Limit Distributions of Maxima and Minima:**

1. **Stationary Sequences:** A sequence $X_1, X_2, \ldots$ of random variables is called stationary if

$$F_{i_1,i_2,\ldots,i_q}(x_1,x_2,\ldots,x_k) = F_{i_1+x_1,i_2+x_2,\ldots,i_q+x_k}(x_1,x_2,\ldots,x_k)$$

(8)

for any integers $k$ and $s$. It is clear that exchangeable variables are stationary.

---

23 Chernick (1981b) shows that condition $D(u_n)$ is satisfied for Markov sequences of order one: Let $\{X_n\}$ be a stationary Markov sequence of order one. Let $F(x)$ be the cdf of $X_n$. Then, $D(u_n)$ is satisfied for any sequence $\{u_n\}$ such that $\lim_{n \to \infty} F(u_n) = 1$. 

Sufficient Conditions for i.i.d limit distributions: Let \( \{X_n\} \) be a stationary sequence and let \( \{a_n\} \) and \( \{b_n\} \) be two sequences of real numbers such that
\[
\lim_{n \to \infty} P[X_{n,n} \leq a_n + b_n x] = G(x) \tag{9}
\]
If the sequence \( \{u_n = a_n + b_n x\} \) satisfies the condition \( D(u_n) \) for each \( x \), then \( G(x) \) is one of the three limit distributions (Frechet, Weibull, Gumbel) for the independent case. Because condition \( D \) implies condition \( D(u_n) \), this theorem remains true if condition \( D \) holds.

2. \( m \)-dependent Sequences: A sequence \( X_1, X_2, \ldots \) of random variables is called \( m \)-dependent if the random vectors \( (X_{i}, X_{i+1}, \ldots, X_{i+m-1}) \) and \( (X_{j}, X_{j+1}, \ldots, X_{j+m-1}) \) where
\[
\min(j_1, j_2, \ldots, j_k) - \max(i_1, i_2, \ldots, i_k) \geq m \tag{10}
\]
are independent. Note that \( X_i \) and \( X_j \) can be dependent if they are close \((|i - j| < m)\) but they are independent if they are far apart.

Asymptotic Distribution of the Maxima of \( m \)-dependent Stationary Sequences: Let \( X_1, X_2, \ldots \) be an \( m \)-dependent stationary sequence with common cdf \( F(x) \) such that
\[
\lim_{n \to \infty} n[1 - F(a_n + b_n x)] = u(x); \quad 0 < u(x) < \infty \tag{11}
\]
Then
\[
\lim_{n \to \infty} P[Z_n < a_n + b_n x] = \exp(-u(x)) \tag{12}
\]
If and only if
\[
\lim_{u \to \omega(F)} \frac{P[X_i \geq u, X_{i+1} \geq u]}{1 - F(u)} = 0; \quad 1 \leq i < m \tag{13}
\]
For \( m \)-dependent sequences, equation (13) implies condition \( D'(u_n) \), for any sequence \( \{u_n\} \) such that \( \lim u_n = \omega(F) \) when \( n \to \infty \).

Finite Moving Average Stationary Models: Any MA(q) model, i.e., any model of the form
\[
X_t = \varepsilon_t + c_1 \varepsilon_{t-1} + c_2 \varepsilon_{t-2} + \ldots + c_q \varepsilon_{t-q} \tag{14}
\]
where \( \{\varepsilon_t\} \) is a sequence of i.i.d. random variables is obviously \( m \)-dependent. Any stationary \( m \)-dependent sequence obviously satisfies the \( D \) condition with \( g(s) \) identically zero for \( s \geq m \).\(^{24}\)

\(^{24}\) When we assume an \( m \)-dependent sequence \( \{X_n\} \) of random variables such that the joint function of any \( m \)-consecutive elements, \( \{X_n, X_{n+1}, \ldots, X_{n+m-1}\} \), in the sequence is the Marshall-Olkin (1983) function then the limit distribution of the maximum of \( \{X_n\} \) coincides with the independent case. Different specifications of the Marshall-Olkin (1983) function lead to Weibull or Gumbel distribution.
3. Moving Average Models: In this section we include one theorem which gives the asymptotic distribution of sequences following moving average models [see Leadbetter et al. (1983)]

Asymptotic Distribution of the Maxima for the Moving Average Model of Stable distributions: Suppose the following moving average model

\[ Y_t = \sum_{i=-\infty}^{\infty} C_i X_{t-i}; \quad t \geq 1 \]  

(15)

where \( X_t \) for \( t > 1 \) are independent and stable random variables, i.e., with characteristic function of the form\(^2\)

\[ \varphi(t) = \exp \left[ -\gamma^{\alpha} |t|^{\alpha} \left( 1 - \frac{i\beta h(t, \alpha)}{|t|} \right) \right] \]  

(16)

where

\[ h(t, \alpha) = \begin{cases} \frac{2 \log |t|}{\pi} & \text{if } \alpha = 1 \\ \tan \left( \frac{\pi \alpha}{2} \right) & \text{otherwise} \end{cases} \]  

(17)

and

\[ \gamma \geq 0; \quad 0 < \alpha \leq 2; \quad |\beta| \leq 1. \]  

(18)

Assume also that the constants \( C_i (\infty < i < \infty) \) satisfy

\[ \sum_{i=\infty}^{\infty} |C_i|^{\alpha} < \infty \]  

