Volatility Regressions with Fat Tails*

Jihyun Kim Nour Meddahi

Toulouse School of Economics

February 6, 2020

Abstract

Nowadays, a common method to forecast integrated variance is to use the fitted value of a simple OLS autoregression of the realized variance. However, non-parametric estimates of the tail index of this realized variance process reveal that its second moment is possibly unbounded. In this case, the behavior of the OLS estimators and the corresponding statistics are unclear. We prove that when the second moment of the spot variance is unbounded, the slope of the spot variance's autoregression converges to a random variable as the sample size diverges. The same result holds when one uses the integrated or realized variance instead of the spot variance. We then consider the class of diffusion variance models with an affine drift, a class which includes GARCH and CEV processes, and we prove that IV estimation with adequate instruments provide consistent estimators of the drift parameters as long as the variance process has a finite first moment regardless of the existence of the second moment. In particular, for the GARCH diffusion model with fat tails, an IV estimation where the instrument equals the sign of the centered lagged value of the variable of interest provides consistent estimators. Simulation results corroborate the theoretical findings of the paper.

Key words: volatility; autoregression; fat tails; random limits.

*We thank the Guest Editor and three anonymous referees for many helpful comments. For useful discussions, we are also grateful to Tim Bollerslev, Jean-Pierre Florens, Arnaud Gloter, Christian Gouriéroux, Jean Jacod, Joon Y. Park, Andrew Patton, Anders Rahbek, Eric Renault, and the participants at 2017 Toulouse Financial Econometrics Conference, 2017 EC^2, 2017 Korean-American Economic Association Conference, 2017 Korean Econometrics Study Group, 2017 French Econometrics Conference, and 2018 SoFiE Conference. Both authors are grateful to the French Government and the ANR for support under the Investissements d'Avenir program, grant ANR-17- EURE-0010. The second author also thanks financial support from the Chair "Marché des risques et création de valeur" Fondation du risque/SCOR.

1. Introduction

In this paper, we are interested in using high-frequency data based measures to forecast future variance. A common practice is to approximate the latent daily integrated variance by realized measures like realized variance (Andersen et al. (2001)) or robust-to-noise measures (Zhang et al. (2005); Barndorff-Nielsen et al. (2008); Jacod et al. (2009)), and then to estimate a simple autoregressive regression of these realized measures by OLS to get a forecast of the integrated variance. This autoregressive regression is often misspecified because the dynamics of an integrated variance is more complex than a simple autoregressive process. For instance, if the true instantaneous (or spot) variance is a square-root process, then the integrated and realized variances are ARMA (1,1) processes (Barndorff-Nielsen and Shephard (2002); Meddahi (2003)). Even if the autoregression model is misspecified, it still provides a very accurate forecast because integrated variance as well as high-frequency realized measures are persistent and therefore, few lags suffice to predict well future volatility (Andersen et al. (2003); Andersen et al. (2004)).

On the other hand, the GARCH approach (Engle (1982); Bollerslev (1986)) based on parametric models of daily data provides very useful information about the variance process. One of them, which is a primary interest in this paper, is fat tails. When one estimates a daily GARCH model on stock returns or exchange rates, one often finds that the returns' fourth moment is not bounded or close to being unbounded. If the fourth moment of the returns is unbounded, then the second moment of the daily realized variance defined as the sum of intra-daily squared returns is also unbounded. Consequently, the interpretation of the autoregressive regression and the OLS estimation, based on L^2 projections, is questionable. Likewise, relying on Gaussian limit theory, the delivered forecast and all the statistical tools, used to assess the quality of the forecast, could be invalid.

The doubt about the finiteness of the returns' fourth moment is based on a parametric model of the volatility. In contrast, an important contribution of the high-frequency volatility literature is that the availability of a lot of data allows one to get non-parametric measures of the variance without relying on parametric models for the volatility. It is therefore necessary to assess the finiteness of the second moment of realized measures in a non-parametric way. The solution hinges on a non-parametric estimation of tail indexes. We use Hill's (1975) estimator for our data, and we get the same result. More precisely, we find that the Hill estimator of the tail index of the daily returns is close to four, while it is close to two for the daily realized variance, and two other popular measures which are robust in the presence of jumps, namely the bipower variance of Barndorff-Nielsen and Shephard (2004b, 2006) and the threshold variance estimator of Jacod (2008, 2012) and Mancini (2009).

In this paper, we revisit the results of the autoregressive regression of the variance process like Andersen et al. (2004) when the second moment of the spot variance is possibly unbounded, which implies that the second moment of integrated and realized variances are unbounded.[1] When the instantaneous variance has an unbounded second moment, then the results in Andersen et al. (2004) are no longer valid because one cannot compute the population autoregression parameters.

We study empirical regressions instead of population regressions. More precisely, we analyze the asymptotic behavior of the OLS estimator of the autoregressions. We consider autoregressions of three variables: the spot variance, the integrated variance and the realized variance. Of course, the first two autoregressions are not feasible because the variables are not observed, but still the two autoregressions provide good benchmarks. In particular, the third autoregression will try to mimic the second one.

The asymptotic behavior of an OLS estimator under fat tails is ambiguous; it could be deterministic or random. For instance, when one considers an autoregressive process of order one with i.i.d. errors and unbounded variance, the OLS estimator is consistent whether the autoregressive parameter is smaller than one (Hannan and Kanter (1977); Knight (1987)) or equals one (Chan and Tran (1989); Phillips (1990)). Likewise, when one considers an infinite MA process with i.i.d. errors and unbounded variance and assumes some summability restrictions, the empirical autocorrelation parameters converge to deterministic values; see (Brockwell and Davis, 1991, p. 538). In contrast, when one considers an ARCH or GARCH process, the empirical autocorrelation parameters of the squared process converge to random variables when the fourth moment of the process is unbounded (Davis and Mikosch (1998); Mikosch and Starica (2000)).

[1]When one considers a continuous time model without jumps and without market microstructure noise, the fourth moment of the intra-day returns is unbounded if and only if the second moment of the instantaneous variance is infinite.

When we study the autoregressions, we consider two types of asymptotic approaches. Two time-dimension parameters will play a role in these asymptotic analyses: Δ which is the length of sub-periods and the time span denoted T. In the first asymptotic approach, we assume that $\Delta \to 0$ while T is fixed or diverges to ∞. We consider this type of asymptotic approach because we want to characterize the behavior of the OLS estimators without making a parametric model assumption as in Andersen et al. (2004). In the second asymptotic approach, we keep Δ fixed and allow $T \to \infty$ at the cost of making a parametric assumption of the variance diffusion process.

In the first asymptotic analysis with $\Delta \to 0$, we characterize the behavior of the OLS estimators of the three regressions' slopes. When the spot variance process has a bounded second moment, we prove that the OLS estimators converge to finite quantities, which are the same ones as the population parameters derived in Andersen et al. (2004). In contrast, when the spot variance has an unbounded second moment, we prove that the OLS estimators converge to random variables. Both the simulations and the comparison with the results in Andersen et al. (2004), when the spot variance has a finite second moment, corroborate the good quality of our approach.

These results are obviously negative. Providing positive results in a general context is not easy because one needs to specify the object of interest. We therefore consider a class of variance models based on diffusion processes having an affine form of drift, where the class includes GARCH and CEV processes, with possibly unbounded second moment. For this semiparametric class of models, we follow the literature on regressions with fat tails like Blattberg and Sargent (1971) and Samorodnitsky et al. (2007) by considering instrumental variable (IV) estimations. We prove that the IV estimators become consistent estimators of the drift parameters when instruments are chosen appropriately.

Samorodnitsky et al. (2007) studied the estimation of linear regression models where the explanatory and the noise variables have fat tails. It considered estimators that have an instrumental variable interpretation where the instrument is a signed power of the explanatory variable, with the OLS being a particular case. The choice of the power is selected for either efficiency purposes or for getting an estimator with a normal asymptotic distribution, which is often not the case of the OLS estimator when it is consistent. However, in this paper, we select the instruments for consistency purposes of the drift parameters. The asymptotic distribution of the estimator as well as the efficiency question are not studied and left for future research.

When Δ is fixed, unlike the asymptotics with $\Delta \to 0$, we need a conditional moment restriction for the asymptotics of IV estimators. It is well known that for a stationary diffusion with affine drift, the conditional mean is affine as long as the diffusion has a bounded second moment (see Meddahi and Renault (2004)). We prove that for a GARCH diffusion, the result is still valid when the second moment is unbounded. We then show that the IV estimation with adequate instruments leads to consistent estimators of the drift parameters, and we provide a feasible inference under additional restrictions. A particular instrument we study is the sign of the lagged value of the centered spot variance, corresponding to power zero of the signed power instrument mentioned above. This estimator was first proposed by Cauchy (1836), and is often referred to as the "Cauchy estimator" (see, e.g., So and Shin (1999); Phillips et al. (2004); Choi et al. (2016) for recent uses of the Cauchy estimator).

Interestingly, Jean-Marie Dufour used the sign-based methods in several studies for inference purposes, especially for exact inference in a finite sample. In particular, he used such an approach in Coudin and Dufour (2009, 2017) in order to provide inferences about the slope parameter in a linear regression model without making moment restrictions on the disturbance errors and therefore allowing for fat tails. The assumption made in these papers is a median restriction on the errors conditional on the explanatory variables. In other words, we are using the same approach with a slightly different framework because we assume that the (conditional) first moment of the errors exists and equals zero, but we do not make assumptions on higher moments.

The paper is organized as follows. The next section provides the setup, an empirical motivation for fat tails, and various regressions. In Section 3, we analyze the asymptotic behavior of the OLS estimators when $\Delta \to 0$. Section 4 studies the IV estimation. In Section 5, we analyze the volatility regressions under a fixed Δ, and provide a feasible inference based on instrumental variables. Section 6 provides simulations to assess the finite sample properties of the estimators, while the last section concludes. All the proofs are provided in an Online Appendix.

Throughout the paper we use "$P_T \sim Q_T$" to denote $P_T = Q_T(1+o(1))$. Similarly, "$P_T \sim_p Q_T$" and "$P_T \sim_d Q_T$" mean $P_T = Q_T(1+o_p(1))$ and $P_T =_d Q_T(1+o_p(1))$, respectively. These notations, as well as other standard notations used in asymptotics, will be used frequently throughout the paper without further references.

2. Model and Preliminaries

2.1. Spot, Integrated and Realized Variances

We consider a price process $(P_t, 0 \leq t \leq T)$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Our basic assumption is that P_t is a Brownian semimartingale with the following form:

$$d\log(P_t) = D_t dt + V_t^{1/2} dW_t^P,$$

where W_t^P is a Brownian motion, D_t and V_t are adapted processes with càdlàg paths. For a Δ-interval, we define the spot variance (v_i), integrated variance (x_i) and realized variance (y_i) of the price process (P_t) as

$$v_i = V_{i\Delta}, \qquad x_i = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} V_t dt, \qquad y_i = \frac{1}{\Delta}\sum_{j=1}^{n}\left(r^{(\delta)}_{(i-1)\Delta+j\delta}\right)^2, \tag{2.1}$$

for $i = 1, \cdots, N$ with $N\Delta = T$, where $r^{(\delta)}$ is the δ-period return defined as $r^{(\delta)}_{(i-1)\Delta+j\delta} = \log(P_{(i-1)\Delta+j\delta}) - \log(P_{(i-1)\Delta+(j-1)\delta})$ for $j = 1, \cdots, n$ with $n\delta = \Delta$. It is well known that the realized variance y is a noisy measure of the integrated variance x, and satisfies

$$(n/2)^{1/2}(y_i - x_i) \to_d \eta_i \mathbb{N}(0,1), \tag{2.2}$$

where $\eta_i^2 = \Delta^{-1}\int_{(i-1)\Delta}^{i\Delta} V_t^2 dt$, as $n \to \infty$ for fixed Δ and for each $i = 1, \cdots, N$. See, e.g., Barndorff-Nielsen and Shephard (2004a). Moreover, the convergence (2.2) holds jointly for $i = 1, \cdots, N$ if $T = N\Delta$ is fixed (see, e.g., Jacod and Protter (1998)).

In this paper, we analyze the asymptotic properties of various estimators for the volatility regression. Specifically, we consider the autoregression

$$z_{i+1} = \alpha_z + \beta_z^{(k)} z_{i-k} + u_{i+1} \quad \text{with} \quad k \geq 0 \tag{2.3}$$

for $z = v, x, y$, and estimate the slope coefficient $\beta_z^{(k)}$ using OLS or IV method. Our asymptotics for $z = v, x$ involve two parameters, the sampling interval Δ and the time span T, and it is developed under the assumption that $\Delta \to 0$ and $T \to$

∞ simultaneously. On the other hand, the asymptotics for $z = y$ involve three parameters, the sampling interval Δ at low-frequency, the sampling interval δ at high-frequency, and the time span T. In this case, the asymptotics are developed under the assumption that $\delta/\Delta \to 0$, $\Delta \to 0$ and $T \to \infty$ simultaneously.

To effectively analyze the large T asymptotics, we assume that the underlying variance process V is a diffusion process on $\mathcal{D} = (\underline{v}, \overline{v}) \subset \mathbb{R}$ driven by

$$dV_t = \mu(V_t)dt + \sigma(V_t)dW_t, \tag{2.4}$$

where W is a Brownian motion, and μ and σ are respectively drift and diffusion functions of V. To obtain more explicit asymptotic results, we mainly consider a pure diffusion V without having leverage effects, i.e., each of V and D is independent of W^P, unless we mention that they are dependent. We believe that the implications of our results under no leverage effect remain valid for the model with leverage effects.

We let s be the scale function defined as

$$s(v) = \int_y^v \exp\left(-\int_y^x \frac{2\mu(z)}{\sigma^2(z)}dz\right)dx, \tag{2.5}$$

where the lower limits of the integrals can be arbitrarily chosen to be any point $y \in \mathcal{D}$. Defined as such, the scale function s is uniquely identified up to any increasing affine transformation, i.e., if s is a scale function, then so is $as + b$ for any constants $a > 0$ and $-\infty < b < \infty$. We also define the speed density

$$m(v) = \frac{1}{(\sigma^2 s')(v)} \tag{2.6}$$

on $\mathcal{D}$, where s' is the derivative of s, often called the scale density, which is assumed to exist. The speed density is defined to be the measure on $\mathcal{D}$ given by the speed density with respect to the Lebesgue measure.

Throughout this paper, we assume

Assumption 2.1. *(a) $\sigma^2(v) > 0$ for all $v \in \mathcal{D}$, and (b) $\mu(v)/\sigma^2(v)$ and $1/\sigma^2(v)$ are locally integrable at every $v \in \mathcal{D}$.*

Assumption 2.1 provides a simple sufficient set of conditions to ensure that a weak solution to the stochastic differential equation (2.4) exists uniquely up to an explosion time. See, e.g., Theorem 5.5.15 in Karatzas and Shreve (1991). Note, under Assumption 2.1, that both the scale function s and speed density m are well defined, and that the scale function is strictly increasing, on $\mathcal{D}$. Consequently, the natural scale diffusion V^s of V, where $V^s = s(V)$, is well defined with speed density $m_s = (m/s') \circ s^{-1}$. It follows immedaitely from Ito's lemma that the natural scale diffusion V^s has no drift term. Following Kim and Park (2017), we use the natural scale representation in the development of our long span asymptotics.

2.2. Population Regressions with GARCH Diffusions

In this section, we study the volatility regressions in population when $\mathbb{E}(V_t^2) < \infty$ as did Andersen et al. (2004). These authors considered the Eigenfunction Stochastic Volatility (ESV) model of Meddahi (2001) to derive analytical forecast results. Examples of ESV include the square-root model, the log-normal stochastic volatility model and the GARCH diffusion model. We focus here on the GARCH diffusion model of Nelson (1990) because it allows for unbounded second moments while the two other examples do not. More precisely, we assume that the spot variance V_t, defined on $(0, \infty)$, is given by

$$dV_t = \kappa(\mu - V_t)dt + \sigma V_t dW_t. \tag{2.7}$$

When V_t is assumed to be stationary, one can easily prove that the second moment of V_t is bounded if and only if $\sigma^2 < 2\kappa$.

