Robust Jump Regressions*

Jia Li† Viktor Todorov‡ George Tauchen§

December 28, 2015

Abstract

We develop robust inference methods for studying linear dependence between the jumps of discretely observed processes at high frequency. Unlike classical linear regressions, jump regressions are determined by a small number of jumps occurring over a fixed time interval and the rest of the components of the processes around the jump times. The latter are the continuous martingale parts of the processes as well as observation noise. By sampling more frequently the role of these components, which are hidden in the observed price, shrinks asymptotically. The robustness of our inference procedure is with respect to outliers, which are of particular importance in the current setting of relatively small number of jump observations. This is achieved by using non-smooth loss functions (like L_1) in the estimation. Unlike classical robust methods, the limit of the objective function here remains non-smooth. The proposed method is also robust to measurement error in the observed processes which is achieved by locally smoothing the high-frequency increments. In an empirical application to financial data we illustrate the usefulness of the robust techniques by contrasting the behavior of robust and OLS-type jump regressions in periods including disruptions of the financial markets such as so called "flash crashes."

Keywords: high-frequency data, jumps, microstructure noise, robust regression, semimartingale.

---

*Li's and Todorov's research have been partially supported by NSF grants SES-1326819 and SES-0957330 and SES-1530748, respectively. Yuan Xue provided excellent research assistance.

†Department of Economics, Duke University.

‡Department of Finance, Northwestern University.

§Department of Economics, Duke University.

1 Introduction

When major events occur in the economy, asset prices often respond with abrupt large moves. These price moves are typically modeled as jumps in continuous-time semimartingale models ([22], [7]). Understanding the dependence between asset prices at times of market jumps sheds light on how firm values respond to market-wide information, which is of interest both for researchers and practitioners; see, e.g., [27] and [6]. More generally, jumps in semimartingales are used to model spike-like, or "bursty," phenomena in engineering and neuroscience; see, e.g., chapter 10 of [25]. The goal of the current paper is to develop robust inference techniques for regressions that connect jumps in multivariate semimartingales observed at high frequency. High-frequency data are well suited for studying jumps because they give a microscopic view of the process's dynamics around jump-inducing events. Robustness is needed to guard against both potential outliers and measurement errors.

The statistical setup here differs in critical dimensions from that of the classical linear regression model. The asymptotic behavior of the estimator is driven by the local behavior of the observed process at a finite number of jump times. The observed high-frequency increments around the jumps also contain the non-jump components of the price, i.e., the drift, the continuous martingale part, and possibly observation error. The drift component of the process is dominated at high frequencies by the continuous martingale part. The latter component around the jump times is approximately a sum of conditionally Gaussian independent random variables with conditional variances proportional to the length of the interval and the levels of the stochastic volatility of the process before and after the jump times. By sampling more frequently, this component of the price around the jump times shrinks asymptotically. When observation error is present, the precision does not increase as we sample more frequently. Nonetheless, smoothing techniques explained below will make it behave asymptotically similarly to the continuous martingale component of the price. Our setting thus shares similarities to one with a signal and asymptotically shrinking noise (e.g., section VII.4 in [12]).

In this paper, we pursue robustness for the jump regression in two dimensions. The first is robustness in the sense of Huber ([11]). The initial analysis considers a general class of extremum estimators using possibly non-quadratic and non-smooth loss functions. This framework accommodates, among others, the L_1-type estimators analogous to the least absolute deviation (LAD) and quantile regression estimators ([18]) of the classical setting; the results extend those of [20] for the least-squares jump regressions. In view of the different statistical setup, the asymptotic theory for robust estimators in the jump regression setting is notably different from that of classical extremum estimation. In the classical case, the sample objective function need not be smooth but the limiting objective function is smooth. In contrast, here both the sample and the limiting objective functions are non-smooth, because the kinks in the loss function are not "smoothed away" when the losses are aggregated over a fixed number of temporally separate jumps over a fixed sample span. Therefore, unlike the classical setting, the limiting objective function is not locally quadratic, and the asymptotic properties of the proposed extremum estimator need to be gleaned indirectly from the asymptotic behavior of the limiting objective function. We derive a feasible inference procedure in our setting which is very easy to implement.