(19)

\[ \sum_{i=\infty}^{\infty} C_i \log |C_i| < \infty; \quad \text{if } \alpha = 1, \beta \neq 0 \]  

(20)

Then we have the asymptotic distribution of the maxima,

\[ \lim_{n \to \infty} P[X_{n,\alpha} \leq n^{1/\alpha} x] = \begin{cases} \exp \{-K_\alpha [c^+ (1 + \beta) + c^- (1 - \beta)] x^{-\alpha}\} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \]  

(21)

where \( c^+ = \max_{-\infty < i < \infty} \max(0, C_i) \), \( c^- = \max_{-\infty < i < \infty} \max(0, -C_i) \), and \( K_\alpha = \frac{\Gamma(\alpha) \sin(\alpha \pi / 2)}{\pi} \)

\(^2\) Note that any ARMA model is a moving average model of the form (15) where the coefficients \( C_i \) can be obtained by inverting the original ARMA model.

\(^2\) Here \( \gamma \) is a scale parameter, \( \alpha \) is called the index of the distribution, and \( \beta \) is the symmetry parameter. If \( \beta = 0 \) the distribution is symmetric, while if \( |\beta| \leq 1 \) and \( \alpha < 2 \) the distribution is said to be completely asymmetric. For \( \alpha < 1 \), the completely asymmetric stable distributions are concentrated on the positive real line if \( \beta = 1 \), and on the negative real line if \( \beta = -1 \). If \( \alpha = 2 \) the distribution is clearly normal. If \( \gamma = \alpha = 1 \) and \( \beta = 0 \), then \( \varphi(t) \) is the characteristic function of a Cauchy distribution.
Davis and Resnick (1985) study the limit distribution of extremes of moving average models of random variables belonging to a Frechet-type domain of attraction and give conditions under which the extremes belong to a Frechet-type domain of attraction. In particular, Davis and Resnick (1985) derive asymptotic joint distribution of maxima and minima and show that a necessary and sufficient condition for their asymptotic independence is that all coefficients $C_i$ in equation (15) must have the same sign.

4. **Gaussian Sequences:** An important case of stationary sequences is that of Gaussian sequences for which we have the following results [for the proof, the reader is referred to Galambos (1978)]

**Asymptotic Distribution of Stationary Normal Sequences:** If a sequence $X_1, X_2, \ldots$ is a stationary sequence of standard normal $N(0,1)$ random variables with correlation function

$$
\rho_m = E[X_j X_{j+m}]
$$

we have that

(i) If

$$
\lim_{m \to \infty} \rho_m \log m = 0
$$

then

$$
\lim_{n \to \infty} P[Z_n < a_n + b_n x] = H_{3,0}(x)
$$

where $H_{3,0}(x)$ is the Gumbel distribution for maxima.

(ii) If

$$
\lim_{m \to \infty} \rho_m \log m = \tau; \quad 0 < \tau < \infty
$$

then

$$
\lim_{n \to \infty} P[Z_n < a_n + b_n x] = H(x)
$$

where $H(x)$ is the convolution of $H_{3,0}(x + \tau)$ and $\Phi[2(2\tau)^{-1/2}]

(iii) If

$$
\lim_{m \to \infty} \rho_m \log m = \infty
$$

and $\rho_m$ is decreasing

then

$$
\lim_{n \to \infty} P[Z_n < (1 - \rho_n)^{1/2} a_n + x b_n^{1/2}] = \Phi(x)
$$

where $a_n$ and $b_n$ are given by

$$
a_n = \frac{1}{b_n} - \frac{b_n [\log \log n + \log(4\pi) + 2]}{2}
$$

$$
b_n = (2 \log n)^{-1/2}
$$
**Autoregressive Moving Average Models:** Let us assume a sequence \( \{X_n\} \) such that it follows an ARMA\((p,q)\) autoregressive moving average model of orders \( p \) and \( q \) of the form

\[
X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}
\]

where \( \phi_1, \phi_2, \ldots, \phi_p \) and \( \theta_1, \theta_2, \ldots, \theta_q \) are constants and \( \{\varepsilon_t\} \) is a sequence of i.i.d. \( N(0, \sigma^2) \) random variables. Assume that the model above is stationary and invertible, i.e., that the roots of the polynomials

\[
P(x) = x^p - \phi_1 x^{p-1} - \ldots - \phi_p \quad \text{and} \quad Q(x) = x^q - \theta_1 x^{q-1} - \ldots - \theta_q
\]

lie inside the circle.

The autocorrelation function satisfies the difference equation:

\[
\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p} \quad k > q + 1
\]

which has a solution

\[
\rho_k = \sum_{s=1}^{r} \left( m_s \right)^k \left[ \left( \sum_{j=0}^{p-1} c_{ij} k^j \right) \cos(k \alpha_s) + \left( \sum_{j=0}^{q-1} d_{ij} k^j \right) \sin(k \alpha_s) \right]
\]

where \( c_{ij} \) and \( d_{ij} \) are constants, \( m_s \) and \( \alpha_s \) are the modulus and argument, respectively, of the \( r \) different roots of the polynomial \( P(x) \), and \( p_s \) are their degrees of multiplicity, which satisfy the equation

\[
\sum_{s=1}^{r} p_s = p
\]

It is clear that

\[
\lim_{m \to \infty} \rho_m \log m = 0
\]

Then, if \( \sigma^2 \) is chosen such that \( X_t \) is \( N(0,1) \), equations (24)-(25) guarantee that the limit distribution of \( Z_n = \max(X_1, X_2, \ldots, X_n) \) is

\[
\lim_{n \to \infty} P[Z_n < a_n + b_n x] = \exp[-\exp(-x)]
\]

with \( a_n \) and \( b_n \) given by equations (31) and (32), respectively.