2.2.1 GARCH Diffusions with $\mathbb{E}(V_t^2) < \infty$

Andersen et al. (2004) computed the population values of the autocovariances of spot (v), integrated (x) and realized variances (y) under $\mathbb{E}(V_t^2) < \infty$. From these quantities, one gets the corresponding autoregressive coefficients β_v, β_x and β_y. In particular, one has

$$\beta_v = \exp(-\kappa\Delta), \quad \beta_x = \frac{1}{2}\frac{(1-\exp(-\kappa\Delta))^2}{\exp(-\kappa\Delta)+\kappa\Delta-1}, \quad \beta_y = \frac{a_1^2}{\Delta^2\kappa^2}\frac{(1-\exp(-\kappa\Delta))^2}{Var(y)},$$

where

$$Var(y) = 2\frac{a_1^2}{\Delta^2\kappa^2}(\exp(-\kappa\Delta) + \kappa\Delta - 1) + \frac{4}{\delta\Delta}\left(\frac{a_0^2\delta^2}{2} + \frac{a_1^2}{\kappa^2}(\exp(-\kappa\delta) + \kappa\delta - 1)\right),$$

with $a_0 = \mathbb{E}(V_t) = \mu$ and $a_1^2 = Var(V_t) = \mu^2\sigma^2/(2\kappa - \sigma^2)$.

One should notice that in this example, the spot variance is an AR(1) process while both integrated and realized variances are ARMA(1,1) processes. In addition, the three processes have the same autoregressive root which equals $\exp(-\kappa\Delta)$.

When Δ is small, one gets

$$\beta_v = 1 - \kappa\Delta + o(\Delta), \qquad \beta_x = 1 - \frac{2}{3}\kappa\Delta + o(\Delta).$$

Likewise, when both Δ and δ/Δ are small, one gets

$$\beta_y = 1 - \frac{2}{3}\kappa\Delta - 2\frac{\delta}{\Delta}\frac{\mathbb{E}(V_t^2)}{Var(V_t)} + o(\Delta) + o(\delta/\Delta)$$

with $Var(V_t) = \mu^2\sigma^2/(2\kappa - \sigma^2)$ and $\mathbb{E}(V_t^2) = 2\kappa\mu^2/(2\kappa - \sigma^2)$. It is interesting to notice that, as $\delta/\Delta, \Delta \to 0$, we have

$$\beta_v - 1 \sim -\Delta\kappa, \qquad \beta_x - 1 \sim -\Delta\frac{2}{3}\kappa, \qquad \beta_y - 1 \sim -\Delta\frac{2}{3}\kappa - 4\frac{\delta}{\Delta}\frac{\kappa}{\sigma^2}, \tag{2.8}$$

that is, integrated variance has a larger first order autocorrelation than the spot and realized variances.

2.2.2 GARCH Diffusions with $\mathbb{E}(V_t) < \infty$

One can easily prove that

$$V_{t+\Delta} = \mu + \exp(-\kappa\Delta)(V_t - \mu) + \varepsilon_{t+\Delta}, \quad \varepsilon_{t+\Delta} = \sigma\int_t^{t+\Delta}\exp(-\kappa(t + \Delta - u))V_u du.$$

When $\mathbb{E}(V_t^2) < \infty$, $\varepsilon_{t+\Delta}$ is a martingale-difference-sequence (m.d.s.), which implies that

$$\mathbb{E}[V_{t+\Delta} \mid V_t] = \mu + \exp(-\kappa\Delta)(V_t - \mu). \tag{2.9}$$

However, the m.d.s. result is not valid when $\mathbb{E}(V_t^2) = \infty$ because $\int_0^{t+\Delta} \exp(-\kappa(t+\Delta-u))V_u du$ is not a martingale but a local martingale. Interestingly, we are able to prove that (2.9) is still valid when V_t is a stationary GARCH diffusion with $\mathbb{E}(V_t) < \infty$, whether $\mathbb{E}(V_t^2)$ is finite or not.[2]

Lemma 2.1. *For any $\Delta > 0$, we have*

$$\mathbb{E}[(v_{i+1} - \mu) - \exp(-\kappa\Delta)(v_i - \mu)|v_i] = 0. \tag{2.10}$$

When $\mathbb{E}(V_t^2) < \infty$, the previous result implies that V_t is an AR(1), from which one can estimate $\exp(-\kappa\Delta)$ by using an autoregression of order one of the spot variance. However, both the integrated and realized variances are ARMA(1,1) processes, which implies that first order autoregressions of these variables will not deliver a consistent estimator of the autoregressive parameter $\exp(-\kappa\Delta)$. However, Meddahi (2003) derived multi-period moment restrictions fulfilled by the integrated and realized variances when $\mathbb{E}(V_t^2) < \infty$. The following result proves that these multi-period moment restrictions are still valid when $\mathbb{E}(V_t^2) = \infty$.

Proposition 2.2. *Let $\Delta > 0$. (a) For $z = v, x$, we have*

$$\mathbb{E}[(z_{i+1} - \mu) - \exp(-\kappa\Delta)(z_i - \mu)|z_{i-1}] = 0. \tag{2.11}$$

(b) If $D_t = 0$ almost surely for all $t \geq 0$, then the result in Part (a) holds for $z = y$.

Proposition 2.2 will allow us to estimate consistently the coefficient $\exp(-\kappa\Delta)$ even when $\mathbb{E}(V_t^2) = \infty$ by using the following corollary:

Corollary 2.3. *Let $r : \mathbb{R} \to \mathbb{R}$ be bounded such that $\mathbb{E}[(z_i - \mu)r(z_{i-1} - \mu)] \neq 0$ for a given $\Delta > 0$. If $D_t = 0$ almost surely for all $t \geq 0$, we have for $z = v, x, y$ that*

$$\frac{\mathbb{E}[(z_{i+1} - \mu)r(z_{i-1} - \mu)]}{\mathbb{E}[(z_i - \mu)r(z_{i-1} - \mu)]} = \exp(-\kappa\Delta).$$

2.3. *Empirical Evidence of Fat Tails*

We now assess the magnitude of tails of empirical data. We use trade data on the SPDR S&P 500 ETF (SPY), which is an exchange traded fund (ETF) that tracks

[2] We are very grateful to Jean Jacod for providing us the proof of the result.

the S&P 500 index. Our primary sample comprises 10 years of trade data on SPY starting from June 15, 2004 through June 13, 2014 as available in the New York Stock Exchange Trade and Quote (TAQ) database. This tick-by-tick dataset has been cleaned according to the procedure outlined by Barndorff-Nielsen et al. (2008). We also remove short trading days leaving us with 2,497 days of trade data. In addition, we consider three subperiods: Before Crisis, from June 15, 2004 through August 29, 2008 (1,053 trading days); During Crisis, from September 2, 2008 through May 29, 2009 (185 trading days), and After Crisis, from June 1, 2009 through June 13, 2014 (1,259 trading days).

We estimate the tail index of the daily open-to-close returns and daily realized variance based on five minutes intra-day returns. Because we could have jumps that may affect the tail of the realized variance data, we also consider daily bipower variation which is a consistent estimator of integrated variance under the presence of jumps (see Barndorff-Nielsen and Shephard (2003, 2004b, 2006); Barndorff-Nielsen et al. (2005); Barndorff-Nielsen et al. (2006)) as well as the threshold estimator of integrated variance (see Jacod (2008, 2012); Mancini (2009); Jacod and Rosenbaum (2013)).

We estimate the tail index by using Hill's (1975) estimator. Let $(X_i)_{i=1}^n$ be a stationary time series with

$$\mathbb{P}[X_i > x] \sim x^{-\alpha}\ell(x), \quad x \to \infty,$$

for some slowly varying function ℓ. The Hill's estimator for α^{-1} which arose in the i.i.d. context as a conditional MLE is defined as

$$h = \frac{1}{k_n}\sum_{i=1}^{k_n} \log(X_{(i)}/X_{(k_n)}),$$

where $(X_{(i)})_{i=1}^n$ is the order statistics $X_{(n)} \leq \cdots \leq X_{(k_n)} \leq \cdots \leq X_{(1)}$ for some $k_n \leq n$ such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

The results by Hsing (1991) and Resnick and Stărică (1995) indicate that the Hill estimator is asymptotically quite robust with respect to deviations from independence; Resnick and Stărică (1998) prove consistency under ARCH-type dependence. See also Hill (2010) for some other processes including ARFIMA, FIGARCH, explosive

GARCH, nonlinear ARMA-GARCH and etc.

Valid standard errors of the Hill estimator are available only for some specific models with serial correlation. Therefore, we will not provide any of them. Instead, we follow the literature by providing Hill's plots, that is by varying the integer k_n. A flat area is viewed as a good estimator of the tail index. As usual, we truncate k_n. In practice we start with $k = 25$.

Figure 1 (a) depicts the Hill index of the returns and three volatility measures over the whole period for k_n between 25 and 500. The left panel provides the estimator for the returns, which is clearly below four. The right panel depicts the tail index of the three volatility measures. The plots suggest that the tail of these measures is below two. Observe that the three plots have flat areas, with a tail index between 1.2 and 1.4. One should notice that the plots for the three volatility measures are quite close.

The period considered in the previous figure includes the financial crisis. A natural question is whether the strong empirical evidence of fat tails is driven by the crisis' period. We therefore carry the Hill estimators for the periods before, during, and after crisis, as explained above. Given the length of the crisis period (185 trading days), we vary k_n from 25 to 150. Figure 1 (b) depicts the Hill index of the returns on the left panel and the realized volatility on the right panel for the three periods, while Figure 1 (c) depicts those of the bipower (left panel) and threshold (right panel) measures. Clearly, the crisis period exhibits fatter tails than the other two periods for the four variables. However, both the periods before and after the crisis suggest very fat tails with a tail index slightly below four for the returns and around two for the three volatility measures. Therefore, the evidence of fat tails and unbounded second moment for the volatility measures is quite strong.

3. Least Square Estimates

In this section, we consider the OLS estimator $\hat{\beta}_z^{(k)}$ for $\beta_z^{(k)}$ in (2.3) given by

$$\hat{\beta}_z^{(k)} = \frac{\sum_{i=k+1}^{N-1}(z_{i-k} - \overline{z}_N)z_{i+1}}{\sum_{i=k+1}^{N-1}(z_{i-k} - \overline{z}_N)^2}$$

where $\overline{z}_N$ is the sample mean of $(z_{i-k} : i = k+1, \cdots, N-1)$. For $k = 0$, we simply write $\beta_z^{(0)} = \beta_z$ and $\hat{\beta}_z^{(0)} = \hat{\beta}_z$.

3.1. *Primary Asymptotics*

Recall that $T = N\Delta$ and $\Delta = n\delta$. For our asymptotics here we let $\delta/\Delta, \Delta \to 0$, with T being fixed or $T \to \infty$ simultaneously as $\delta/\Delta, \Delta \to 0$. In case we have $\delta/\Delta, \Delta \to 0$ and $T \to \infty$ simultaneously, we assume that $\delta/\Delta, \Delta \to 0$ sufficiently fast relative to $T \to \infty$. It is indeed more relevant in a majority of practical applications, which rely on observations collected at small sampling intervals over moderately long span.

In our asymptotics, we frequently deal with various functional transforms of D and V over time interval $[0, T]$. To effectively handle such functional transforms, we define

$$T_D = \max_{0 \le t \le T} |D_t| \quad \text{and} \quad T_V(f) = \max_{0 \le t \le T} |f(V_t)|$$

for some function $f : \mathcal{D} \to \mathbb{R}$. We also denote by ι the identity function on $\mathcal{D}$, and $\iota(v) = v$ for all $v \in \mathcal{D}$. Consequently, we have $T_V(\iota) = \max_{0 \le t \le T} |V_t|$ for the identity function. Obviously, T_D and $T_V(\iota)$ are the asymptotic orders of extremal process of D and V, respectively. The order of the extremal process is known for a wide class of diffusions. For instance, under some regularity conditions, the extremal process of a stationary diffusion V is of order $O_p(s^{-1}(T))$, where s is the scale function of V, to which the reader is referred, e.g., Davis (1982). More generally, we may obtain the exact rate of $T_V(f)$ from the asymptotic behavior of extremal process. In particular, if f is regularly varying and c_T is the order of the extremal process, then the asymptotic order of $T_V(f)$ is given by $O_p(f(c_T))$.

Assumption 3.1. *(a) σ^2 is twice continuously differentiable on $\mathcal{D}$, and (b) for $f = \mu$, σ^2, $\sigma^{2\prime}$, $\sigma^{2\prime\prime}$ and ι, there is a locally bounded function $\omega : \mathcal{D} \to \mathbb{R}$ such that $|f(v)| \le \omega(v)$ for all $v \in \mathcal{D}$.*

The differentiability condition of σ^2 in Assumption 3.1 (a) is routinely assumed in the study of diffusion models. Under Assumption 3.1 (a), the majorizing function ω in Assumption 3.1 (b) always exists as long as μ is locally bounded.

Assumption 3.2. *For ω in Assumption 3.1, $\Delta T_V(\omega^8) T^2 \log(T/\Delta) \to_p 0$.*

Assumption 3.3. *For ω in Assumption 3.1,* $(\delta/\Delta)T_V(\omega^8)T^2\log^3(T/\delta)\to_p 0$.

Assumption 3.4. $(\delta/\Delta)T_D^4T\to_p 0$.

Assumption 3.5. $(\delta/\Delta^2)=O(1)$.

Assumption 3.2 is similar to Assumption 5.1 in Kim and Park (2017), and provides a sufficient condition for our primary asymptotics of spot variance (v_i) and integrated variance (x_i). On the other hand, the asymptotics of realized variance (y_i) involve three parameters, δ, Δ and T, and require Assumptions 3.3-3.5 in addition to Assumption 3.2. The role of Assumption 3.3 is to analyze the asymptotic effect of the errors (x_i-y_i) in the OLS estimates. On the other hand, Assumption 3.4 is a condition to control the effects from the drift part (D_t) in (P_t) so that (D_t) has no asymptotic impact in the asymptotics of the OLS estimates with (y_i). Lastly, Assumption 3.5 is to exclude less interesting cases where the errors (x_i-y_i) dominate the signals (x_i) in the OLS estimates with (y_i). In particular, if $\delta/\Delta^2\to\infty$, then the error components may have bigger stochastic order than the signals.

Assumptions 3.2-3.4 make it necessary to have $\Delta\to 0$ and $\delta/\Delta\to 0$. For a fixed T, a set of necessary and sufficient conditions for Assumptions 3.2-3.4 is $\Delta\to 0$ and $(\delta/\Delta)\log^3(1/\delta)\to 0$. Our asymptotics in the paper are derived under the conditions $\Delta\to 0$, $\delta/\Delta\to 0$ and $T\to\infty$ jointly. For Assumptions 3.2-3.4 to hold, it requires $\Delta\to 0$ and $\delta/\Delta\to 0$ sufficiently fastly as $T\to\infty$. For instance, (i) Assumption 3.2 holds as long as $\Delta=O(T^{-2-\epsilon})$, and (ii) Assumptions 3.3-3.4 hold as long as $(\delta^{1-\epsilon}/\Delta)T^{2+\epsilon}$ for some $\epsilon>0$, if V and D are bounded with $T_V(\omega^8)=O_p(1)$ and $T_D=O_p(1)$. For example, if daily realized variances (y_i) are obtained using 5 minutes returns over five years, then $\delta/\Delta=1/288$, $\Delta=1/250$ and $T=5$. In this case, both ΔT^2 and $(\delta/\Delta)T^2$ are small ($\Delta T^2=1/10$ and $(\delta/\Delta)T^2=25/288$). Our asymptotics in this section hold jointly in δ, Δ and T under Assumptions 3.1-3.5, and we do not use sequential asymptotics, requiring $\delta/\Delta\to 0$, $\Delta\to 0$ and $T\to\infty$ sequentially.