The second sense of robustness is with respect to the observation error in high-frequency data. It is well-known that the standard semimartingale model is inadequate for modeling financial asset returns sampled at very high frequency. This is due to the fact that at such frequencies market microstructure frictions are no longer negligible ([26], [10]). Such frictions are typically treated as measurement errors statistically, and referred to as "microstructure noise." A large literature has been developed in the noisy setting for estimating integrated variance and covariances for the diffusive price component ([28], [4], [13] , [1]). Noise-robust inference concerning jumps is restricted to the estimation of power variations ([14], [19]).

In Section 3, we further extend the extremum estimation theory to a setting where the observations are contaminated with noise. We adopt the pre-averaging approach of [14] and locally smooth the data before conducting the robust jump regressions. That is, we form blocks of asymptotically increasing number of observations but with shrinking time span over which we average the data, and then we use these averages to detect the jumps and conduct the robust jump regressions. The local smoothing reduces the effect of the noise around the jump times to the second order.

We show that our robust jump regression techniques have very good finite sample properties on simulated data from models calibrated to real financial data. In an empirical application we study the reaction of Microsoft to big market jumps over the period 2007 – 2014. We find strong dependence between the Microsoft stock and the market at the time of market jumps. We examine the sensitivity of the robust jump regression with respect to two episodes in the data which are associated with potential market disruptions known as "flash crashes." We show that the robust jump regression estimates have very little sensitivity towards these events. This is unlike the least-squares estimates based on the detected jumps, which are very sensitive to the inclusion of these two episodes in the estimation.

This paper is organized as follows. Section 2 describes the baseline results in the setting without observation noise, which are extended to the noisy setting in Section 3. Section 4 contains a Monte Carlo evaluation and Section 5 provides an empirical example. Section 6 concludes. Technical assumptions are collected in the appendix. The online supplement contains all proofs, as well as MATLAB codes that are used in our numerical work.

2 The case without noise

In this section, we present the jump regression theory in the setting without noise. This theory extends that in [20] towards a setting with general (possibly non-smooth) loss functions, and serves as the baseline framework for the noise-robust theory that we further develop in Section 3.

2.1 The model

We consider two càdlàg (i.e., right continuous with left limit) adapted semimartingale processes Y and $\mathbf{Z}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, which respectively take values in $\mathbb{R}$ and $\mathbb{R}^{d-1}$. Let $\mathbf{X} \equiv (Y, \mathbf{Z})$. The jump of the d-dimensional process $\mathbf{X}$ at time t is denoted by $\Delta\mathbf{X}_t \equiv \mathbf{X}_t - \mathbf{X}_{t-}$, where $\mathbf{X}_{t-} \equiv \lim_{s\uparrow t} \mathbf{X}_s$.

The jump regression concerns the following population relationship between jumps of Y and $\mathbf{Z}$:

$$\Delta Y_\tau = \boldsymbol{\beta}^{*\top} \Delta\mathbf{Z}_\tau, \quad \tau \in \mathcal{T}, \tag{1}$$

where τ is a jump time of the process $\mathbf{Z}$, $\mathcal{T}$ is a collection of such times and $^\top$ denotes matrix transposition. We refer to the coefficient $\boldsymbol{\beta}^* \in \mathbb{R}^{d-1}$ as the *jump beta*, which is the parameter of interest in our statistical analysis. [20] provide empirical evidence that this simple model provides an adequate approximation for stock market data.

The model restriction (1) can be understood as a type of orthogonality condition. Indeed, if we define the residual process as

$$U_t^* = Y_t - \boldsymbol{\beta}^{*\top}\mathbf{Z}_t, \tag{2}$$

model (1) amounts to saying that U^* does not jump at the same time as $\mathbf{Z}$. We remark that this model does not impose any restriction on the diffusive components of $\mathbf{X}$ nor on the idiosyncratic jumps of Y (i.e., jumps that occur outside of the times in $\mathcal{T}$).

The inference for $\boldsymbol{\beta}^*$ is complicated by the fact that the jumps are not directly observable from data, where the process $\mathbf{X}$ is only sampled on the discrete time grid $\mathcal{I}_n \equiv \{i\Delta_n : i = 0, \ldots, \lfloor T/\Delta_n \rfloor\}$ and $\lfloor \cdot \rfloor$ denotes the floor function. We account for the sampling uncertainty in an infill asymptotic setting where the time span $[0, T]$ is fixed and the sampling interval $\Delta_n \to 0$ asymptotically. This setup is applicable in situations where, for the available sampling frequency, the volatility remains approximately constant over the sampling interval and one can identify with good accuracy the jumps from the large increments of $\mathbf{X}$ (the cost of coarser sampling frequency is that the small jumps cannot be separated from the diffusive component of $\mathbf{X}$).