Leadbetter et al. (1983) show that for normal sequences that if

\[
\lim_{m \to \infty} \rho_m \log m = 0
\]

and \( n[1-\Phi(\mu_n)] \) is bounded, then conditions \( D(u_n) \) and \( D'(u_n) \) hold. This implies that ARMA\((p,q)\) stationary and invertible models satisfy these conditions for \( u_n = a_n + b_n x \).
Appendix B. Moments of the GEV Distribution

Let \( X \) be a GEV distributed maxima with cdf \( F_X(x) \), where
\[
F_X(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}.
\] (1)

Consider the transformed variable \( Y \) given by
\[
Y = 1 + \xi \left( \frac{X - \mu}{\sigma} \right),
\] (2)
and compute the cdf of \( Y \), \( F_Y(y) \):
\[
F_Y(y) = \Pr(Y < y)
\]
\[
= \Pr \left( 1 + \xi \left( \frac{X - \mu}{\sigma} \right) < y \right)
\]
\[
= \Pr \left( X < \mu + \frac{\sigma}{\xi} (y - 1) \right),
\] (4)
where in (4) we assume that \( \xi > 0 \). Equation (1) then implies that
\[
F_Y(y) = \exp \left\{ -y^{-1/\xi} \right\}.
\] (5)

Now consider the transformation
\[
Z = Y^{-1/\xi},
\] (6)
we have that
\[
\Pr(Z < z) = \Pr(Y^{-1/\xi} < z)
\]
\[
= \Pr(Y > z^{-\xi})
\]
\[
= 1 - \Pr(Y < z^{-\xi})
\]
\[
= 1 - e^{-z}.
\] (7)

So \( Z \) has the exponential distribution with probability density function
\[
f_z(z) = e^{-z}.
\] (8)

Then the \( r \)th moment of \( Y \) about zero is given by
\[
E(Y^r) = E(Z^{-\xi r}) = \int_0^\infty z^{-\xi r} e^{-z} \, dz = \Gamma(1 - \xi r).
\] (9)

It follows from (9) that
\[
E(Y) = \Gamma(1 - \xi),
\] (10)
and
\[
\text{var}(Y) = E(Y^2) - (E(Y))^2 = \Gamma(1 - 2\xi) - [\Gamma(1 - \xi)]^2.
\] (11)

From (2) we have that
\[
X = \mu + \frac{\sigma}{\xi} (Y - 1).
\] (12)

Equations (10) and (12) imply that
\[
E(X) = \mu + \frac{\sigma}{\xi} (E(Y) - 1) = \mu + \frac{\sigma}{\xi} (\Gamma(1 - \xi) - 1),
\] (13)
and equations (11) and (12) imply that
\[
\text{var}(X) = \frac{\sigma^2}{\xi^2} \text{var}(Y) = \frac{\sigma^2}{\xi^2} \{ \Gamma(1 - 2\xi) - [\Gamma(1 - \xi)]^2 \}.
\] (14)
References


Table 1
Descriptive Statistics

This table shows the descriptive statistics for 5-minute and daily logarithmic returns on S&P 100, and for daily realized, implied (VIX), GED TS-GARCH, and extreme value volatility estimators (EVT). The time period of investigation extends from January 5, 1987 to August 31, 2000, giving a total of 3,425 daily and 269,731 5-minute return observations. The skewness and excess kurtosis statistics are reported for testing the distributional assumption of unconditional normality. ADF denotes the Augmented Dickey-Fuller (ADF) unit root statistic with a 1% critical value of -3.4354. \( \rho \) represents the autocorrelation coefficient of order \( j \). *, (**), [***] denote the 1 %, (5 %), [10 %] level of significance.