To effectively explain our asymptotics, we apply the summation by parts to the numerator of $\hat{\beta}_z^{(k)}$, and rewrite it as

$$\hat{\beta}_z^{(k)} - 1 = \frac{1}{2}\frac{\sum_{j=0}^{k}\left((z_{N-j}^2 - z_{1+j}^2) - \overline{z}_N(z_{N-j} - z_{1+j})\right)}{\sum_{i=k+1}^{N-1}(z_{i-k} - \overline{z}_N)^2} - \frac{1}{2}\frac{\sum_{i=k+1}^{N-1}(z_{i+1} - z_{i-k})^2}{\sum_{i=k+1}^{N-1}(z_{i-k} - \overline{z}_N)^2}. \tag{3.1}$$

For each term in (3.1), we have the following continuous time approximations when $\Delta \to 0$ and $\delta/\Delta \to 0$ such that Assumptions 3.1 and 3.2 holds.

Lemma 3.1. *Let Assumptions 3.1-3.5 hold.*

(a) For $k \geq 0$, we have

$$\sum_{j=0}^{k}\left((z_{N+j}^2 - z_{1+j}^2) - \overline{z}_N(z_{N+j} - z_{1+j})\right) \sim_p (1+k)\left(V_T^2 - V_0^2 - \overline{V}_T(V_T - V_0)\right),$$

$$\sum_{i=k+1}^{N-1}(z_{i-k} - \overline{z}_N)^2\Delta \sim_p \int_0^T (V_t - \overline{V}_T)^2 dt,$$

where $\overline{V}_T = T^{-1}\int_0^T V_t dt$, for $z = v, x, y$.

(b) For $k \geq 0$, we have

$$\sum_{i=k+1}^{N-1}(z_{i+1} - z_{i-k})^2 \sim_p \sum_{i=1}^{N-1}(z_{i+1} - z_i)^2 + k\sum_{i=1}^{N-1}(v_{i+1} - v_i)^2$$

for $z = v, x, y$, and

$$\sum_{i=1}^{N-1}(z_{i+1} - z_i)^2 \sim_p \begin{cases} [V]_T, & \textit{for } z = v \\ (2/3)[V]_T, & \textit{for } z = x \\ (2/3)[V]_T + (4\delta/\Delta^2)\int_0^T V_t^2 dt, & \textit{for } z = y. \end{cases} \tag{3.2}$$

Remark 3.1. (a) The continuous time approximations of the sum of squared increments (SSI), $\sum_{i=k+1}^{N-1}(z_{i+1} - z_i)^2$, in Lemma 3.1 (b) are depending upon z.[3] In partic-

[3]Note that (3.2) is an extension of the "integral-to-spot device" in Mykland and Zhang (2017) to the large T setting.

ular, we have

$$\sum_{i=k+1}^{N-1}(x_{i+1}-x_i)^2 < \sum_{i=k+1}^{N-1}(v_{i+1}-v_i)^2, \tag{3.3}$$

$$\sum_{i=k+1}^{N-1}(x_{i+1}-x_i)^2 < \sum_{i=k+1}^{N-1}(y_{i+1}-y_i)^2 \tag{3.4}$$

with probability approaching one as $\delta/\Delta, \Delta \to 0$ under Assumptions 3.1-3.5. An intuitive explanation for the inequalities in (3.3) and (3.4) are as follow. We can naturally expect that the integrated variance (x_i) has a smoother sample path compare to that of the spot variance (v_i). As a result, the SSI of (x_i) tends to be smaller than that of (v_i), and we have the first inequality in (3.3). This downward bias is often referred to as smoothing bias. For the use of this terminology, see Stoker (1993) in nonparametric density estimations, and Aït-Sahalia et al. (2013) and Mykland and Zhang (2017) in the high frequency setting. On the other hand, the realized variance (y_i) is a noisy measure of the integrated variance (x_i), and the error component in (y_i) generates additional variations. Consequently, a sample path of (y_i) becomes rougher compare to that of (x_i), and hence, (3.4) holds.

(b) Unlike Lemma 3.1 (b), the continuous time approximations in Lemma 3.1 (a) are identical for all $z = v, x, y$. The results in Lemma 3.1 (a) are well expected since $|z_i - V_{(i-1)\Delta}| \to_p 0$ for all z as long as δ/Δ and Δ are sufficiently small relative to T.

(c) It follows from Lemma 3.1 with $k = 0$ that

$$\begin{aligned}\sum_{i=1}^{N-1}(z_i-\overline{z}_N)(z_{i+1}-z_i) &= \frac{1}{2}\{z_N^2 - z_1^2 - \overline{z}_N(z_N - z_1)\} - \frac{1}{2}\sum_{i=1}^{N-1}(z_{i+1}-z_i)^2 \\ &\sim_p \frac{1}{2}\{V_T^2 - V_0^2 - \overline{V}_T(V_T - V_0)\} - \begin{cases} (1/2)[V]_T, & \text{for } z = v \\ (1/3)[V]_T, & \text{for } z = x \\ (1/3)[V]_T + (2\delta/\Delta^2)\int_0^T V_t^2 dt, & \text{for } z = y \end{cases} \\ &= \begin{cases} \int_0^T (V_t - \overline{V}_T) dV_t, & \text{for } z = v \\ \int_0^T (V_t - \overline{V}_T) dV_t + (1/6)[V]_T, & \text{for } z = x \\ \int_0^T (V_t - \overline{V}_T) dV_t + (1/6)[V]_T - (2\delta/\Delta^2)\int_0^T V_t^2 dt, & \text{for } z = y, \end{cases}\end{aligned} \tag{3.5}$$

where the last equality follows from Ito's lemma, as $\delta/\Delta, \Delta \to 0$ under Assumptions

3.1-3.5. The result (3.5) for $z = v$ is quite natural and expected by the asymptotic negligibility of discretization errors when $\Delta \to 0$. In a similar argument, one may expect

$$\sum_{i=1}^{N-1}(z_i - \overline{z}_N)(z_{i+1} - z_i) \sim_p \int_0^T (V_t - \overline{V}_T)dV_t \quad \text{for} \quad z = x, y \tag{3.6}$$

since $\sup_{0\leq i\leq N} |z_i - v_i| \to_p 0$ as $\delta/\Delta, \Delta \to 0$. However, we have (3.5), and the conjecture (3.6) is not true. This is not surprising at all since the convergence of stochastic process does not necessarily imply the convergence of stochastic integral associated with the stochastic process (see, e.g., Kurtz and Protter (1991)). In particular, the different asymptotics depending on z in (3.5) are caused by the fact that the SSIs of x and y are biased estimators for the quadratic variation $[V]$ of V, whereas the SSI of v is unbiased.

The primary asymptotics for $\hat{\beta}_z^{(k)}$ can be easily obtained by successively applying Lemma 3.1 and Ito's lemma to (3.1).

Proposition 3.2. *Under Assumptions 3.1-3.5, we have*

$$\begin{aligned}
\hat{\beta}_v - 1 &\sim_p \Delta \frac{\int_0^T (V_t - \overline{V}_T)dV_t}{\int_0^T (V_t - \overline{V}_T)^2 dt}, \\
\hat{\beta}_x - 1 &\sim_p \Delta \frac{\int_0^T (V_t - \overline{V}_T)dV_t + (1/6)[V]_T}{\int_0^T (V_t - \overline{V}_T)^2 dt}, \\
\hat{\beta}_y - 1 &\sim_p \Delta \frac{\int_0^T (V_t - \overline{V}_T)dV_t + (1/6)[V]_T - (2\delta/\Delta^2)\int_0^T V_t^2 dt}{\int_0^T (V_t - \overline{V}_T)^2 dt}
\end{aligned}$$

and $\hat{\beta}_z^{(k)} - 1 \sim_p (\hat{\beta}_z - 1) + k(\hat{\beta}_v - 1)$.

Remark 3.2. (a) As explained in Remark 3.1 (a), (x_i) has smoother sample paths than (v_i), and hence, we have (3.3). In Proposition 3.2, we have $\hat{\beta}_v < \hat{\beta}_x$ which implies that (x_i) tends to have more persistent sample paths than (v_i). This result is a consequence of (3.3). Moreover, $\hat{\beta}_y$ is downward biased with $\hat{\beta}_y < \hat{\beta}_x$ which is induced by the errors in (y_i).[4]

[4] A similar downward bias in volatility regressions, but for a discrete time ARMA process, can be found in Hansen and Lunde (2014).

(b) Note that Assumptions 3.3-3.5 do not necessarily imply $\delta/\Delta^2 \to 0$. Therefore, the speeds of $\delta \to 0$ and $\Delta \to 0$ are important in the asymptotic negligibility of the estimation errors of (y_i). In particular, if $\delta/\Delta^2 \to 0$ sufficiently quickly, then the errors of (y_i) become asymptotically negligible, and hence, we may have $\hat{\beta}_y - 1 \sim_p \hat{\beta}_x - 1$. In Section 4.3, we analyze the asymptotics negligibility of the estimation errors of (y_i) for more general class of estimators, including the OLS and IV estimators, under the presence of leverage effects.

3.2. *Long Span Asymptotics*

The primary asymptotics in Proposition 3.2 do not require $T \to \infty$. In particular, if T is fixed, then $(N/T)(\hat{\beta}_z - 1) = (1/\Delta)(\hat{\beta}_z - 1)$ is random for all $z = v, x, y$, and is determined by a particular realization of the underlying variance process V. Under the fixed T asymptotic scheme, the law of motion of V is less important. In particular, the results in Proposition 3.2 require neither certain moment conditions nor stationarity. However, the underlying probabilistic structure of V is crucial in the development of the large T asymptotics.

In our long span asymptotics, we only consider a stationary diffusion V to effectively analyze consequences of fat tails in the volatility regressions. Throughout the paper we assume that the scale function s in (2.5) and the speed density m in (2.6) satisfy the following conditions:

Assumption 3.6. *(a)* $s(\underline{v}) = -\infty$ *and* $s(\overline{v}) = \infty$, *and (b)* $\int_{\mathcal{D}} m(v)dv < \infty$.

Under Assumption 3.6 (a), V becomes recurrent. Moreover, Assumption 3.6 implies that V is positive recurrent. Positive recurrent diffusions have time invariant distributions, and if they are started from the time invariant distributions they become stationary. The time invariant density of the positive recurrent diffusion V is given by $\pi(v) = m(v)/\int_{\mathcal{D}} m(v)dv$. Therefore, conditions on unconditional moments are characterized by corresponding m-integrability conditions. For instance, $\mathbb{E}(f(V_t)) < \infty$ if and only if f is m-integrable, since $\mathbb{E}(f(V_t)) = \int_{\mathcal{D}} f(v)\pi(v)dv$ and $\pi(v) = m(v)/\int_{\mathcal{D}} m(v)dv$ with $\int_{\mathcal{D}} m(v)dv < \infty$.

Since we allow fat tails, we consider not only integrable functions but also nonintegrable functions with respect to the speed density m of V. We will not require any regularity conditions for m-integrable functions. To effectively analyze m-nonintegrable functions, however, we need some regularity conditions. Following Kim and Park (2017), it will be maintained throughout the paper that all m-nonintegrable functions f are m-regularly varying, i.e., mf is regularly varying on $\mathcal{D}$. For a m-nonintegrable function f, we say that f is m-strongly nonintegrable if $f\ell$ is not m-integrable for any slowly varying function ℓ on $\mathcal{D}$. On the other hand, we say that f is m-nearly integrable if $f\ell$ is m-integrable for some slowly varying function ℓ on $\mathcal{D}$.

We assume that

Assumption 3.7. *(a) s' is regularly varying or rapidly varying with index $c \neq -1$, (b) σ^2 is regularly varying, and (c) $f = \sigma^2, \iota^2$ is either m-integrable or m-strongly nonintegrable.*

Assumption 3.7 (a) and (b) appear in Kim and Park (2018), and are mild enough to include most diffusion processes used in practice. The reader is also referred to Bingham et al. (1993) for more discussions about the regularly and rapidly varying functions. In Assumption 3.7 (c), we assume that σ^2 and ι^2 are m-strongly nonintegrable as long as they are not m-integrable. This assumption is a technical condition to simplify our discussions below. Our subsequent theory can also be developed under the m-near integrablility at the cost of more involved analysis (see Kim and Park (2017, 2018) for the related discussions).

In the following, we let $f_s = f \circ s^{-1}$ for any function f on $\mathcal{D}$ other than m.[5] Moreover, for a regularly varying function f on $\mathbb{R}$, we define its limit homogeneous function $\overline{f}$ as $f(\lambda v)/f(\lambda) \to \overline{f}(v)$ as $\lambda \to \infty$ for all $v \neq 0$.

We define numerical sequences p_T and q_T as

$$p_T = \begin{cases} T & \text{if } \sigma^2 \text{ is } m\text{-integrable} \\ T^2(m_s\sigma_s^2)(T) & \text{if } \sigma^2 \text{ is } m\text{-strongly nonintegrable} \end{cases}$$

$$q_T = \begin{cases} T & \text{if } \iota^2 \text{ is } m\text{-integrable} \\ T^2(m_s\iota_s^2)(T) & \text{if } \iota^2 \text{ is } m\text{-strongly nonintegrable,} \end{cases}$$

[5] In Section 2.1, m_s is defined as $m_s = (m/s') \circ s^{-1}$ which is the speed density of natural scale diffusion $V^s = s(V)$ of the underlying diffusion V.

and let

$$P = \begin{cases} \mathbb{E}(\sigma^2(V_t)) & \text{if } \sigma^2 \text{ is } m\text{-integrable} \\ \int_0^\tau \overline{m_s \sigma_s^2}(B_t)dt & \text{if } \sigma^2 \text{ is } m\text{-strongly nonintegrable} \end{cases}$$

$$Q = \begin{cases} \mathbb{E}(V_t^2) & \text{if } \iota^2 \text{ is } m\text{-integrable} \\ \int_0^\tau \overline{m_s \iota_s^2}(B_t)dt & \text{if } \iota^2 \text{ is } m\text{-strongly nonintegrable} \end{cases}$$

$$S = \begin{cases} \mathbb{E}(V_t^2) - (\mathbb{E}(V_t))^2 & \text{if } \iota^2 \text{ is } m\text{-integrable} \\ \int_0^\tau \overline{m_s \iota_s^2}(B_t)dt & \text{if } \iota^2 \text{ is } m\text{-strongly nonintegrable,} \end{cases}$$

where B is Brownian motion and $\tau = \inf\{t \,\big|\, L(t,0) > 1/\int_{\mathcal{D}} m(v)dv\}$ with Brownian local time $L(\cdot, 0)$ of B at the origin (i.e., $L(t,0) = \lim_{\epsilon\to 0}(2\epsilon)^{-1}\int_0^t 1\{|B_s| < \epsilon\}ds$). Under Assumption 3.7, both (p_T, q_T) and (P, Q, S) are well defined (see Kim and Park (2017)).

Lemma 3.3. *Let Assumption 3.7 hold. As $T \to \infty$, we have $Tp_T/q_T \to \infty$ and*

$$\frac{1}{p_T}[V]_T \to_d P, \qquad \frac{1}{p_T}\int_0^T (V_t - \overline{V}_T)dV_t \to_d -\frac{P}{2},$$
$$\frac{1}{q_T}\int_0^T V_t^2 dt \to_d Q, \qquad \frac{1}{q_T}\int_0^T (V_t - \overline{V}_T)^2 dt \to_d S.$$

Under the m-integrability of $f = \iota^2, \sigma^2$, Lemma 3.3 becomes a standard law of large numbers of stationary diffusions. However, if $f = \iota^2, \sigma^2$ is not m-integrable, the standard limit theory is not applicable and we have completely different limit theory. In particular, the limit distribution $\int_0^\tau \overline{m_s f_s}(B_t)dt$ is not Gaussian and is highly nonstandard. Moreover, the normalizing sequence $T^2(m_s f_s)(T)$ diverges faster than T since the function $m_s f_s$ becomes a regularly varying function with index $c > -1$ as long as f is not m-integrable and Assumption 3.7 holds. The reader is referred to Kim and Park (2017) for more detailed discussions about the asymptotics of diffusion functionals.

The long span asymptotics for $\hat{\beta}_z$ follow immediately from Proposition 3.2 with Lemma 3.3.