We denote the increments of $\mathbf{X}$ by $\Delta_i^n \mathbf{X} \equiv \mathbf{X}_{i\Delta_n} - \mathbf{X}_{(i-1)\Delta_n}$. The returns that contain jumps are collected by

$$\mathcal{J}_n^* \equiv \{i : \tau \in ((i-1)\,\Delta_n, i\Delta_n] \text{ for some } \tau \in \mathcal{T}\}. \tag{3}$$

The sample counterpart of (1) is then given by

$$\Delta_i^n Y = \boldsymbol{\beta}^{*\top} \Delta_i^n \mathbf{Z} + \Delta_i^n U^*, \quad i \in \mathcal{J}_n^*. \tag{4}$$

The error term $\Delta_i^n U^*$ contains the diffusive moves of the asset prices and plays the role of random disturbances in the jump regression. In contrast to the population relationship (1), (4) depicts a noisy relationship for the data, just like in classical regression settings.

That noted, we clarify some important differences between the jump regression and the classical regression from the outset. Firstly, we note that (4) only concerns jump returns, which form a small and unobserved subset of all high-frequency returns. Secondly, the cardinality of the set $\mathcal{J}_n^*$ is bounded by the number of jumps and, hence, does not diverge even in large samples because the time span is fixed. Consequently, the intuition underlying the law of large numbers in classical asymptotic settings does not apply here. Instead, the asymptotic justification for jump regressions is based on the fact that the error terms $\Delta_i^n U^*$ are asymptotically small because the diffusive price moves shrink at high frequencies.

2.2 The estimator and its implementation

To estimate $\boldsymbol{\beta}^*$ in (4), we first uncover the (unobservable) set $\mathcal{J}_n^*$. We use a standard thresholding method ([21]). To this end, we pick a sequence of truncation threshold $\mathbf{u}_n = (u_{j,n})_{1\leq j\leq d-1}$, such that for all $1 \leq j \leq d-1$,

$$u_{j,n} \asymp \Delta_n^{\varpi}, \quad \varpi \in (0, 1/2) . \tag{5}$$

The thresholding estimator for $\mathcal{J}_n^*$ is then given by

$$\mathcal{J}_n \equiv \mathcal{I}_n \setminus \{i : -\mathbf{u}_n \leq \Delta_i^n \mathbf{Z} \leq \mathbf{u}_n\} . \tag{6}$$

The rate condition (5) ensures the separation between the small diffusive increments (which are of order $\Delta_n^{1/2}$) and the jumps. Theoretically, any truncation threshold that satisfies (5) will work (including the same one for $j = 1, ..., d$). In practice, however, it is important to set $u_{j,n}$ in an adaptive way which takes into account the fact that the diffusive volatility changes over time (intuitively, what constitutes a big or small in magnitude move for the diffusive component of $\mathbf{Z}$ depends on the level of its volatility). That is, we recommend using different thresholds over the sample period that track the diffusive volatility path (for which we form estimates), so that at each point $u_{j,n}$ represents several local standard deviations of the diffusive component. For this reason we also recommend to use separate truncation thresholds for the different components of $\mathbf{Z}$ to account for their possibly different levels of diffusive volatility at any point in time. We refer to Section 4 for implementation details.

We estimate the unknown parameter $\boldsymbol{\beta}^*$ using

$$\hat{\boldsymbol{\beta}}_n \equiv \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{i \in \mathcal{J}_n} \rho\left(\Delta_i^n Y - \mathbf{b}^\top \Delta_i^n \mathbf{Z}\right), \tag{7}$$

where the loss function $\rho(\cdot)$ is convex. The least-squares estimator of [20] corresponds to the special case with $\rho(u) = u^2$.