<table>
<thead>
<tr>
<th></th>
<th>5-min Return</th>
<th>Daily Return</th>
<th>Realized</th>
<th>EVT</th>
<th>VIX</th>
<th>GED TS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>2.2463%</td>
<td>8.5307%</td>
<td>6.7200%</td>
<td>6.4660%</td>
<td>9.4610%</td>
<td>7.4992%</td>
</tr>
<tr>
<td>Minimum</td>
<td>-2.2468%</td>
<td>-23.781%</td>
<td>0.0850%</td>
<td>0.0742%</td>
<td>0.5690%</td>
<td>0.1659%</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0007%</td>
<td>0.0570%</td>
<td>0.7539%</td>
<td>0.9765%</td>
<td>1.2845%</td>
<td>1.0056%</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.0966%</td>
<td>1.1272%</td>
<td>0.4073%</td>
<td>0.4023%</td>
<td>0.5110%</td>
<td>0.4938%</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.1305*</td>
<td>-3.0328*</td>
<td>3.2433*</td>
<td>3.3544*</td>
<td>4.3845*</td>
<td>6.1654*</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>25.967*</td>
<td>66.522*</td>
<td>27.465*</td>
<td>23.654*</td>
<td>50.792*</td>
<td>63.540*</td>
</tr>
<tr>
<td>ADF statistic</td>
<td>-55.102*</td>
<td>-27.864*</td>
<td>-10.735*</td>
<td>-8.7203*</td>
<td>-6.5631*</td>
<td>-7.8404*</td>
</tr>
<tr>
<td>( \rho_1 )</td>
<td>0.194*</td>
<td>-0.019</td>
<td>0.710*</td>
<td>0.935*</td>
<td>0.942*</td>
<td>0.970*</td>
</tr>
<tr>
<td>( \rho_2 )</td>
<td>0.025*</td>
<td>-0.054*</td>
<td>0.635*</td>
<td>0.870*</td>
<td>0.893*</td>
<td>0.936*</td>
</tr>
<tr>
<td>( \rho_3 )</td>
<td>-0.003*</td>
<td>-0.043*</td>
<td>0.589*</td>
<td>0.831*</td>
<td>0.876*</td>
<td>0.900*</td>
</tr>
<tr>
<td>( \rho_4 )</td>
<td>-0.006*</td>
<td>-0.033*</td>
<td>0.553*</td>
<td>0.796*</td>
<td>0.855*</td>
<td>0.865*</td>
</tr>
<tr>
<td>( \rho_5 )</td>
<td>-0.005*</td>
<td>0.023*</td>
<td>0.545*</td>
<td>0.758*</td>
<td>0.834*</td>
<td>0.832*</td>
</tr>
<tr>
<td>( \rho_{10} )</td>
<td>-0.008*</td>
<td>0.008*</td>
<td>0.464*</td>
<td>0.595*</td>
<td>0.728*</td>
<td>0.673*</td>
</tr>
<tr>
<td>( \rho_{21} )</td>
<td>0.001*</td>
<td>-0.012**</td>
<td>0.408*</td>
<td>0.439*</td>
<td>0.637*</td>
<td>0.471*</td>
</tr>
<tr>
<td>( \rho_{63} )</td>
<td>0.007*</td>
<td>0.007*</td>
<td>0.268*</td>
<td>0.182*</td>
<td>0.458*</td>
<td>0.228*</td>
</tr>
<tr>
<td>( \rho_{126} )</td>
<td>-0.005*</td>
<td>-0.005*</td>
<td>0.249*</td>
<td>0.149*</td>
<td>0.385*</td>
<td>0.182*</td>
</tr>
<tr>
<td>( \rho_{252} )</td>
<td>-0.001*</td>
<td>-0.006***</td>
<td>0.182*</td>
<td>0.053*</td>
<td>0.309*</td>
<td>0.096*</td>
</tr>
</tbody>
</table>
Table 2
Maximum Likelihood Estimates of the GEV Distribution

This table presents the maximum likelihood estimates of the time-varying location, scale, and shape parameters of the GEV distribution. The intraday (5-minute) returns data cover the period from January 2, 1987 through August 31, 2000. The 1-day maximal and minimal returns on S&P 100 are obtained from 5-minute returns observed in a day. 3,426 daily extremes are used in maximum likelihood estimation. Asymptotic $t$-statistics are given in parentheses. The maximized likelihood values (Log-L) are reported for each model.

\[
GEV \text{ distribution: } H(M_i; \mu_i, \sigma_i, \xi_i) = \exp \left\{ -1 + \frac{\xi_i}{\xi} \left(\frac{M_i - \mu_i}{\sigma_i}\right)^{-1/\xi_i} \right\}
\]

Location: $\mu_i = \mu + \phi M_{i-1}, \mu_i = \mu$ when $\phi = 0$
Scale: $\sigma_i = \sigma + \lambda_1 \xi_{i-1} + \lambda_2 \sigma_{i-1}, \sigma_i = \sigma$ when $\lambda_1 = \lambda_2 = 0$
Shape: $\xi_i = \xi + \gamma_1 \xi_{i-1} + \gamma_2 \xi_{i-1}, \xi_i = \xi$ when $\gamma_1 = \gamma_2 = 0$