Theorem 3.4. *Let Assumptions 3.1-3.7 hold. As $\delta/\Delta, \Delta \to 0$ and $T \to \infty$, we have*

$$\hat{\beta}_v - 1 \sim_d -\Delta \frac{p_T}{q_T} \frac{P}{2S}, \qquad \hat{\beta}_x - 1 \sim_d -\Delta \frac{p_T}{q_T} \frac{P}{3S}, \qquad \hat{\beta}_y - 1 \sim_d -\Delta \frac{p_T}{q_T} \frac{P}{3S} - \frac{\delta}{\Delta} \frac{2Q}{S}.$$

As shown in Lemma 3.3 that P, Q and S become constants only when both ι^2 and σ^2 are m-integrable. The relation $\sim_d$ in Theorem 3.4 becomes $\sim_p$ if P, Q and S are all constants. On the other hand, if ι^2 and σ^2 are not m-integrable, then P, Q and S remain random. In this case, Theorem 3.4 implies that $\hat{\beta}_z - 1$ is random for all small Δ.

Remark 3.3. The results in Theorem 3.4 can be applied to a broad class of volatility processes used in the literature.

(a) If both σ^2 and ι^2 are m-integrable, then $p_T = q_T = T$ and

$$\hat{\beta}_v - 1 \sim_p -\Delta \frac{\mathbb{E}(\sigma^2(V_t))}{2Var(V_t)}, \quad \hat{\beta}_x - 1 \sim_p \frac{2}{3}(\hat{\beta}_v - 1), \quad \hat{\beta}_y - 1 \sim_p (\hat{\beta}_x - 1) - \frac{\delta}{\Delta} \frac{2\mathbb{E}(V_t^2)}{Var(V_t)}.$$

(b) For a stationary Ornstein-Uhlenbeck process V, given as

$$dV_t = \kappa(\mu - V_t)dt + \sigma dW_t,$$

we have $\mathbb{E}(\sigma^2(V_t)) = \sigma^2$, $Var(V_t) = \sigma^2/(2\kappa)$ and $\mathbb{E}(V_t^2) = \sigma^2/(2\kappa) + \mu^2$. Therefore,

$$\hat{\beta}_v - 1 \sim_p -\Delta\kappa, \quad \hat{\beta}_x - 1 \sim_p -\Delta \frac{2}{3}\kappa, \quad \hat{\beta}_y - 1 \sim_p -\Delta \frac{2}{3}\kappa - 2\frac{\delta}{\Delta}\left(1 + \frac{2\kappa\mu^2}{\sigma^2}\right). \quad (3.7)$$

(c) Let V be a stationary GARCH diffusion (2.7) with $\sigma^2 < 2\kappa$ so that $\mathbb{E}(V_t^2) < \infty$. In this case, we have $\mathbb{E}(\sigma^2(V_t)) = \sigma^2\mathbb{E}(V_t^2)$, $Var(V_t) = \mu^2\sigma^2/(2\kappa - \sigma^2)$ and $\mathbb{E}(V_t^2) = 2\kappa\mu^2/(2\kappa - \sigma^2)$, and hence, Theorem 3.4 implies

$$\hat{\beta}_v - 1 \sim_p -\Delta\kappa, \qquad \hat{\beta}_x - 1 \sim_p -\Delta \frac{2}{3}\kappa, \qquad \hat{\beta}_y - 1 \sim_p -\Delta \frac{2}{3}\kappa - 4\frac{\delta}{\Delta}\frac{\kappa}{\sigma^2}. \qquad (3.8)$$

It is interesting to note that the results (3.8) are the same as (2.8) for population regressions derived by Andersen et al. (2004).

(d) Let V be a stationary GARCH diffusion (2.7) with $2\kappa < \sigma^2$ so that $\mathbb{E}(V_t^2) =$

$\mathbb{E}(\sigma^2(V_t)) = \infty$. In this case, $p_T = \sigma^2 q_T$ and $P = Q = S$, and therefore, we have

$$\hat{\beta}_v - 1 \sim_p -\Delta \frac{1}{2}\sigma^2, \qquad \hat{\beta}_x - 1 \sim_p -\Delta \frac{1}{3}\sigma^2, \qquad \hat{\beta}_y - 1 \sim_p -\Delta \frac{1}{3}\sigma^2 - 2\frac{\delta}{\Delta}. \tag{3.9}$$

Under $\mathbb{E}(V_t^2) < \infty$, as shown in Remark 3.3 (c), the limits of $(\hat{\beta}_z - 1)/\Delta$ are mainly determined by the mean reversion parameter κ in the drift function $\mu(v)$. Under $\mathbb{E}(V_t^2) = \infty$, the limits $(\hat{\beta}_z - 1)/\Delta$ are still constant, but they are determined by the diffusion parameter σ^2 in the diffusion function $\sigma^2(v)$.

We also note that GARCH diffusion is a special example that $(\hat{\beta}_z - 1)/\Delta$ has a degenerated constant limit even under $\mathbb{E}(V_t^2) = \infty$, which is induced by the relationship $v^2 \propto \sigma^2(v)$ between the quadratic function v^2 and the diffusion function $\sigma^2(v)$. For any other models which do not satisfy $\iota^2(v) \propto \sigma^2(v)$ asymptotically, $(\hat{\beta}_z - 1)$ has a random limit, after proper normalization, as long as $\mathbb{E}(V_t^2) = \infty$ or $\mathbb{E}(\sigma^2(V_t)) = \infty$. This is the case for the CEV process considered below.

(e) Let V be a stationary CEV process

$$dV_t = \kappa(\mu - V_t)dt + \sigma V_t^\gamma dW_t.$$

If $\kappa, \mu, \sigma > 0$ and $1 < \gamma < 3/2$, then $\mathbb{E}(V_t^2) = \infty$ and $\mathbb{E}(\sigma^2(V_t)) = \infty$ since $m(v) \sim v^{-2\gamma}$ as $v \to \infty$. For the CEV process, we have

$$\begin{aligned} p_T &= \sigma T^2 (m_s \iota_s^{2\gamma})(T), & q_T &= T^2 (m_s \iota_s^2)(T), \\ P &= \int_0^\tau \overline{m_s \iota_s^{2\gamma}}(B_t)dt, & Q = S &= \int_0^\tau \overline{m_s \iota_s^2}(B_t)dt, \end{aligned}$$

where $p_T/q_T = \sigma \iota_s^{2\gamma-2}(T) \to \infty$ as $T \to \infty$, since $\gamma > 1$ and $\iota_s = s^{-1}$ is monotonically increasing by the recurrence property. Clearly, $P \neq S$ for any $\gamma \in (1, 3/2)$, and hence, P/S remains random unlike the GARCH diffusion. Therefore, $\hat{\beta}_z - 1$ has random limit for all sufficiently small Δ.

(f) Our example in Remark 3.3 (d) should be contrasted to the limit theory for the sample autocorrelations of GARCH(1,1) processes with fat tails obtained in Mikosch and Starica (2000). Let

$$X_i = \sigma_i Z_i \quad \text{with} \quad \sigma_i^2 = \alpha_0 + \beta_1 \sigma_{i-1}^2 + \alpha_1 X_{i-1}^2 \quad \text{for} \quad i = 1, 2, \cdots, N,$$

where (Z_i) is a sequence of i.i.d. symmetric random variables with $\mathbb{E}Z_i^2 = 1$. Under some assumptions, which imply that the vector (X_i, σ_i) is regularly varying with index $p > 0$, it is shown that for $p \in (0, 4)$ the variance process (σ_i^2) has unbounded variance and satisfies, for any $k \geq 1$,

$$\left(\frac{\sum_{i=1}^{N-k} X_i^2 X_{i+k}^2}{\sum_{i=1}^{N} X_i^4} - 1, \frac{\sum_{i=1}^{N-k} \sigma_i^2 \sigma_{i+k}^2}{\sum_{i=1}^{N} \sigma_i^4} - 1 \right) \sim_d \left(\frac{\Sigma_{1,X^2} - \Sigma_{0,X^2}}{\Sigma_{0,X^2}}, \frac{\Sigma_{1,\sigma^2} - \Sigma_{0,\sigma^2}}{\Sigma_{0,\sigma^2}} \right),$$

where the limit distribution is nondegenerated since the vector $(\Sigma_{m,X^2}, \Sigma_{m,\sigma^2})_{m=0,1}$ is $p/2$-stable. This contrasts with our result for a GARCH diffusion with unbounded variance (see Remark 3.3 (d)), in which $(\hat{\beta}_z - 1)/\Delta$ has a constant limit for $z = v, x$. We think that the difference between our results and those of Mikosch and Starica (2000) is due to the fact that we allow $\Delta \to 0$. We conjecture that the result would be the same when Δ is fixed.

4. Instrumental Variable Estimations

In this section we will study various IV estimators of $\beta_z^{(k)}$ in (2.3). When one has a model like

$$z_{i+1} = \alpha + \beta z_i + u_{i+1}, \quad |\beta| \leq 1$$

where (u_i) is i.i.d. with possible fat tails, it is well known that in general the OLS estimator of β is consistent (see Hannan and Kanter (1977); Knight (1987) for $|\beta| < 1$, and Chan and Tran (1989); Phillips (1990) for $\beta = 1$). However, the OLS is not necessarily efficient and its asymptotic distribution could be non-Gaussian. An alternative method that could lead to more efficient estimators or asymptotically Gaussian ones is to consider a signed power estimator defined as[6]

$$\tilde{\beta} = \frac{\sum \operatorname{sign}(z_i) \mid z_i \mid^c (z_{i+1} - \bar{z}_N)}{\sum \operatorname{sign}(z_i) \mid z_i \mid^c (z_i - \bar{z}_N)}.$$

The reader is referred to Samorodnitsky et al. (2007) for the asymptotics of OLS and the signed power estimator. One can easily prove that this estimator is indeed the

[6]The signed power estimator can be modified by replacing $\operatorname{sign}(z_i)|z_i|^c$ with $\operatorname{sign}(z_i - \bar{z}_N)|z_i - \bar{z}_N|^c$ so that the resulting estimator can adjust its sample mean.

empirical counterpart of the IV estimator defined by

$$\mathbb{E}\left[\begin{pmatrix} 1 \\ \operatorname{sign}(z_i)|z_i|^c \end{pmatrix}(z_{i+1}-\alpha-\beta z_i)\right]=0.$$

Typically, the power c is smaller than one in order to reduce the tails of the moments involved in the estimation method. An extreme case is the Cauchy estimator which corresponds to $c = 0$, that is when the instrument equals the sign of z_i.

In the following subsection, we will study IV estimators of $\beta_z^{(k)}$ in (2.3) which have the form

$$\tilde{\beta}_z^{(k)} = \frac{\sum_{i=k+1}^{N-1} r(z_{i-k})(z_{i+1}-\overline{z}_N)}{\sum_{i=k+1}^{N-1} r(z_{i-k})(z_{i-k}-\overline{z}_N)},$$

where we use a functional transformation $r(z_{i-k})$ of z_{i-k} as an instrument. We prove below that the IV estimator, with a proper choice of instrument, is robust to fat tails.

Interestingly, Jean-Marie Dufour used the sign-based methods in several studies for inference purposes, especially for exact inference in finite sample. In particular, he used such an approach in Coudin and Dufour (2009, 2017) in order to provide inference about the slope parameter in a linear regression model without making moment restrictions on the disturbance errors and therefore allowing for fat tails. The assumption made in these papers is a median restriction on the errors conditional on the explanatory variables. In other words, we are using the same approach with a slightly different framework because we assume that the conditional first moment of the errors exists and equals zero, but we do not make assumptions on higher moments.

4.1. *IV Estimator $\tilde{\beta}_z^{(k)}$ with a Current Instrument*

Let r be continuously differentiable, and define $r_1(z) = \int_{z_0}^{z} r(x)dx$ for some $z_0 \in \mathcal{D}$. Then, by Taylor expansion, we have

$$\begin{aligned}\sum_{j=0}^{k}(r_1(z_{N-j})-r_1(z_{1+j})) &= \sum_{i=k+1}^{N-1}(r_1(z_{i+1})-r_1(z_{i-k})) \\ &= \sum_{i=k+1}^{N-1} r(z_{i-k})(z_{i+1}-z_{i-k}) + \frac{1}{2}\sum_{i=k+1}^{N-1} r'(z_{i-k}^*)(z_{i+1}-z_{i-k})^2\end{aligned}$$

for $(z_i^*)_{i=1}^{N-1}$ such that $z_i^* \in [z_{i-k}, z_{i+1}]$. Using the expansion, we may rewrite $\tilde{\beta}_z^{(k)}$ as

$$\tilde{\beta}_z^{(k)} - 1 = \frac{\sum_{j=0}^{k}(r_1(z_{N-j}) - r_1(z_{1+j}))}{\sum_{i=k+1}^{N-1} r(z_{i-k})(z_{i-k} - \overline{z}_N)} - \frac{1}{2}\frac{\sum_{i=k+1}^{N-1} r'(z_{i-k}^*)(z_{i+1} - z_{i-k})^2}{\sum_{i=k+1}^{N-1} r(z_{i-k})(z_{i-k} - \overline{z}_N)}. \tag{4.1}$$

As in Lemma 3.1, we may obtain the continuous time approximation for each term in (4.1). For the approximation, we require

Assumption 4.1. *(a) r is three times continuously differentiable on $\mathbb{R}$ with $r'(z) > 0$ for all $z \in \mathbb{R}$, and (b) r and its derivatives are all majorized by the function ω in Assumption 3.1.*

The role of Assumption 4.1 is similar to Assumption 3.1, and make it convenient to develop the continuous time approximations if combined with the conditions on δ, Δ and T in Assumptions 3.2-3.5.

Proposition 4.1. *Let Assumptions 3.1-3.5 and 4.1 hold.*

(a) For $k \geq 0$, we have

$$\sum_{j=0}^{k}(r_1(z_{N-j}) - r_1(z_{1+j})) \sim_p (1+k)(r_1(V_T) - r_1(V_0)),$$
$$\sum_{i=k+1}^{N-1} r(z_{i-k})(z_{i-k} - \overline{z}_N)\Delta \sim_p \int_0^T r(V_t)(V_t - \overline{V}_T)dt$$

for $z = v, x, y$.

(b) For $k \geq 0$, we have

$$\sum_{i=k+1}^{N-1} r'(z_{i-k}^*)(z_{i+1} - z_{i-k})^2 \sim_p \sum_{i=1}^{N-1} r'(z_i)(z_{i+1} - z_i)^2 + k\sum_{i=1}^{N-1} r'(v_i)(v_{i+1} - v_i)^2$$

for $z = v, x, y$, and

$$\sum_{i=1}^{N-1} r'(z_i)(z_{i+1} - z_i)^2 \sim_p \begin{cases} \int_0^T r'(V_t)d[V]_t, & \text{for } z = v \\ (2/3)\int_0^T r'(V_t)d[V]_t, & \text{for } z = x \\ (2/3)\int_0^T r'(V_t)d[V]_t + (4\delta/\Delta^2)\int_0^T r'(V_t)V_t^2 dt, & \text{for } z = y \end{cases}$$

(c) We have

$$\tilde{\beta}_v - 1 \sim_p \Delta \frac{\int_0^T r(V_t) dV_t}{\int_0^T r(V_t)(V_t - \overline{V}_T) dt},$$
$$\tilde{\beta}_x - 1 \sim_p \Delta \frac{\int_0^T r(V_t) dV_t + (1/6) \int_0^T r'(V_t) d[V]_t}{\int_0^T r(V_t)(V_t - \overline{V}_T) dt},$$
$$\tilde{\beta}_y - 1 \sim_p \Delta \frac{\int_0^T r(V_t) dV_t + (1/6) \int_0^T r'(V_t) d[V]_t - (2\delta/\Delta^2) \int_0^T r'(V_t) V_t^2 dt}{\int_0^T r(V_t)(V_t - \overline{V}_T) dt}$$

and

$$\tilde{\beta}_z^{(k)} - 1 \sim_p (\tilde{\beta}_z - 1) + k(\tilde{\beta}_v - 1).$$

We note that if $r(z) = z$, then $\tilde{\beta}_z^{(k)}$ becomes the OLS estimator $\hat{\beta}_z^{(k)}$ in Section 3. Proposition 4.1 (a) and (b) are generalizations of Lemma 3.1 (a) and (b), respectively. Similarly, Proposition 4.1 (c) is a generalization of Proposition 3.2. Moreover, one may want to adjust the sample mean of z_i in the IV estimation by using $r(z_{i-k} - \overline{z}_N)$ as an instrument for $\tilde{\beta}_z^{(k)}$ instead of $r(z_{i-k})$. The following corollary provides the corresponding asymptotic approximations of the IV estimators $\tilde{\beta}_z^{(k)}$.