Our main motivation for deviating from the benchmark least-squares setting is due to a concern of robustness in the sense of [11]. Robustness is of particular interest in the jump regression setting because the number of large market moves is typically small within a given sample period; consequently, an outlying observation may be overly influential in the least-squares estimation. Such outliers can be due to “flash crashes” in financial markets, which we investigate empirically in Section 5, and more generally can result from some rare liquidity-related issues on financial markets, see e.g., [3]. We are particularly interested in the LAD estimation that corresponds to $\rho(u) = |u|$, where the non-smoothness of the objective function poses a nontrivial complication for the statistical inference. The extremum estimation theory, below, is thus distinct from prior work in a nontrivial way.

We assume that $\rho(\cdot)$ satisfies the following assumption, which allows for L_q loss functions, $q \geq 1$, as well as asymmetric loss functions used in regression quantiles ([18]).

Assumption 1. *(a) $\rho(\cdot)$ is a convex function on $\mathbb{R}$; (b) for some $q \in [1, 2]$, $\rho(au) = a^q \rho(u)$ for all $a > 0$ and $u \in \mathbb{R}$.*

The proposed estimation procedure is very simple to implement. The least-squares estimator admits a closed-form solution. The LAD estimator can be computed using standard software for quantile regressions. More generally, since $\rho(\cdot)$ is convex, the estimator can be computed efficiently using convex minimization software over observations indexed by the (typically small) set $\mathcal{J}_n$.

2.3 Discussion on regularity conditions

We now briefly discuss the regularity conditions on $\mathbf{X}$. The formal statement of these conditions are defered to the appendix. We assume that $\mathbf{X}$ is a d-dimensional Itô semimartingale of the form

$$\mathbf{X}_t = \int_0^t \mathbf{b}_s ds + \int_0^t \boldsymbol{\sigma}_s d\mathbf{W}_s + \mathbf{J}_t, \tag{8}$$

where $\mathbf{b}$ is the drift, $\boldsymbol{\sigma}$ denotes stochastic volatility, $\mathbf{W}$ is a multivariate standard Brownian motion and $\mathbf{J}$ is a pure-jump process with finite activity. The spot covariance matrix of $\mathbf{X}$ at time t is denoted by $\mathbf{c}_t \equiv \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\top$. We need some mild pathwise regularity conditions for these processes; see Assumption 2 in the appendix.

Clearly, the identification of $\boldsymbol{\beta}^*$ requires that the collection of jump vectors $(\Delta \mathbf{Z}_\tau)_{\tau \in \mathcal{T}}$ has full rank, so it is necessary that the number of jumps is not less than the number of regressors. We remind the reader that we are interested in uncovering *pathwise* properties of the studied processes as is typical in the infill asymptotic setting. Therefore, we confine our analysis to the event $\Omega_0 \equiv \{|\mathcal{T}| \geq d - 1\}$, where $|\mathcal{T}|$ denotes the cardinality of $\mathcal{T}$. When $\mathbf{Z}$ is scalar-valued, the identification condition automatically holds in Ω_0. In the multivariate setting (i.e., $d > 2$), the full rank condition is satisfied almost surely when the jump sizes have a continuous distribution, as stated in Assumption 3 in the appendix.

One caveat is that the full rank condition in the multivariate setting can break down when the jumps in the vector process $\mathbf{Z}$ have a low-dimensional linear factor structure. In this case, redundant regressors should be removed as in conventional regression analysis.

2.4 The asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$

We observe from (7) that $\hat{\boldsymbol{\beta}}_n$ is the solution to a convex minimization problem. Therefore, we can adapt a convexity argument ([16], [24], [8], [9]) to deduce the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$ from the finite-dimensional convergence of the objective function. To do so, we reparametrize the problem (7) via $\mathbf{h} = \Delta_n^{-1/2}(\mathbf{b} - \boldsymbol{\beta}^*)$ and consider the localized objective function

$$M_n\left(\mathbf{h}\right) \equiv \Delta_n^{-q/2} \sum_{i \in \mathcal{J}_n} \rho(\Delta_i^n Y - (\boldsymbol{\beta}^* + \Delta_n^{1/2}\mathbf{h})^\top \Delta_i^n \mathbf{Z}). \tag{9}$$

Note that $M_n(\cdot)$ is minimized by $\hat{\mathbf{h}}_n \equiv \Delta_n^{-1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)$.