<table>
<thead>
<tr>
<th>Maxima</th>
<th>$\mu$</th>
<th>$\phi$</th>
<th>$\sigma$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\xi$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>Log-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>0.00148</td>
<td>0.2798</td>
<td>0.000018</td>
<td>0.0325</td>
<td>0.9263</td>
<td>0.0049</td>
<td>0.0453</td>
<td>0.8724</td>
<td>18726.68</td>
</tr>
<tr>
<td></td>
<td>(41.568)</td>
<td>(20.938)</td>
<td>(4.2332)</td>
<td>(5.4829)</td>
<td>(82.655)</td>
<td>(2.3106)</td>
<td>(2.4779)</td>
<td>(28.339)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi$</td>
<td>0.00146</td>
<td>0.2834</td>
<td>0.000024</td>
<td>0.0525</td>
<td>0.8873</td>
<td>0.1331</td>
<td>0.0</td>
<td>0.0</td>
<td>18712.44</td>
</tr>
<tr>
<td></td>
<td>(41.148)</td>
<td>(20.918)</td>
<td>(5.5538)</td>
<td>(8.4675)</td>
<td>(73.263)</td>
<td>(8.9569)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi$</td>
<td>0.00151</td>
<td>0.2982</td>
<td>0.00064</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2050</td>
<td>0.0</td>
<td>0.0</td>
<td>18540.12</td>
</tr>
<tr>
<td></td>
<td>(44.495)</td>
<td>(29.113)</td>
<td>(36.557)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(12.692)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi$</td>
<td>-0.00137</td>
<td>0.3044</td>
<td>0.000025</td>
<td>0.0501</td>
<td>0.8937</td>
<td>0.0097</td>
<td>0.0513</td>
<td>0.8428</td>
<td>18565.21</td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi$</td>
<td>-0.00134</td>
<td>0.3133</td>
<td>0.000045</td>
<td>0.1020</td>
<td>0.8541</td>
<td>0.0459</td>
<td>0.0</td>
<td>0.0</td>
<td>18550.20</td>
</tr>
<tr>
<td></td>
<td>(-39.011)</td>
<td>(23.406)</td>
<td>(6.5863)</td>
<td>(10.707)</td>
<td>(61.465)</td>
<td>(4.412)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi$</td>
<td>-0.00224</td>
<td>0.0067</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2224</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>18311.81</td>
</tr>
<tr>
<td></td>
<td>(112.42)</td>
<td>(35.786)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(13.637)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minima</th>
<th>$\mu$</th>
<th>$\phi$</th>
<th>$\sigma$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\xi$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>Log-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>-0.00137</td>
<td>0.3044</td>
<td>0.000053</td>
<td>0.0433</td>
<td>0.8527</td>
<td>0.0021</td>
<td>0.0548</td>
<td>0.9350</td>
<td>18612.21</td>
</tr>
<tr>
<td></td>
<td>(-39.477)</td>
<td>(23.389)</td>
<td>(3.5818)</td>
<td>(5.1826)</td>
<td>(28.951)</td>
<td>(2.2875)</td>
<td>(2.4994)</td>
<td>(45.128)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>-0.00134</td>
<td>0.3133</td>
<td>0.000045</td>
<td>0.0540</td>
<td>0.8492</td>
<td>0.1669</td>
<td>0.0</td>
<td>0.0</td>
<td>18605.82</td>
</tr>
<tr>
<td></td>
<td>(-39.011)</td>
<td>(23.406)</td>
<td>(5.5623)</td>
<td>(7.3657)</td>
<td>(44.101)</td>
<td>(11.372)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>-0.00140</td>
<td>0.3159</td>
<td>0.00065</td>
<td>0.0</td>
<td>0.2102</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>18494.69</td>
</tr>
<tr>
<td></td>
<td>(-46.82)</td>
<td>(30.764)</td>
<td>(36.699)</td>
<td>(0.0)</td>
<td>(13.259)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>-0.00204</td>
<td>0.3165</td>
<td>0.000055</td>
<td>0.0535</td>
<td>0.8465</td>
<td>0.00034</td>
<td>0.0403</td>
<td>0.9425</td>
<td>18416.71</td>
</tr>
<tr>
<td></td>
<td>(-107.96)</td>
<td>(30.764)</td>
<td>(3.9203)</td>
<td>(6.3508)</td>
<td>(33.151)</td>
<td>(2.135)</td>
<td>(2.3601)</td>
<td>(59.774)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>-0.00200</td>
<td>0.0941</td>
<td>0.0941</td>
<td>0.8397</td>
<td>0.0541</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>18320.12</td>
</tr>
<tr>
<td></td>
<td>(-123.56)</td>
<td>(8.5382)</td>
<td>(51.281)</td>
<td>(4.6377)</td>
<td>(51.281)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
<tr>
<td>$\mu_i, \sigma_i, \xi_i$</td>
<td>-0.00217</td>
<td>0.0941</td>
<td>0.0941</td>
<td>0.8397</td>
<td>0.0541</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>18247.45</td>
</tr>
<tr>
<td></td>
<td>(-102.92)</td>
<td>(36.413)</td>
<td>(36.413)</td>
<td>(36.413)</td>
<td>(36.413)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td>(0.0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. In-Sample Performance of Alternative Volatility Models

This table shows the in-sample performance of alternative volatility models. The predictive power of alternative models is evaluated by estimating an OLS regression:

$$ Realized_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t-1} + \omega_3 GARCH_{t-1} + u_t, $$

where $Realized_t$ is the square-root of the sum of squared 5-minute returns on S&P 100 index at time $t$, $VIX_t$ is the daily implied volatility index at time $t$, $EVT_{t-1}$ is the daily extreme value volatility estimator at time $t$, $GARCH_{t-1}$ is the daily volatility of index returns at time $t$, and $u_t$ is the forecast error. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by $VIX_t$, $EVT_{t-1}$, and $GARCH_{t-1}$. The Newey-West (1987) adjusted $t$-statistics of the regression coefficients are given in parentheses. *, (**), [***] denote statistical significance at least at the 10% (5%) [1%] level. Theil Inequality Coefficient (TIC) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) measure how far the estimated volatilities are away from the realized (or integrated) volatility.