Corollary 4.2. *Let the conditions in Proposition 4.1 holds. If $r(z_{i-k} - \overline{z}_N)$ is used as an instrument for $\tilde{\beta}_z^{(k)}$ instead of $r(z_{i-k})$, then the results in Proposition 4.1 (c) hold with $r(V_t - \overline{V}_T)$ and $r'(V_t - \overline{V}_T)$ in place of $r(V_t)$ and $r'(V_t)$, respectively.*

Now we develop the large T asymptotics of $\tilde{\beta}_z$. To effectively control the fat tails in V, we impose the following conditions on r.

Assumption 4.2. *The function $r : \mathcal{D} \to \mathbb{R}$ satisfies that $\mathbb{E}[r(V_t)], \mathbb{E}[r(V_t)V_t], \mathbb{E}[r'(V_t)V_t^2]$, and $\mathbb{E}[r'(V_t)\sigma^2(V_t)]$ are all bounded, and $\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t] \neq 0$.*

Assumption 4.2 provides simple sufficient conditions to ensure that the IV estimator $\tilde{\beta}_z$ has a constant limit involving parameters in μ and σ^2. An example of r satisfying Assumption 4.2 is $r(v) = \arctan(v)$. Clearly, r is monotonically increasing, bounded and continuously differentiable with $r'(v) = 1/(1 + v^2)$. Therefore, Assumption 4.2 holds if $\mathbb{E}|V_t| < \infty$ and $\mathbb{E}[r'(V_t)\sigma^2(V_t)] < \infty$. When $\mathcal{D} = (0, \infty)$ and $r(v) = \arctan(v)$, we have $\mathbb{E}[r'(V_t)\sigma^2(V_t)] < \infty$ as long as $\sigma^2(v)/v^3 = O(v^\epsilon)$ as $v \to \infty$ for some $\epsilon > 0$.

Theorem 4.3. *Let Assumptions 3.1-3.5 and 4.1-4.2 hold. If* $\mathbb{E}|V_t| < \infty$*, then*

$$\tilde{\beta}_v - 1 \sim_p -\Delta \frac{1}{2} \frac{\mathbb{E}[r'(V_t)\sigma^2(V_t)]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]},$$
$$\tilde{\beta}_x - 1 \sim_p -\Delta \frac{1}{3} \frac{\mathbb{E}[r'(V_t)\sigma^2(V_t)]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]},$$
$$\tilde{\beta}_y - 1 \sim_p -\Delta \frac{1}{3} \frac{\mathbb{E}[r'(V_t)\sigma^2(V_t)]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]} - \frac{\delta}{\Delta} \frac{2\mathbb{E}[r'(V_t)V_t^2]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]}.$$

As well expected, $\tilde{\beta}_z - 1$ has a well defined constant limit under the moment conditions in Assumption 4.2. For a given parametric diffusion model, we may explicitly compute the limit of $\tilde{\beta}_z - 1$. As an example, we consider a stationary diffusion V defined on $\mathcal{D} = (0, \infty)$ having a linear drift

$$dV_t = \kappa(\mu - V_t)dt + \sigma(V_t)dW_t \tag{4.2}$$

with $\mathbb{E}(V_t) = \mu$ and $\sigma^2(v)/v^2 = O(1)$ as $v \to \infty$.

Corollary 4.4. *Let Assumptions 3.1-3.5 and 4.1-4.2 hold for* V *in* (4.2)*. Then*

$$\frac{\mathbb{E}[r'(V_t)\sigma^2(V_t)]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]} = 2\kappa \quad \text{and} \quad \frac{1}{\Delta}(\tilde{\beta}_v - 1) \to_p -\kappa.$$

regardless of the finiteness of $\mathbb{E}(V_t^2)$*. Moreover, if* V *in* (4.2) *is a GARCH diffusion with* $\sigma^2(v) = \sigma^2 v^2$*, then*

$$\frac{\mathbb{E}[r'(V_t)V_t^2]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]} = 2\frac{\kappa}{\sigma^2}$$

and

$$\tilde{\beta}_x - 1 \sim_p -\Delta \frac{2}{3}\kappa, \qquad \tilde{\beta}_y - 1 \sim_p -\Delta \frac{2}{3}\kappa - 4\frac{\delta}{\Delta}\frac{\kappa}{\sigma^2}$$

regardless of the finiteness of $\mathbb{E}(V_t^2)$*.*

Corollary 4.4 implies that the IV estimates $\tilde{\beta}_z$ do not depend on the finiteness of $\mathbb{E}(V_t^2)$, and are equivalent to those of the OLS estimates $\hat{\beta}_z$ in (3.7), which are obtained under $\mathbb{E}(V_t^2) < \infty$. In contrast, the OLS estimates $\hat{\beta}_z$ have different limits depending upon $\mathbb{E}(V_t^2) < \infty$ holds or not (see the discussions in Remark 3.3 (c) and (d)). Therefore, we may say that the instrumental variable approach can effectively control the fat tails as long as r is appropriately chosen.

If the transformation r satisfies some additional integrability conditions, we may obtain the asymptotic normality of the IV estimator. For the asymptotic normality, we may use the asymptotics of the centered additive functionals (see, e.g., Mandl (1968); van der Vaart and van Zanten (2005)) so that we have

$$\sqrt{T}\left(\frac{1}{T}\int_0^T (r'\sigma^2)(V_t)dt - \mathbb{E}[r'(V_t)\sigma^2(V_t)]\right) \to_d \mathbb{N}(0, \Sigma_r), \tag{4.3}$$

provided that the asymptotic variance

$$\Sigma_r = 4\left(\int_{\mathcal{D}} m(v)dv\right)\left[\int_{\mathcal{D}}\left(\int_{\underline{v}}^{v}\left\{(r'\sigma^2)(v) - \mathbb{E}[r'(V_t)\sigma^2(V_t)]\right\}\pi(u)du\right)^2 ds(v)\right]$$

is finite, where m, π and s are the speed density, time invariant distribution and scale function, respectively. Therefore, if r is appropriately chosen such that $\Sigma_r < \infty$, we may deduce from Proposition 4.1 (c), Assumption 4.2 and (4.3) that

$$\sqrt{T}\left(\tilde{\beta}_v - 1 + \frac{\Delta}{2}\frac{\mathbb{E}[r'(V_t)\sigma^2(V_t)]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]}\right) \to_d (2\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t])^{-1}\,\mathbb{N}(0, \Sigma_r),$$
$$\sqrt{T}\left(\tilde{\beta}_x - 1 + \frac{\Delta}{3}\frac{\mathbb{E}[r'(V_t)\sigma^2(V_t)]}{\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t]}\right) \to_d (3\mathbb{E}[r(V_t)V_t] - \mathbb{E}[r(V_t)]\mathbb{E}[V_t])^{-1}\,\mathbb{N}(0, \Sigma_r),$$

and $\tilde{\beta}_y - 1$ has the same asymptotic distribution as $\tilde{\beta}_x - 1$ if $\delta/\Delta^2 \to 0$. On the other hand, if r does not satisfy $\Sigma_r < \infty$, then (4.3) does not hold, and the limit distributions of $\tilde{\beta}_z - 1$ are not Gaussian (see Theorem 3.6 of Kim and Park (2017)).

Heuristically, we may consider the Cauchy estimator by using $r(z - \overline{z}_N)$, where $r(z) = \text{sign}(z)$, as an instrument in $\tilde{\beta}_z$. Clearly, r is not differentiable, and hence, our results Proposition 4.1 and Theorem 4.3 are not directly applicable. By the standard approximation method with Tanaka's formula, however, we may obtain the asymptotics of the Cauchy estimator. Given Proposition 4.1 (c), we conjecture that

$$\tilde{\beta}_v - 1 \sim_p \Delta \frac{\int_0^T \operatorname{sign}(V_t - \overline{V}_T) dV_t}{\int_0^T |V_t - \overline{V}_T| dt},$$
$$\tilde{\beta}_x - 1 \sim_p \left(\tilde{\beta}_v - 1\right) + \frac{\Delta}{3} \frac{\sigma^2(\overline{V}_T) L_V(T, \overline{V}_T)}{\int_0^T |V_t - \overline{V}_T| dt},$$
$$\tilde{\beta}_y - 1 \sim_p \left(\tilde{\beta}_x - 1\right) - \frac{4\delta}{\Delta} \frac{(\overline{V}_T)^2 L_V(T, \overline{V}_T)}{\int_0^T |V_t - \overline{V}_T| dt},$$

where $L_V(\cdot, v)$ is the local time of V at $v \in \mathcal{D}$, defined as $L_V(T, v) = \lim_{\epsilon \to 0} (2\epsilon)^{-1} \int_0^T 1\{|V_t - v| < \epsilon\} dt$. The large T asymptotics then follow immediately from the law of large numbers, and they are given by

$$\tilde{\beta}_v - 1 \sim_p -\Delta \frac{\sigma^2(\mu)\pi(\mu)}{\mathbb{E}|V_t - \mu|}, \quad \tilde{\beta}_x - 1 \sim_p -\Delta \frac{2\sigma^2(\mu)\pi(\mu)}{3\mathbb{E}|V_t - \mu|}, \quad \tilde{\beta}_y - 1 \sim_p (\tilde{\beta}_x - 1) - \frac{4\delta}{\Delta} \frac{\mu^2 \pi(\mu)}{\mathbb{E}|V_t - \mu|},$$

since $\overline{V}_T \to_p \mathbb{E}[V_t] = \mu$, $T^{-1} L_V(T, \overline{V}_T) \to_p \pi(\mu)$, $T^{-1} \int_0^T |V_t - \overline{V}_T| dt \to_p \mathbb{E}|V_t - \mu|$ and

$$\frac{1}{T} \int_0^T \operatorname{sign}(V_t - \overline{V}_T) dV_t \to_p -\sigma^2(\mu)\pi(\mu).$$

If V is a GARCH diffusion, it can be shown as in Corollary 4.4 that

$$\frac{\sigma^2(\mu)\pi(\mu)}{\mathbb{E}|V_t - \mu|} = \kappa, \qquad \frac{\mu^2 \pi(\mu)}{\mathbb{E}|V_t - \mu|} = \frac{\kappa}{\sigma^2}.$$

Therefore, we conjecture that Corollary 4.4 holds even when the Cauchy estimator is used. We will formally analyze the asymptotics, under a fixed Δ, of the Cauchy estimator in Section 5.2.

4.2. IV Estimator $\breve{\beta}_z^{(k)}$ with a Lagged Instrument

In the GARCH diffusion case, Corollary 4.4 means that the proposed IV estimator converges to the object of interest when one uses the spot variance while one gets a bias estimation when one uses integrated or realized variance. The reason is that the integrated and realized variances are ARMA(1,1) processes and therefore this IV estimator converges to the first order autocorrelation (when the second moment of these variables are bounded). As mentionned above, a solution to this problem is to consider the multi-period moment restriction (2.11), which in turn corresponds to consider a lagged instrument in the estimation of $\beta_z^{(k)}$ in (2.3). More precisely, in this subsection we study the estimator $\breve{\beta}_z^{(k)}$ defined as

$$\breve{\beta}_z^{(k)} = \frac{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i+1} - \overline{z}_N)}{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - \overline{z}_N)}.$$

In other words, the IV estimator $\tilde{\beta}_z^{(k)}$ studied in the previous subsection uses $r(z_{i-k})$ as an instrument for $(z_{i-k} - \overline{z}_N)$, whereas $\breve{\beta}_z^{(k)}$ employs $r(z_{i-k-1})$ as an instrument for the same object $(z_{i-k} - \overline{z}_N)$.

For the asymptotics, we write

$$\begin{aligned}\breve{\beta}_z^{(k)} - 1 &= \frac{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i+1} - z_{i-k})}{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - \overline{z}_N)} \\ &= \frac{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i+1} - z_{i-k-1})}{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - \overline{z}_N)} - \frac{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - z_{i-k-1})}{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - \overline{z}_N)} \\ &\equiv \phi_z^{(k)} - \psi_z^{(k)}. \end{aligned} \tag{4.4}$$

For the denominator of $\phi_z^{(k)}$ and $\psi_z^{(k)}$ with a fixed $k \geq 0$, we may show that

$$\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - \overline{z}_N)\Delta \sim_p \sum_{i=k+2}^{N-1} r(z_{i-k})(z_{i-k} - \overline{z}_N)\Delta \sim_p \int_0^T r(V_t)(V_t - \overline{V}_T)dt$$

as long as δ/Δ and Δ are sufficiently small. We then may deduce from Proposition 4.1 with (4.1) that

$$\begin{aligned}\phi_z^{(k)} &\sim_p \Delta \frac{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i+1} - z_{i-k-1})}{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k-1} - \overline{z}_N)\Delta} \sim_p \tilde{\beta}_z^{(1+k)} - 1, \\ \psi_z^{(k)} &\sim_p \Delta \frac{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k} - z_{i-k-1})}{\sum_{i=k+2}^{N-1} r(z_{i-k-1})(z_{i-k-1} - \overline{z}_N)\Delta} \sim_p \tilde{\beta}_z^{(0)} - 1\end{aligned}$$

as long as δ/Δ and Δ are sufficiently small. We formally have

Theorem 4.5. *Let Assumptions 3.1-3.5 and 4.1 hold. For $z = v, x, y$, we have*

$$\check{\beta}_z^{(k)} - 1 \sim_p (\tilde{\beta}_z^{(1+k)} - 1) - (\tilde{\beta}_z^{(0)} - 1) \sim_p (1+k)(\tilde{\beta}_v - 1).$$

Unlike $\hat{\beta}_z$ and $\tilde{\beta}_z$, the limits of $\check{\beta}_z - 1$ are given by $\tilde{\beta}_v - 1$ regardless of $z = v, x, y$. Consequently, if V is a linear drift diffusion in (4.2), then $(\check{\beta}_z - 1)/\Delta \to_p -\kappa$ for all $z = v, x, y$, rather than the limits of $\tilde{\beta}_z$ in Corollary 4.4. Therefore, we may say that the IV estimator $\check{\beta}_z$ is a consistent estimator for the mean reversion parameter κ of linear drift diffusions, and is robust to not only fat tails in V but also errors $(v_i - x_i)$ and $(v_i - y_i)$ in, respectively, the integrated variance and realized variance.

For the linear transformation $r(z) = z$, we may easily see that $\tilde{\beta}_z = \hat{\beta}_z$. In this case, the IV estimator $\check{\beta}_z$ becomes a simple IV estimator with an instrument z_{i-1} for $(z_i - \overline{z}_N)$, and it follows from Theorem 4.5 that

$$\check{\beta}_z^{(k)} - 1 \sim_p (1+k)(\hat{\beta}_v - 1) \tag{4.5}$$

for $z = v, x, y$. If Assumption 3.7 holds in addition to the conditions in Theorem 4.5, then (4.5) becomes $\check{\beta}_z - 1 \sim_d -\Delta(p_T/q_T)(P/(2S))$ for $z = v, x, y$ by Theorem 3.4. On the other hand, Assumption 4.2 holds for $r(z) = z$ if and only if $\mathbb{E}(V_t^2)$ and $\mathbb{E}(\sigma^2(V_t))$ are finite. Consequently, when $r(z) = z$, we have $(\check{\beta}_z - 1)/\Delta \not\to_p -\kappa$ for a linear drift diffusion (4.2) satisfying either $\mathbb{E}(V_t^2) = \infty$ or $\mathbb{E}(\sigma^2(V_t)) = \infty$. For instance, a GARCH diffusion satisfies $(\check{\beta}_z - 1)/\Delta \to_p -\kappa$ if $\mathbb{E}(V_t^2) < \infty$ with $\sigma^2 < 2\kappa$, whereas $(\check{\beta}_z - 1)/\Delta \to_p -\sigma^2/2$ if $\mathbb{E}(V_t^2) = \infty$ with $2\kappa < \sigma^2$.

As a conclusion of this section, let us remark that there is a large literature considering autoregressions in discrete time models and allowing for heavy tails. In particular, Hill (2015) and Hill and Prokhorov (2016) propose a robust generalized empirical likelihood (GEL) method for estimation and inference of an autoregression that may have a heavy tailed heteroscedastic error. We expect that the GEL estimator can also be robust to fat tails in continuous time models. However, it is questionable whether the GEL estimator can be robust to the non-Markovianity of (x_i) and (y_i) in our framework. We leave the asymptotic properties of GEL methods in volatility regression for future research.

4.3. Asymptotic Negligibility of Errors in Realized Variance

In the previous subsections, we analyze the asymptotic behaviors of the IV estimators under the assumption that each of V and D is independent of W^P. In reality, however, it is widely believed that there exists the leverage effect, which corresponds to a negative correlation between past returns and future volatility. As an extension of our previous results, we allow arbitrary dependences among V, D and W^P, and analyze the asymptotic negligibility of the errors in the realized variance.

Assumption 4.3. *(a) For ω in Assumption 3.1, $(\delta/\Delta^2)T_V(\omega^6)T\log^3(T/\delta) \to_p 0$, and (b) $\Delta T_D^2 \to_p 0$.*

It can be seen from the primary asymptotics in Proposition 4.1 (c) that the impact of errors in (y_i) may become asymptotically negligible as long as $\delta/\Delta^2 \to 0$ sufficiently quickly. Assumption 4.3 (a) is a sufficient condition for the asymptotic negligibility of the error, and requires faster rate of convergence $\delta \to 0$ than Assumption 3.3. On the other hand, Assumption 4.3 (b) has a similar role to Assumption 3.4, and provides a sufficient condition for the asymptotic negligibility of the drift part (D_t) in the IV estimation with (y_i).

Proposition 4.6. *Under Assumptions 3.1-3.2, 4.1, and 4.3,*

$$\tilde{\beta}_y^{(k)} - 1 \sim_p \tilde{\beta}_x^{(k)} - 1 \quad \textit{and} \quad \check{\beta}_y^{(k)} - 1 \sim_p \check{\beta}_x^{(k)} - 1.$$

Unlike Proposition 4.1 (c), $\tilde{\beta}_y^{(k)}$ becomes asymptotically equivalent to $\tilde{\beta}_x^{(k)}$ regardless of the presence of leverage effects, especially, when $\delta/\Delta^2 \to 0$ sufficiently quickly. It is also true that, under the conditions in Proposition 4.6, $\hat{\beta}_y^{(k)} - 1 \sim_p \hat{\beta}_x^{(k)} - 1$ since $\hat{\beta}_z^{(k)}$ is a special case of $\tilde{\beta}_z^{(k)}$ with $r(z) = z$.

5. Volatility Regressions Under Fixed Δ

In this section, we consider a volatility regression under fixed Δ, and provide an IV based inference procedure.

5.1. *IV Based Inference*

Motivated by the multi-period moment restriction (2.11) for GARCH diffusions, we consider a stationary time series (z_i), defined on $\mathcal{D}_z = (\underline{z}, \overline{z})$, satisfy $\mathbb{E}|z_i| < \infty$ and

$$\mathbb{E}[(z_{i+1} - \mu_z) - \beta_z(z_i - \mu_z)|\mathcal{G}_{i-1}] = 0, \tag{5.1}$$

where $\mathcal{G}_{i-1}$ is the sigma-field generated by $(z_{i-k}, k = 1, 2, \cdots)$. For a GARCH diffusion (2.7), $\mu_z = \mu$ and $\beta_z = \exp(-\kappa\Delta)$ for $z = v, x$ as well as for $z = y$ if $D_t = 0$ almost surely.

An equivalent representation of (z_i) satisfying the multi-period moment restriction (5.1) is that

$$z_{i+1} = \alpha_z + \beta_z z_i + u_{i+1} \quad \text{with} \quad \alpha_z = \mu_z(1 - \beta_z), \quad \mathbb{E}(u_{i+1}|\mathcal{G}_{i-1}) = 0. \tag{5.2}$$

Clearly, (z_i) can be non-autoregressive since we allow $\mathbb{E}(u_{i+1}|z_i) \neq 0$.

Now we propose an IV based inference procedure for β_z in (5.2). We consider a measurable function $r : \mathcal{D}_z^l \to \mathbb{R}$ for some $1 \leq l < \infty$, and construct the IV estimator as

$$\breve{\beta}_z = \frac{\sum_{i=l+1}^{N-1} r(Z_{i-1,i-l})(z_{i+1} - \overline{z}_N)}{\sum_{i=l+1}^{N-1} r(Z_{i-1,i-l})(z_i - \overline{z}_N)}, \quad \text{where} \quad Z_{i-1,i-l} = (z_{i-1}, z_{i-2}, \cdots, z_{i-l})'.$$

Assumption 5.1. *There exists a measurable function $r : \mathcal{D}_z^l \to \mathbb{R}$ for some $1 \leq l < \infty$ such that (a) $\mathbb{E}|r(Z_{i-1,i-l})(z_i - \mu_z)| < \infty$ with $\mathbb{E}[r(Z_{i-1,i-l})(z_i - \mu_z)] \neq 0$, and (b) $\mathbb{E}(r^2(Z_{i-1,i-l})z_{i+1}^2), \mathbb{E}(r^2(Z_{i-1,i-l})z_i^2) < \infty$.*

Assumption 5.1 is quite general and includes the situations where $\mathbb{E}(z_i^2) = \infty$. If r is bounded, then $\mathbb{E}[r(Z_{i-1,i-l})(z_i - \mu_z)]$ is well defined since $\mathbb{E}|z_i| < \infty$, and hence, Assumption 5.1 (a) is always satisfied as long as r is bounded such that the identification condition $\mathbb{E}[r(Z_{i-1,i-l})(z_i - \mu_z)] \neq 0$ is met. However, Assumption 5.1 (b) imposes more restrictions on the tail behaviors of the function r.

Under Assumption 5.1 with some regularity conditions, we have

$$\frac{1}{N}\sum_{i=l+1}^{N-1} r(Z_{i-1,i-l})u_{i+1} \to_p \mathbb{E}[r(Z_{i-1,i-l})u_{i+1}] = 0, \tag{5.3}$$

$$\frac{1}{N}\sum_{i=l+1}^{N-1} r(Z_{i-1,i-l})(z_i-\mu_z) \to_p \mathbb{E}[r(Z_{i-1,i-l})(z_i-\mu_z)], \tag{5.4}$$

$$\frac{1}{N}\sum_{i=l+1}^{N-1} r^2(Z_{i-1,i-l})u_{i+1}^2 \to_p \mathbb{E}[r^2(Z_{i-1,i-l})u_{i+1}^2], \tag{5.5}$$

$$\frac{1}{N^{1/2}}\sum_{i=l+1}^{N-1} r(Z_{i-1,i-l})u_{i+1} \to_d \mathbb{N}(0, \mathbb{E}[r^2(Z_{i-1,i-l})u_{i+1}^2]) \tag{5.6}$$

due to the LLN and the martingale CLT.

Theorem 5.1. *(a) If Assumption 5.1 (a) and (5.3)-(5.4) hold, then $\check{\beta}_z \to_p \beta_z$. (b) If Assumption 5.1 and (5.3)-(5.6) hold, then*

$$\sqrt{N}(\check{\beta}_z - \beta_z) \to_d \mathbb{N}(0, \Sigma_l(r)), \quad \textit{where} \quad \Sigma_l(r) = \frac{\mathbb{E}[r^2(Z_{i-1,i-l})u_{i+1}^2]}{\mathbb{E}[r(Z_{i-1,i-l})(z_i-\mu_z)]^2},$$

and $\check{\Sigma}_l(r) \to_p \Sigma_l(r)$, where

$$\check{\Sigma}_l(r) = N\frac{\sum_{l+1}^{N-1} r^2(Z_{i-1,i-l})\check{u}_{i+1}^2}{\left(\sum_{l+1}^{N-1} r(Z_{i-1,i-l})(z_i-\bar{z}_N)\right)^2} \quad \textit{with} \quad \check{u}_{i+1} = (z_{i+1}-\bar{z}_N) - \check{\beta}_z(z_i-\bar{z}_N).$$

Theorem 5.1 (a) implies that $\check{\beta}_z \to_p \beta_z$ as long as r is bounded. For a GARCH diffusion, in particular, $\beta_z = \exp(-\kappa\Delta)$ for $z = v, x, y$, and hence, the IV estimator $\check{\beta}_z$ for $z = v, x, y$ with instrument $r(z_{i-1} - \bar{z}_N)$ has the same limit $\exp(-\kappa\Delta)$ as long as r is bounded. So the Cauchy estimator with $r(z) = \text{sign}(z)$ becomes consistent for $\exp(-\kappa\Delta)$.

If we further impose some restrictions on the tail behaviors of r, then the IV estimator $\check{\beta}_z$ is asymptotically normal as shown in Theorem 5.1 (b). In GARCH diffusion example, Assumption 5.1 (b) holds for $z = v, x$ if $r(z) \sim \text{sign}(z)|z|^c$ with $c < -1/2$. Clearly, $r(z) = \text{sign}(z)$ does not satisfy Assumption 5.1 (b) unless z has finite variance, and hence, the Cauchy estimator is not asymptotically normal, though it is consistent, under $\mathbb{E}(V_t^2) = \infty$.

5.2. *Fixed Δ and Small Δ Asymptotics for GARCH Diffusion*

In Section 4.2, we obtained general asymptotics of the IV estimator $\check{\beta}_z$, which is robust to fat tails as well as errors in observed variance measures, in particular, under the assumption that $\Delta \to 0$. In our asymptotics, the main motivation of introducing the small Δ assumption is to effectively handle general variance processes V having potentially unbounded moments. In practice, however, the volatility measure, (x_i) or (y_i), is often computed on a daily basis, and Δ is commonly fixed to a length of day. Under the fixed Δ, one may be interested in the quality of our approximation based on $\Delta \to 0$.

To see the usefulness of our asymptotics under $\Delta \to 0$, we again consider a GARCH diffusion (2.7). It then follows from Theorem 5.1 that the fixed Δ asymptotics are approximately equivalent to the asymptotics obtained under $\Delta \to 0$ since $\exp(-\kappa\Delta) = 1 - \kappa\Delta + o(\Delta)$, and the leading term $1 - \kappa\Delta$ is equivalent to the limit of $\check{\beta}_z$ obtained under $\Delta \to 0$ in Theorem 4.5. We also note that our asymptotics of $\hat{\beta}_z$ in Section 3 provide the same results as those derived by Andersen et al. (2004) when the required moments are satisfied. Therefore, we may conclude that our asymptotics, obtained under $\Delta \to 0$, provide a useful asymptotic approximation at least for some popular models.

Given the asymptotic assumption of $\Delta \to 0$ as well as continuity of the sample path of V, two consecutive measures z_{i+1} and z_i are supposed to be very close for $z = v, x, y$. Consequently, we always have unit roots in $z = v, x, y$, and $\hat{\beta}_z, \tilde{\beta}_z, \check{\beta}_z \to_p 1$ as long as $\Delta \to 0$ sufficiently quickly, as derived in the previous sections. Indeed, there are many evidences supporting the unit-root like behavior in volatility regressions. In the empirical studies in Hansen and Lunde (2014), for instance, volatility regressions at daily frequency are considered for 29 assets in the Dow Jones industrial average. The range of parameter estimates for the coefficient of the first order autoregression with realized variances are $[0.611, 0.887]$ for the OLS and $[0.895, 1.037]$ for the IV with a lagged realized variance as instrument. They also find that the volatility processes are highly persistent, and they fail to reject the unit root hypothesis at the 1% level for some volatility processes.

6. Simulations

In this section, we study by simulation the behavior of the Hill tail index estimator as well as the OLS and some IV estimators. For our simulations, we use the GARCH diffusion (2.7) with three sets of parameters. The first one is $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0207)$ which implies that the corresponding V_t has a finite second moment since $\psi_0 = \sigma_0^2/(2\kappa_0) = 0.296 < 1$. This set of parameters was used by Andersen and Bollerslev (1998) as implied from the (weak) daily GARCH(1,1) model estimates for the DM/dollar from 1987 through 1992 using the temporal aggregation results of Drost and Nijman (1993) and Drost and Werker (1996); the same parameters were used by Andersen et al. (2004).

To consider a process with an unbounded variance, we consider two other sets of parameters by keeping the same κ_0 and μ_0, while we multiply σ_0^2 by 4 and 16, corresponding to $\psi_0 = 1.183$ and $\psi_0 = 4.732$, that is $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0828)$ and $(0.0350, 0.6360, 0.3312)$. Clearly, the third model has thicker tails than the second one.

The simulation samples are generated by the Euler discretization at 10 seconds for $T = 250, 500, 1000$ days corresponding to 1, 2 and 4 years. We assume that the market is open 24 hours. For each day ($\Delta = 1$), we set the daily spot variance as the spot variance at the end of the day, while we compute the integrated variance by the numerical integration of the simulated spot variance process at 10 seconds. As for the realized variance, we analyze the frequency effects by considering three different frequencies: 10 minutes ($\delta/\Delta = 1/144$), 5 minutes ($\delta/\Delta = 1/288$) and 1 minute ($\delta/\Delta = 1/1440$). For each sample, we get rid of the first five days to reduce the effect of the initial value, and we do 10,000 replications.

6.1. Tail Index

We start by studying the properties of the Hill estimator by estimating the tail index of the returns, spot and integrated volatility of the GARCH diffusion model. In an important contribution, Nelson (1990) proved that when the length of time between observations goes to zero, the returns follow (up to a scaling factor) a Student distribution with degree of freedom $\nu_0 = 2 + 4\kappa_0/\sigma_0^2 = 2 + 2/\psi_0$. Consequently, the tail index of the return is ν_0 which equals 8.75 for Model 1, 3.69 for Model 2 and 2.43 for

Model 3.

Likewise, the stationary distribution of any stationary scalar diffusion process is well known and proportional to the speed density function $m(\cdot)$ defined in (2.6). One can easily show that $m(v) \sim v^{-2-1/\psi_0}$ when $v \to \infty$, implying that the tail index of the spot variance V_t equals $1 + 1/\psi_0$. Consequently, the tail index of the returns equals the double of the spot variance's tail index when the length of time between observations goes to zero.

Unfortunately we do not know the tail indexes of the integrated and realized variances. There is no general result connecting the tail of a process with the tail of the temporal aggregation version of it.

Figure 2 depicts the average estimator of the tail index of the returns, the spot variance, the integrated variance and the three realized volatility measures of the three models. The averages are computed over 10,000 replications of samples with 1,000 observations each. The top-left panel in Figure 2 depicts the tail index of daily returns. If Nelson's approximation is good, the true tail index should be 8.75 for Model 1, 3.69 for Model 2, and 2.43 for Model 3. The simulations suggest that there is negative bias in the Hill estimator, which is quite small for low values of k (we do not use the subscript n) and increases when k increases. However the order of the tails is coherent across models. The bias is maybe genuine, or because Nelson's approximation is not good for our sample frequency. The middle-left panel in Figure 2 depicts the tail index of the spot variance. The tail index should be 4.38 for Model 1, 1.85 for Model 2, and 1.22 for Model 3. There is clearly a positive bias when k is small and the bias decreases when k is moderately large. Again, the order of the tails is coherent across models with the right magnitudes. The bottom-left panel in Figure 2 depicts the tail index of the integrated variance for which we do not know the true tail index. The plots are quite similar to those of the spot variance. The right panels in Figure 2 depict the tail index of the three realized volatility measures for which we do not know the true index. The graphs are close to those of the integrated variance, especially for the bottom-right panel that corresponds to realized variance computed with 1-minute returns.

6.2. OLS and IV Estimations

We now turn to study the empirical distributions of the OLS and IV estimators. We keep the three models of the GARCH diffusion (2.7), with three sample sizes, 250, 500, and 1,000.

We start by considering the regression

$$v_{i+1} = \alpha + \beta v_i + u_{i+1}, \quad \text{with } v_i = V_{i\Delta}.$$

We will focus on the slope parameter β. It is well known that β equals $\exp(-\kappa\Delta)$ when the spot variance has a finite second moment. However, we proved in (2.10) that the same result holds when V_t is stationary and has a finite first moment. Therefore, the slope of interest is $\exp(-\kappa\Delta)$ for the three models considered in this section.