<table>
<thead>
<tr>
<th>In-sample forecast</th>
<th>$\omega_0$</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VIX_t$</td>
<td>0.00041</td>
<td>0.5551</td>
<td>0.0</td>
<td>0.0</td>
<td>0.484690</td>
<td>28.20%</td>
<td>42.52%</td>
<td>45.35%</td>
</tr>
<tr>
<td></td>
<td>(0.8753)</td>
<td>(13.978)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$EVT_t$</td>
<td>0.00104</td>
<td>0.6673</td>
<td>0.0</td>
<td>0.0</td>
<td>0.417374</td>
<td>19.21%</td>
<td>28.36%</td>
<td>30.90%</td>
</tr>
<tr>
<td></td>
<td>(2.4930)**</td>
<td>(14.338)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GARCH_t$</td>
<td>0.00218</td>
<td>0.5332</td>
<td>0.0</td>
<td>0.0</td>
<td>0.401723</td>
<td>20.67%</td>
<td>30.67%</td>
<td>35.14%</td>
</tr>
<tr>
<td></td>
<td>(3.0893)**</td>
<td>(7.1531)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, EVT_t$</td>
<td>-0.000029</td>
<td>0.4007</td>
<td>0.2747</td>
<td>0.0</td>
<td>0.515158</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-0.6283)</td>
<td>(10.279)**</td>
<td>(7.0110)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, GARCH_t$</td>
<td>0.00051</td>
<td>0.4646</td>
<td>0.0</td>
<td>0.1052</td>
<td>0.487925</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(1.0595)</td>
<td>(9.2126)**</td>
<td></td>
<td>(1.2938)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, EVT_t, GARCH_t$</td>
<td>-0.00033</td>
<td>0.4168</td>
<td>0.2830</td>
<td>-0.0242</td>
<td>0.515167</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-0.8839)</td>
<td>(10.089)**</td>
<td>(5.9699)**</td>
<td>(-0.3279)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Out-of-Sample Performance of Alternative Volatility Models

This table shows the empirical performance of alternative volatility models based on their 1-day- and 20-day-ahead forecasts of realized volatility. The predictive power of alternative models is evaluated by estimating an OLS regression: \( \text{Realized}_{t,j} = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_t + \omega_3 GARCH_t + u_t \), where \( j = 1 \) or 20, \( \text{Realized}_{t,j} \) represents realized standard deviation, \( VIX_t \) is the daily implied volatility index at time \( t \), \( EVT_t \) is the daily extreme value volatility estimator at time \( t \), \( GARCH_t \) is the daily volatility of index returns at time \( t \), and \( u_t \) is the forecast error. Adjusted-\( R^2 \) values indicate the proportion of the total variation in daily realized volatility that can be explained by \( VIX_t \), \( EVT_t \), and \( GARCH_t \). It provides information about how well each model is able to forecast the future volatility of stock market returns. The Newey-West (1987) adjusted \( t \)-statistics of the regression coefficients are given in parentheses. *, **, *** denote statistical significance at least at the 10% (5%) [1%] level. Theil Inequality Coefficient (TIC) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) measure how far the estimated volatilities are away from the realized (or integrated) volatility.

<table>
<thead>
<tr>
<th>1-day-ahead forecast</th>
<th>( \omega_0 )</th>
<th>( \omega_1 )</th>
<th>( \omega_2 )</th>
<th>( \omega_3 )</th>
<th>Adjusted-( R^2 )</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( VIX_t )</td>
<td>-0.00143</td>
<td>0.7493</td>
<td>0.0</td>
<td>0.0</td>
<td>0.566899</td>
<td>24.83%</td>
<td>39.64%</td>
<td>42.54%</td>
</tr>
<tr>
<td></td>
<td>(-3.7283)**</td>
<td>(13.978)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( EVT_t )</td>
<td>-0.00069</td>
<td>0.0</td>
<td>0.8662</td>
<td>0.0</td>
<td>0.498706</td>
<td>17.49%</td>
<td>27.31%</td>
<td>31.41%</td>
</tr>
<tr>
<td></td>
<td>(-1.5446)</td>
<td></td>
<td>(16.381)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( GARCH_t )</td>
<td>0.00110</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6653</td>
<td>0.404105</td>
<td>19.54%</td>
<td>29.96%</td>
<td>34.42%</td>
</tr>
<tr>
<td></td>
<td>(2.4735)**</td>
<td></td>
<td></td>
<td>(13.388)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( VIX_t, EVT_t )</td>
<td>-0.00174</td>
<td>0.6076</td>
<td>0.2084</td>
<td>0.0</td>
<td>0.586463</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-4.6896)**</td>
<td>(18.394)**</td>
<td>(5.4951)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( VIX_t, GARCH_t )</td>
<td>-0.00152</td>
<td>0.7094</td>
<td>0.0</td>
<td>0.0594</td>
<td>0.567462</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-3.8818)**</td>
<td>(11.389)**</td>
<td></td>
<td>(0.8370)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( VIX_t, EVT_t, GARCH_t )</td>
<td>-0.00172</td>
<td>0.6205</td>
<td>0.2181</td>
<td>-0.0289</td>
<td>0.586510</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-4.8345)**</td>
<td>(12.515)**</td>
<td>(4.4956)**</td>
<td>(-0.4028)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20-day-ahead forecast</th>
<th>( \omega_0 )</th>
<th>( \omega_1 )</th>
<th>( \omega_2 )</th>
<th>( \omega_3 )</th>
<th>Adjusted-( R^2 )</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( VIX_t )</td>
<td>-0.00037</td>
<td>0.6596</td>
<td>0.0</td>
<td>0.0</td>
<td>0.439539</td>
<td>32.53%</td>
<td>43.61%</td>
<td>47.80%</td>
</tr>
<tr>
<td></td>
<td>(-0.9284)</td>
<td>(16.918)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( EVT_t )</td>
<td>0.00049</td>
<td>0.0</td>
<td>0.7414</td>
<td>0.0</td>
<td>0.365477</td>
<td>24.44%</td>
<td>32.52%</td>
<td>39.20%</td>
</tr>
<tr>
<td></td>
<td>(0.9125)</td>
<td></td>
<td>(12.007)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( GARCH_t )</td>
<td>0.00207</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5640</td>
<td>0.289981</td>
<td>28.76%</td>
<td>34.55%</td>
<td>41.46%</td>
</tr>
<tr>
<td></td>
<td>(4.4660)**</td>
<td></td>
<td></td>
<td>(10.836)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( VIX_t, EVT_t )</td>
<td>-0.00063</td>
<td>0.5420</td>
<td>0.1732</td>
<td>0.0</td>
<td>0.453195</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-1.5493)</td>
<td>(13.499)**</td>
<td>(3.8397)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( VIX_t, GARCH_t )</td>
<td>-0.00054</td>
<td>0.5913</td>
<td>0.0</td>
<td>0.1039</td>
<td>0.440768</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-1.1917)</td>
<td>(10.209)**</td>
<td></td>
<td>(1.3278)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( VIX_t, EVT_t, GARCH_t )</td>
<td>-0.00068</td>
<td>0.5107</td>
<td>0.1523</td>
<td>0.0661</td>
<td>0.453209</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(-1.6506)</td>
<td>(8.3707)**</td>
<td>(2.6353)**</td>
<td>(0.6744)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Empirical Performance of ARMA(p,q) Realized Volatility Model