When Δ is fixed and the second moment of V_t is bounded, the OLS estimator of β is consistent. Characterizing the fixed Δ asymptotics of the OLS is difficult when the second moment of V_t is not finite. However, we may deduce from Theorem 3.4 (see also Remark 3.3 (d)) that the OLS is inconsistent. However, IV method is consistent when the Cauchy estimator is used.

The left column in Figure 3 depicts the empirical distribution of the OLS and IV estimators of the slope coefficient. The top panel deals with Model 1 for which the second moment is bounded, while the middle and bottom panels deal respectively with Models 2 and 3 for which the second moment of V_t is unbounded. The figures are coherent with the theory. For Model 1 (top-left panel), both OLS and IV estimators look consistent with better properties when the sample size increases. However, the OLS estimator presents a bias and looks inconsistent, as expected by our theory, for Model 2 (middle-left panel) and especially Model 3 (bottom-left panel) which present very fat tails. In contrast, the Cauchy estimator looks consistent for the two models, even though there is some bias that decreases when the sample sizes increases.

In practice, the spot variance process is not observed. It is therefore important to focus on feasible methods based on the observed realized variance processes. Accordingly, we consider the multi-period moment restriction (2.11) which is always valid for the spot and integrated variances, and is valid for the realized variance when the drift

D_t is zero as in our simulations.[7] Consequently, we consider the moment condition

$$\mathbb{E}[r(z_{i-2})(z_i - \alpha - \beta z_{i-1})] = 0,$$

where z_i is either the spot, the integrated or one of the three realized variance measures. We consider two IV estimators: the first one is $r(z_{i-2}) = z_{i-2}$ while the second one is the sign of z_{i-2} minus its empirical mean, that is the Cauchy estimator. For the second and third models, the first IV estimator with $r(z_{i-2}) = z_{i-2}$ does not fulfill the restriction $\mathbb{E}[\| \, r(z_{i-2})(z_i - \alpha - \beta z_{i-1}) \, \|] < \infty$, and hence, we may deduce from (3.9) and (4.5) that it is not consistent for $\exp(-\kappa\Delta)$, even when $\Delta \to 0$. However, the corresponding estimator is consistent for the first model. The second IV estimator is the Cauchy one and leads to a consistent estimator for the three models.

The right column of Figure 3 and Figures 4-5 depict respectively the empirical distribution of the slope's estimator for the five volatility measures listed above. For all figures, the Cauchy instrument based estimator looks consistent whether $\mathbb{E}(V_t^2) < \infty$ (top panels) or not (middle and bottom panels), which is coherent with the theory. Importantly, the estimator based on observed volatility measures is consistent, which is practically more relevant. As expected, the IV estimator with $r(z_{i-2}) = z_{i-2}$ looks consistent only for Model 1 (top panels), but it looks inconsistent for Models 2 and 3 (middle and bottom panels), with a larger bias for Model 3 (bottom panels).

7. Conclusion

Fat tails are a well-known empirical fact of financial returns. Surprisingly, the realized volatility literature has ignored this fact. After showing empirically that the second moment of several realized variance measures is probably unbounded, we theoretically studied the limiting behavior of the OLS estimator of simple autoregressions of a spot, integrated and realized variances. We proved that when the second moment of the spot variance is unbounded, the OLS estimators converge to random variables. Our theory is also valid when the second moment of the spot variance is bounded. In this case, the OLS estimates converge to finite and deterministic quantities which

[7] The presence of a drift will introduce a small bias that will disappear when the length of the intra-day returns δ goes to zero.

are the same ones derived by Andersen et al. (2004) in population regressions. Our theoretical results are based on asymptotic approximations. Both the simulations and the comparison with the results in Andersen et al. (2004) when the spot variance has a finite second moment corroborate the good quality of our approach.

In order to derive more positive results, we considered a GARCH diffusion process with an unbounded second moment for the variance process and then we provided a consistent estimation method based on an instrumental variable approach where the instrument is the sign of the lagged value of the variable of interest.

There is an important question that should be addressed. It concerns the forecast that one should compute under fat tails in a non-parametric setting. Various approaches could be considered like different loss functions or nonlinear transforms of the variable of interest. This question is currently under investigation.

References

Aït-Sahalia, Y., J. Fan, and Y. Li (2013): "The leverage effect puzzle: Disentangling sources of bias at high frequency," *Journal of Financial Economics*, 109, 224–249.

Andersen, T. and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

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

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2004): "Analytical Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079–1110.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2005): "A central limit theorem for realised power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance, Festschrift for Albert Shiryaev*, ed. by Y. Kabanov and R. Lipster, Springer, 33–68.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536.

Barndorff-Nielsen, O. E. and N. Shephard (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 253–280.

——— (2003): "Realized power variation and stochastic volatility," *Bernoulli*, 9, 243–265.

——— (2004a): "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

——— (2004b): "Power and bipower variation with stochastic volatility and jumps," *Journal of Financial Econometrics*, 2, 1–37.

——— (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006): "Limit theorems for multipower variation in the presence of jumps," *Stochastic Processes and Their Applications*, 5, 796–806.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1993): *Regular Variation*, Cambridge, UK: Cambridge University Press.

Blattberg, R. and T. Sargent (1971): "Regression with Non-Gaussian Stable Disturbances: Some Sampling Results," *Econometrica*, 39, 501–510.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Brockwell, P. J. and R. A. Davis (1991): *Time Series: Theory and Methods*, Springer-Verlag New York.

Cauchy, A. (1836): "On a new formula for solvingthe problem of interpolation in a manner applicable to physical investigations," *Philosophical Magazine*, 8, 459–468.

Chan, N. H. and L. T. Tran (1989): "On the First-Order Autoregressive Process with Infinite Variance," *Econometric Theory*, 5, 354–362.

Choi, Y., S. Jacewitz, and J. Y. Park (2016): "A reexamination of stock return predictability," *Journal of Econometrics*, 192, 168–189.

Coudin, E. and J.-M. Dufour (2009): "Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form," *The Econometrics Journal*, 12, S19–S49.

——— (2017): "Finite-sample generalized confidence distributions and sign-based robust estimators in median regressions with heterogeneous dependent errors," Working Paper.

Davis, R. and T. Mikosch (1998): "The Sample Autocorrelations of Heavy-Tailed Processes with Applications to ARCH," *The Annals of Statistics*, 26, 2049–2080.

Davis, R. A. (1982): "Maximum and Minimum of One-Dimensional Diffusions," *Stochastic Processes and their Applications*, 13, 1–9.

Drost, F. C. and T. E. Nijman (1993): "Temporal Aggregation of Garch Processes," *Econometrica*, 61, 909–927.

Drost, F. C. and B. J. Werker (1996): "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31–57.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Hannan, E. J. and M. Kanter (1977): "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability*, 14, 411–415.

Hansen, P. R. and A. Lunde (2014): "Estimating the persistence and the autocorrelation function of a time series that is measured with error," *Econometric Theory*, 30, 60–93.

Hill, B. M. (1975): "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3, 1163–1174.

Hill, J. B. (2010): "On Tail Index Estimation for Dependent, Heterogeneous Data," *Econometric Theory*, 26, 1398–1436.

——— (2015): "Robust Generalized Empirical Likelihood for heavy tailed autoregressions with conditionally heteroscedastic errors," *Journal of Multivariate Analysis*, 135, 131–152.

Hill, J. B. and A. Prokhorov (2016): "GEL estimation for heavy-tailed GARCH models with robust empirical likelihood inference," *Journal of Econometrics*, 190, 18–45.

Hsing, T. (1991): "On Tail Index Estimation Using Dependent Data," *The Annals of Statistics*, 19, 1547–1569.

Jacod, J. (2008): "Asymptotic properties of realized power variations and related functionals of semimartingales," *Stochastic Processes and Their Applications*, 118, 517–559.

——— (2012): "Statistics and high frequency data," in *Statistical Methods for Stochastic Differential Equations*, ed. by M. Kessler, A. Lindner, and M. Sorensen, Chapman and Hall, 191–310.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and their Applications*, 119, 2249–2276.

Jacod, J. and P. Protter (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

Jacod, J. and M. Rosenbaum (2013): "Quarticity and other functionals of volatility: efficient estimation," *Annals of Statistics*, 41, 1462–1484.

Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, New York, New York: Springer-Verlag.

Kim, J. and J. Y. Park (2017): "Asymptotics for Recurrent Diffusions with Application to High Frequency Regression," *Journal of Econometrics*, 196, 27–54.

——— (2018): "Unit Root, Mean Reversion and Nonstationarity in Financial Time Series," Working Paper, Indiana University.

Knight, K. (1987): "Rate of Convergence of Centered Estimates of Autoregressive Parameters for Infinite Variance Autoregressions," *Journal of Time Series Analysis*, 8, 51–60.

Kurtz, T. G. and P. Protter (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *The Annals of Probability*, 19, 1035–1070.

Mancini, C. (2009): "Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps," *Scandinavian Journal of Statistics*, 36, 270–296.

Mandl, P. (1968): *Analytical Treatment of One-Dimensional Markov Processes*, Berlin, Germany: Springer.

Meddahi, N. (2001): "An Eigenfunction Approach for Volatility Modeling," CIRANO Working Paper.

——— (2003): "ARMA Representation of Integrated and Realized Variances," *Econometrics Journal*, 6, 335–356.

Meddahi, N. and E. Renault (2004): "Temporal Aggregation of Volatility Models," *Journal of Econometrics*, 119, 355–379.

Mikosch, T. and C. Starica (2000): "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1, 1) Process," *The Annals of Statistics*, 28, 1427–1451.

Mykland, P. A. and L. Zhang (2017): "Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance," *Econometrica*, 85, 197–231.

Nelson, D. B. (1990): "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7–38.

Phillips, P. C. B. (1990): "Time Series Regression with a Unit Rot and Infinite-Variance Errors," *Econometric Theory*, 6, 44–62.

Phillips, P. C. B., J. Y. Park, and Y. Chang (2004): "Nonlinear instrumental variable estimation of an autoregression," *Journal of Econometrics*, 118, 219–246.

Resnick, S. and C. Stărică (1995): "Consistency of Hill's Estimator for Dependent Data," *Journal of Applied Probability*, 32, 139–167.

——— (1998): "Tail Index Estimation for Dependent Data," *The Annals of Applied Probability*, 8, 1156–1183.

Samorodnitsky, G., S. T. Rachev, J.-R. Kurz-Kim, and S. V. Stoyanov (2007): "Asymptotic Distributions of Unbiased Linear Estimators in the Presence of Heavy-Tailed Stochastic Regressors and Residuals," *Probability and Mathematical Statistics*, 27, 275–302.

So, B. S. and D. W. Shin (1999): "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory*, 15, 165–176.

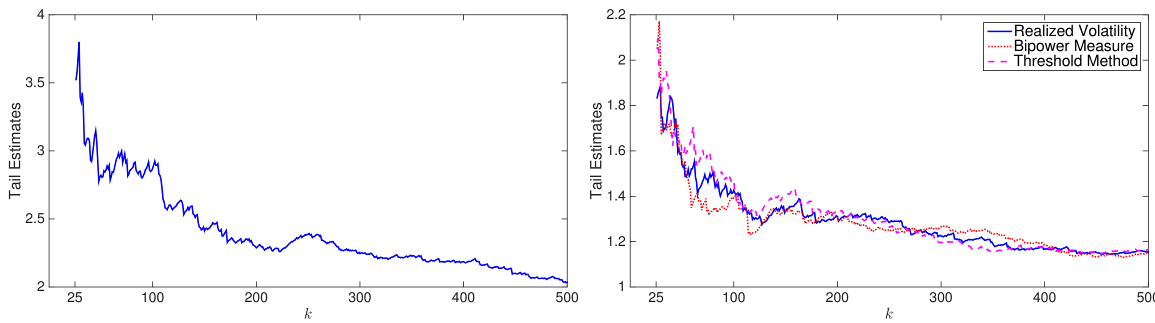
Stoker, T. M. (1993): "Smoothing Bias in Density Derivative Estimation," *Journal of the American Statistical Association*, 88, 855–863.

van der Vaart, A. and H. van Zanten (2005): "Donsker theorems for diffusions: Necessary and sufficient conditions," *Annals of Probability*, 33, 1422–1451.

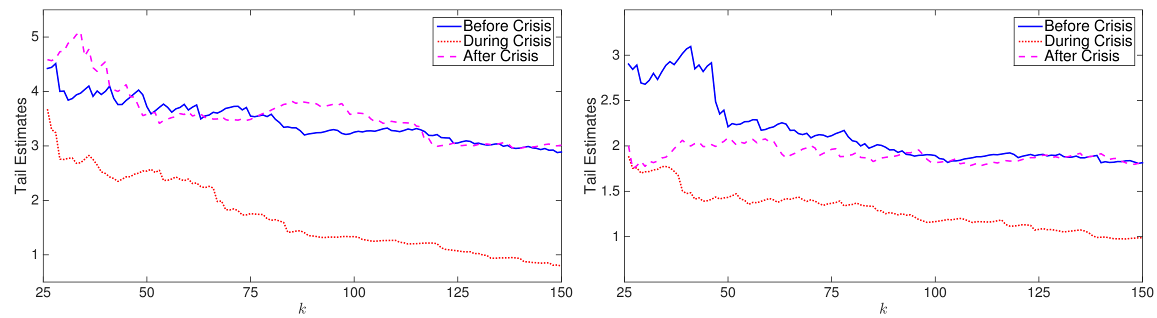
Zhang, L., P. A. Mykland, and Y. Ait-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

Fig. 1. **Estimation of Tail Indexes**

(a) Returns and volatility measures



(b) Returns and realized volatility over sub-periods



(c) Bipower and threshold volatilities over sub-periods

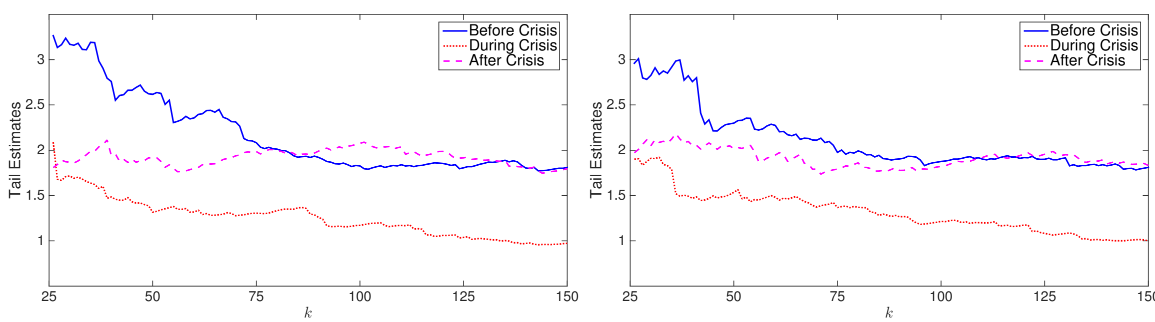


Figure 1 depicts the Hill estimator of the tail index for returns and various volatility measures of the SPDR S&P 500 ETF (SPY), from June 15, 2004 through June 13, 2014. Figure 1 (a) is based on the entire period of the SPY, whereas the full period is divided in three sub-periods: Before Crisis (June 15, 2004 through August 29, 2008), During Crisis (September 2, 2008 through May 29, 2009) and After Crisis (June 1, 2009 through June 13, 2014) in Figure 1 (b) and (c).

Figure 1 (a): The first panel depicts the tail index of the daily return (open-to-close), and the second panel depicts the tail index of the realized volatility, the bipower and the threshold volatility measure.

Figure 1 (b): The first panel depicts the tail index of the daily return (open-to-close), and the second panel depicts the tail index of the realized volatility.

Figure 1 (c): The first panel depicts the tail index of the bipower volatility measure (open-to-close), and the second panel depicts the tail index of the threshold volatility measure.

Fig. 2. **Tail indexes for returns and volatility measures of GARCH diffusions**

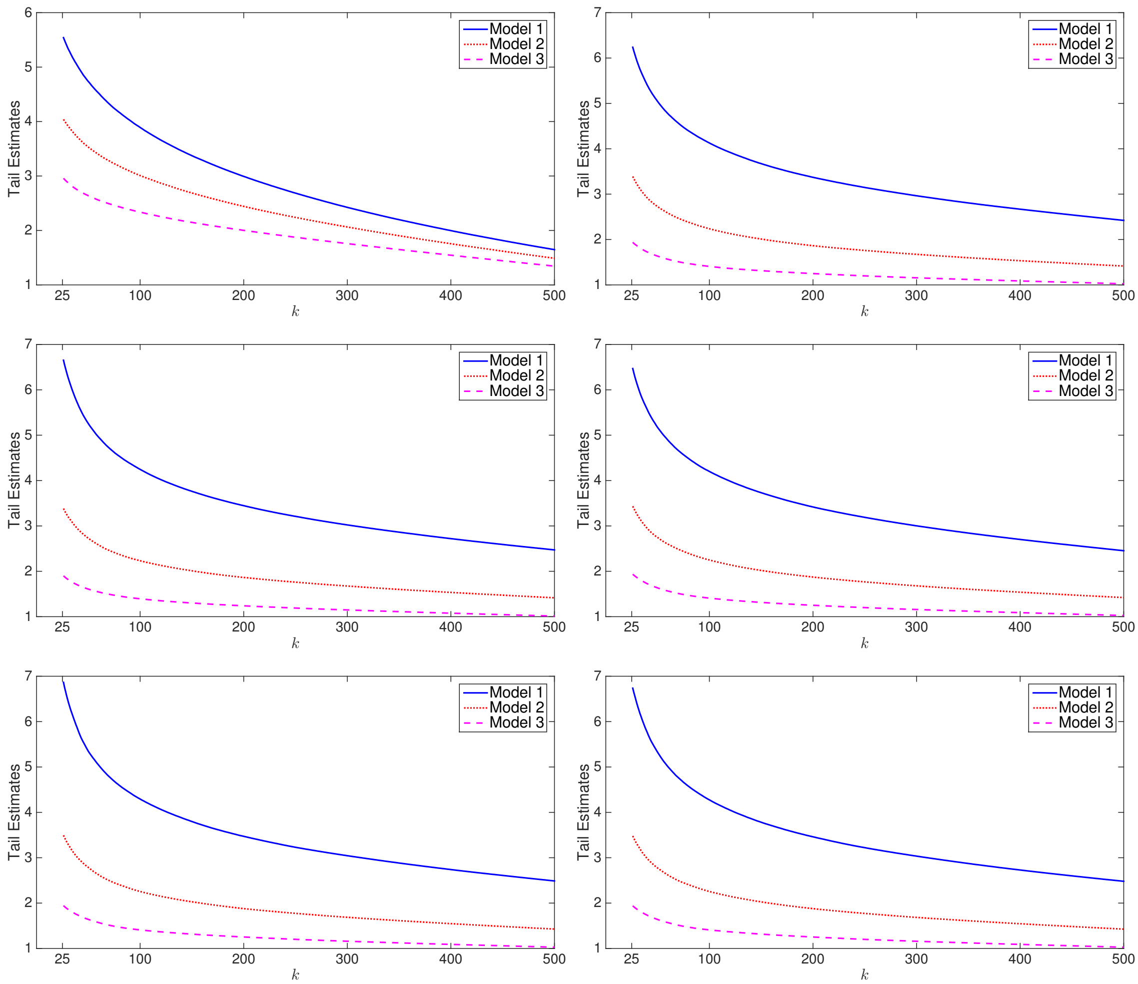


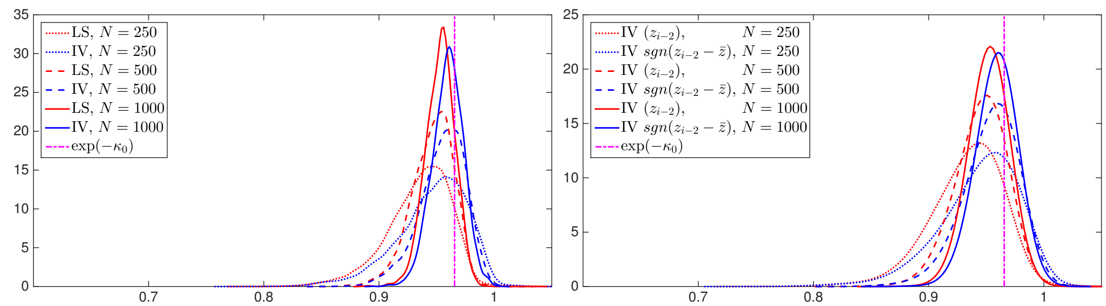
Figure 2 depicts the average estimator of the tail indexes over 10,000 simulations of a sample with 1,000 observations of the GARCH diffusion model. Three designs are considered: Model 1 corresponds to $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0207)$, while Models 2 and 3 correspond respectively to $(0.0350, 0.6360, 0.0828)$ and $(0.0350, 0.6360, 0.3312)$.

Left Column: The first panel depicts the tail index of daily returns; the second panel depicts the tail index of the daily spot volatility while the third one depicts the tail of the daily integrated volatility.

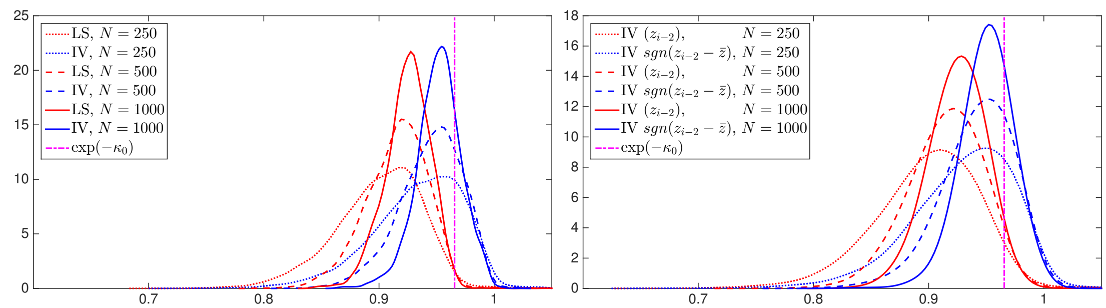
Right Column: The three panels depict the tail index of the daily realized volatility (RV) with different frequencies: the first panel with 10 minute-returns RV; the second panel with 5 minute-returns RV; the last panel with 1 minute-returns RV.

Fig. 3. **OLS and IV estimations with spot volatility**

(a) Model 1 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0207)$



(b) Model 2 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0828)$



(c) Model 3 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.3312)$

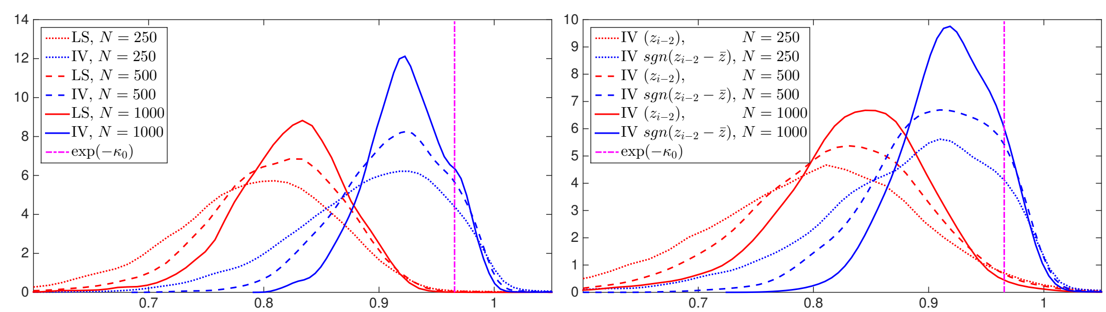


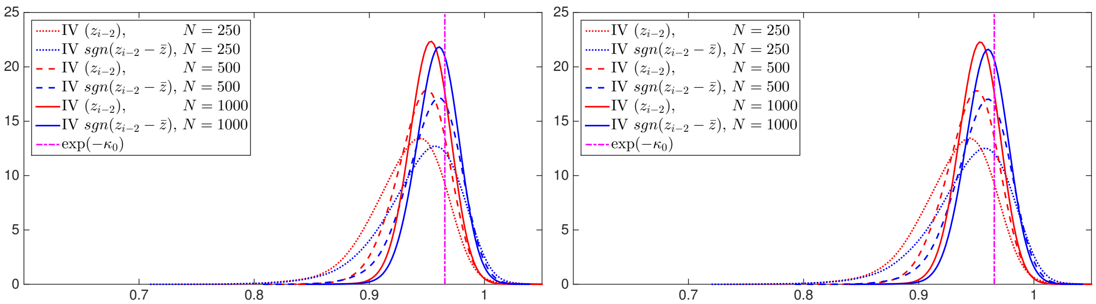
Figure 3 depicts the empirical distribution of the OLS and IV estimators for daily spot volatility of the GARCH diffusion model. The figures are based on 10,000 simulations for three different sample sizes (250, 500 and 1,000).

Left Column: The OLS and IV estimators of the autoregression of order one. The instrument of the IV estimator is the sign of the centered lagged value of the spot volatility.

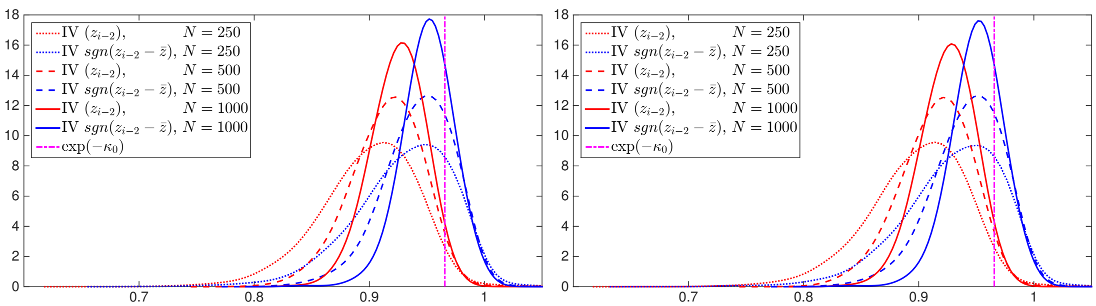
Right Column: Two IV estimators are considered for the multi-period-moment restrictions. The first instrument is the two lags of spot volatility, while the second instrument is the sign of the centered value of the first instrument.

Fig. 4. **IV estimations with integrated and 1-minute realized volatilities**

(a) Model 1 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0207)$



(b) Model 2 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0828)$



(c) Model 3 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.3312)$

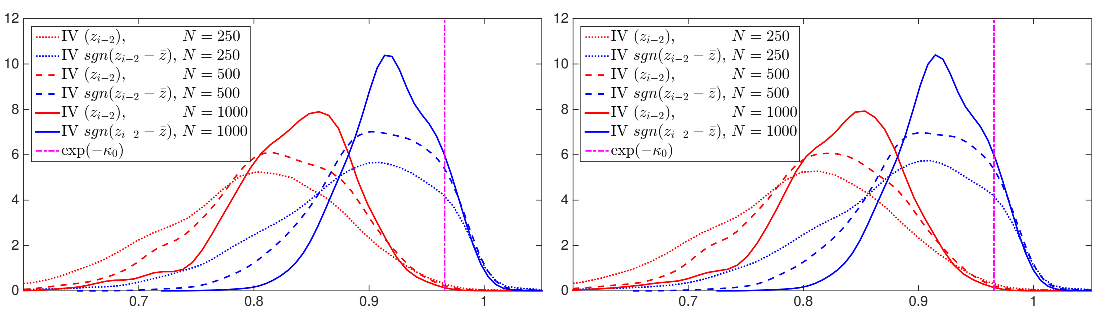
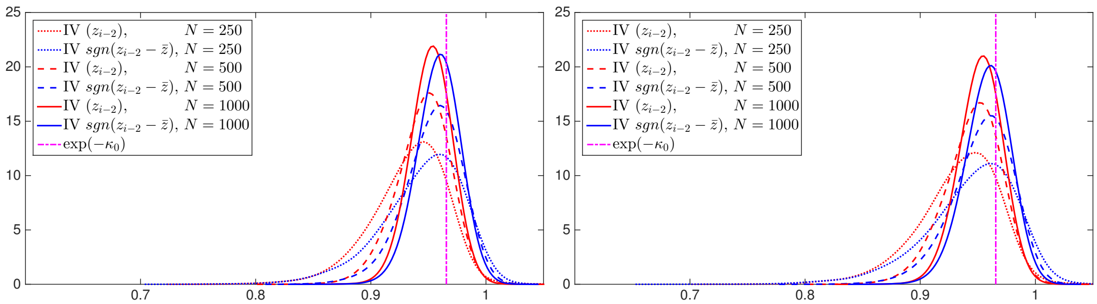


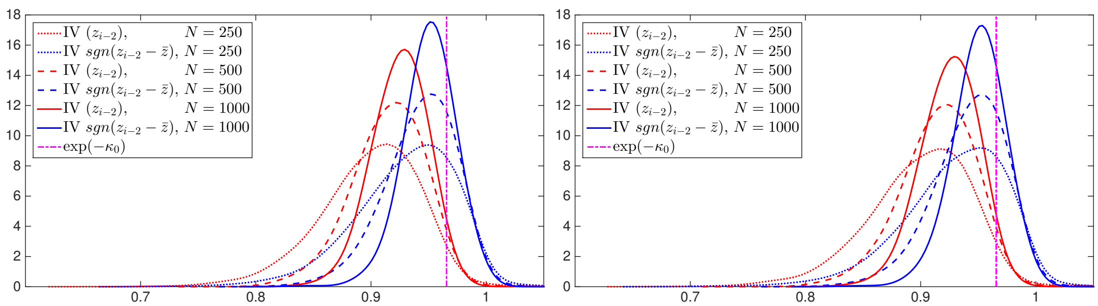
Figure 4 depicts the empirical distribution of the IV estimators for the multi-period-moment restrictions of daily spot volatility of the GARCH diffusion model. The figures are based on 10,000 simulations for three different sample sizes (250, 500 and 1,000). Two IV estimators are considered for the multi-period-moment restrictions. The first instrument is the two lags of spot volatility, while the second instrument is the sign of the centered value of the first instrument. The left and right columns are for the integrated and 1-minute realized volatilities, respectively.

Fig. 5. **IV estimations with 5-minutes and 10-minutes realized volatilities**

(a) Model 1 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0207)$



(b) Model 2 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.0828)$



(c) Model 3 $(\kappa_0, \mu_0, \sigma_0^2) = (0.0350, 0.6360, 0.3312)$

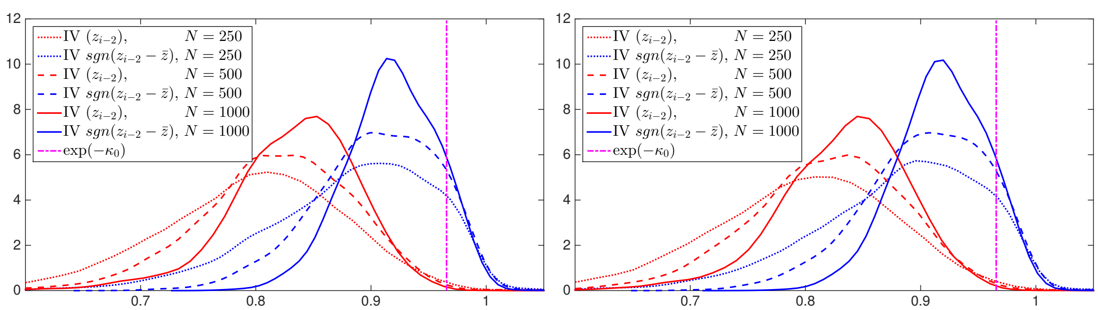


Figure 5 depicts the empirical distribution of the IV estimators for the multi-period-moment restrictions of daily spot volatility of the GARCH diffusion model. The figures are based on 10,000 simulations for three different sample sizes (250, 500 and 1,000). Two IV estimators are considered for the multi-period-moment restrictions. The first instrument is the two lags of spot volatility, while the second instrument is the sign of the centered value of the first instrument. The left and right columns are for the 5-minutes and 10-minutes realized volatilities, respectively.