This table shows the in-sample and out-of-sample performance of ARMA(p,q) realized volatility model. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by ARMA(p,q) model. Theil Inequality Coefficient ($TIC$) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error ($HMAE$) and root mean squared error ($HRMSE$) measure how far the estimated volatilities are away from the realized volatility.

### Panel A. In-Sample Performance of ARMA(p,q) Model

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(1,1)</td>
<td>47.11%</td>
<td>17.80%</td>
<td>24.44%</td>
<td>35.03%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>48.57%</td>
<td>17.54%</td>
<td>24.12%</td>
<td>34.80%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>48.48%</td>
<td>17.55%</td>
<td>24.16%</td>
<td>34.88%</td>
</tr>
<tr>
<td>ARMA(4,4)</td>
<td>48.41%</td>
<td>17.57%</td>
<td>24.19%</td>
<td>35.00%</td>
</tr>
<tr>
<td>ARMA(5,5)</td>
<td>48.54%</td>
<td>17.54%</td>
<td>24.13%</td>
<td>34.84%</td>
</tr>
</tbody>
</table>

### Panel B. Out-of-Sample Performance of ARMA(p,q) Model (1-day-ahead Forecasting)

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(1,1)</td>
<td>43.93%</td>
<td>18.66%</td>
<td>25.71%</td>
<td>38.60%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>45.52%</td>
<td>18.38%</td>
<td>25.31%</td>
<td>38.41%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>45.49%</td>
<td>18.39%</td>
<td>25.32%</td>
<td>38.43%</td>
</tr>
<tr>
<td>ARMA(4,4)</td>
<td>45.43%</td>
<td>18.40%</td>
<td>25.35%</td>
<td>38.46%</td>
</tr>
<tr>
<td>ARMA(5,5)</td>
<td>45.52%</td>
<td>18.39%</td>
<td>25.29%</td>
<td>38.40%</td>
</tr>
</tbody>
</table>

### Panel C. Out-of-Sample Performance of ARMA(p,q) Model (20-day-ahead Forecasting)

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(1,1)</td>
<td>22.56%</td>
<td>22.48%</td>
<td>30.68%</td>
<td>45.39%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>23.94%</td>
<td>22.23%</td>
<td>30.48%</td>
<td>45.32%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>23.89%</td>
<td>22.24%</td>
<td>30.54%</td>
<td>45.42%</td>
</tr>
<tr>
<td>ARMA(4,4)</td>
<td>23.85%</td>
<td>22.26%</td>
<td>30.55%</td>
<td>45.54%</td>
</tr>
<tr>
<td>ARMA(5,5)</td>
<td>23.85%</td>
<td>22.26%</td>
<td>30.57%</td>
<td>45.61%</td>
</tr>
</tbody>
</table>
Table 6. Empirical Performance of Alternative Volatility Models  
(Forecasting Traditional Measure of Realized Volatility)

This table shows the in-sample and out-of-sample performance of alternative volatility models in forecasting the traditional measure of realized volatility. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by alternative volatility models. Theil Inequality Coefficient ($TIC$) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error ($HMAE$) and root mean squared error ($HRMSE$) measure how far the estimated volatilities are away from the realized volatility.

### Panel A. In-Sample Performance of Alternative Volatility Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVT</td>
<td>18.01%</td>
<td>37.36%</td>
<td>56.14%</td>
<td>67.84%</td>
</tr>
<tr>
<td>VIX</td>
<td>17.18%</td>
<td>39.12%</td>
<td>57.92%</td>
<td>70.10%</td>
</tr>
<tr>
<td>GED TS-GARCH</td>
<td>14.81%</td>
<td>40.77%</td>
<td>58.81%</td>
<td>73.22%</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>13.01%</td>
<td>43.92%</td>
<td>67.84%</td>
<td>91.89%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>13.24%</td>
<td>43.72%</td>
<td>67.62%</td>
<td>91.24%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>13.71%</td>
<td>42.86%</td>
<td>67.41%</td>
<td>91.12%</td>
</tr>
<tr>
<td>ARMA(4,4)</td>
<td>13.96%</td>
<td>42.32%</td>
<td>67.02%</td>
<td>89.03%</td>
</tr>
<tr>
<td>ARMA(5,5)</td>
<td>13.81%</td>
<td>42.94%</td>
<td>67.22%</td>
<td>89.52%</td>
</tr>
</tbody>
</table>

### Panel B. Out-of-Sample Performance of Alternative Volatility Models (1-day-ahead Forecasting)

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVT</td>
<td>17.89%</td>
<td>38.38%</td>
<td>57.20%</td>
<td>68.21%</td>
</tr>
<tr>
<td>VIX</td>
<td>16.62%</td>
<td>40.10%</td>
<td>58.82%</td>
<td>71.56%</td>
</tr>
<tr>
<td>GED TS-GARCH</td>
<td>13.20%</td>
<td>41.19%</td>
<td>60.17%</td>
<td>75.05%</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>11.04%</td>
<td>45.99%</td>
<td>70.18%</td>
<td>93.66%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>11.29%</td>
<td>45.91%</td>
<td>69.86%</td>
<td>93.50%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>11.79%</td>
<td>45.36%</td>
<td>69.45%</td>
<td>93.61%</td>
</tr>
<tr>
<td>ARMA(4,4)</td>
<td>12.01%</td>
<td>44.71%</td>
<td>69.00%</td>
<td>92.00%</td>
</tr>
<tr>
<td>ARMA(5,5)</td>
<td>11.92%</td>
<td>44.88%</td>
<td>69.23%</td>
<td>92.43%</td>
</tr>
</tbody>
</table>

### Panel C. Out-of-Sample Performance of Alternative Volatility Models (20-day-ahead Forecasting)

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVT</td>
<td>7.44%</td>
<td>41.41%</td>
<td>59.87%</td>
<td>74.49%</td>
</tr>
<tr>
<td>VIX</td>
<td>6.16%</td>
<td>43.02%</td>
<td>61.58%</td>
<td>79.28%</td>
</tr>
<tr>
<td>GED TS-GARCH</td>
<td>5.33%</td>
<td>44.58%</td>
<td>64.07%</td>
<td>86.19%</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>3.53%</td>
<td>47.90%</td>
<td>73.07%</td>
<td>121.21%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>3.82%</td>
<td>47.54%</td>
<td>72.84%</td>
<td>119.80%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>4.01%</td>
<td>47.35%</td>
<td>72.71%</td>
<td>118.88%</td>
</tr>
<tr>
<td>ARMA(4,4)</td>
<td>4.97%</td>
<td>46.80%</td>
<td>72.09%</td>
<td>116.37%</td>
</tr>
<tr>
<td>ARMA(5,5)</td>
<td>4.51%</td>
<td>47.11%</td>
<td>72.43%</td>
<td>116.95%</td>
</tr>
</tbody>
</table>
This table shows the in-sample and out-of-sample performance of EVT and GED TS-GARCH models. The realized volatility is the sum of squared standardized 5-minute returns. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by EVT and GED TS-GARCH models. Theil Inequality Coefficient ($TIC$) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error ($HMAE$) and root mean squared error ($HRMSE$) measure how far the estimated volatilities are away from the realized volatility.

### Panel A. In-Sample Performance of EVT and GED TS-GARCH

<table>
<thead>
<tr>
<th>Models</th>
<th>Adjusted-$R^2$</th>
<th>$TIC$</th>
<th>$HMAE$</th>
<th>$HRMSE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVT</td>
<td>26.06%</td>
<td>16.62%</td>
<td>18.28%</td>
<td>11.58%</td>
</tr>
<tr>
<td>GED TS-GARCH</td>
<td>10.23%</td>
<td>22.85%</td>
<td>24.32%</td>
<td>15.85%</td>
</tr>
</tbody>
</table>

### Panel B. Out-of-Sample Performance of EVT and GED TS-GARCH (1-day-ahead Forecasting)

<table>
<thead>
<tr>
<th>Models</th>
<th>Adjusted-$R^2$</th>
<th>$TIC$</th>
<th>$HMAE$</th>
<th>$HRMSE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVT</td>
<td>23.30%</td>
<td>17.66%</td>
<td>18.79%</td>
<td>11.81%</td>
</tr>
<tr>
<td>GED TS-GARCH</td>
<td>7.95%</td>
<td>23.14%</td>
<td>26.52%</td>
<td>18.41%</td>
</tr>
</tbody>
</table>

### Panel C. Out-of-Sample Performance of EVT and GED TS-GARCH (20-day-ahead Forecasting)

<table>
<thead>
<tr>
<th>Models</th>
<th>Adjusted-$R^2$</th>
<th>$TIC$</th>
<th>$HMAE$</th>
<th>$HRMSE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVT</td>
<td>17.69%</td>
<td>18.28%</td>
<td>19.33%</td>
<td>11.92%</td>
</tr>
<tr>
<td>GED TS-GARCH</td>
<td>5.39%</td>
<td>25.61%</td>
<td>28.63%</td>
<td>20.40%</td>
</tr>
</tbody>
</table>
Figure 1. Realized Volatility versus Implied Volatility (VIX)

Figure 2. Realized Volatility versus Extreme Value Volatility Estimator (EVT)

Figure 3. Realized Volatility versus GED TS-GARCH