We need some notations for describing the asymptotic distribution of $M_n\left(\cdot\right)$ and, subsequently, that of $\hat{\mathbf{h}}_n$. We denote the successive jump times of the process $\mathbf{Z}$ by $(\tau_p)_{p\geq 1}$ and collect them using the set $\mathcal{P} \equiv \{p \geq 1 : \tau_p \in [0, T]\}$. Let $(\kappa_p, \xi_{p-}, \xi_{p+})_{p\geq 1}$ be mutually independent random variables that are also independent of $\mathcal{F}$, such that κ_p is uniformly distributed on $[0,1]$ and the variables (ξ_{p-}, ξ_{p+}) are standard normal. We denote the spot variance of the residual process U_t^* by $\Sigma_t \equiv \left(1, -\boldsymbol{\beta}^{*\top}\right) \mathbf{c}_t \left(1, -\boldsymbol{\beta}^{*\top}\right)^\top$. We then set

$$\varsigma_p \equiv \sqrt{\kappa_p \Sigma_{\tau_p -}}\xi_{p-} + \sqrt{\left(1-\kappa_p\right)\Sigma_{\tau_p}}\xi_{p+}. \tag{10}$$

The variable ς_p represents the asymptotic distribution of the residual term $\Delta_i^n U^*$ for the unique i such that $\tau_p \in ((i-1)\Delta_n, i\Delta_n]$. Finally, we set

$$M(\mathbf{h}) \equiv \sum_{p \in \mathcal{P}} \rho\left(\varsigma_p - \mathbf{h}^\top \Delta \mathbf{Z}_{\tau_p}\right). \tag{11}$$

The main result of this section is the following theorem, where $\xrightarrow{\mathcal{L}\text{-}s}$ denotes stable convergence in law; see [15] for additional details about stable convergence.

Theorem 1. *Under Assumptions 1–3,*

$$\left(M_n\left(h_k\right)\right)_{1\leq k\leq \bar{k}} \xrightarrow{\mathcal{L}\text{-}s} \left(M(h_k)\right)_{1\leq k\leq \bar{k}}, \tag{12}$$

for any $h_k \in \mathbb{R}$, $1 \leq k \leq \bar{k}$ and $\bar{k} \geq 1$. Consequently, if $M(\cdot)$ has a unique minimum almost surely in restriction to Ω_0, then $\hat{\mathbf{h}}_n \xrightarrow{\mathcal{L}\text{-}s} \hat{\mathbf{h}} \equiv \mathrm{argmin}_{\mathbf{h}}\, M(\mathbf{h})$.

We remark on an important non-standard feature of Theorem 1. When $\rho(\cdot)$ is non-smooth, the limit objective function $M(\cdot)$ is also non-smooth. For example, $M(\mathbf{h}) = \sum_{p\in\mathcal{P}} |\varsigma_p - \mathbf{h}^\top \Delta \mathbf{Z}_{\tau_p}|$ in the LAD estimation, where the kink of the absolute value function is not "smoothed away" in the sum over a fixed number of jumps. This is unlike the classical LAD regression and quantile regressions, where the limit function would be smooth and locally quadratic. Here, the asymptotic distribution of $\hat{\mathbf{h}}_n$ is characterized as the exact distribution of the regression median from regressing the mixed Gaussian variables $(\varsigma_p)_{p\in\mathcal{P}}$ on the jump sizes $(\Delta \mathbf{Z}_{\tau_p})_{p\in\mathcal{P}}$. This distribution is non-standard and generally not mixed Gaussian. That noted, feasible inference is easily implemented as shown in Section 2.5.

The uniqueness of the minimum of $M(\cdot)$ can be verified in specific settings. A sufficient condition is the strict convexity of $\rho(\cdot)$. The LAD case does not verify strict convexity, but the uniqueness can be verified using finite-sample results for regression quantiles; see, for example, Theorem 2.1 of [17].

2.5 Feasible inference on the jump beta

Since the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$ shown in Theorem 1 is generally not $\mathcal{F}$-conditionally Gaussian, estimating consistently its asymptotic $\mathcal{F}$-conditional covariance matrix is not enough for constructing confidence sets of $\boldsymbol{\beta}^*$. We instead provide a simulation-based algorithm for approximating this non-standard asymptotic distribution.

The first step is to nonparametrically estimate the spot variance Σ_t before and after each detected jump. To this end, we pick an integer sequence m_n of block sizes such that $m_n \to \infty$ and $m_n \Delta_n \to 0$. We also pick a real sequence v_n of truncation thresholds that satisfies $v_n \asymp \Delta_n^{\varpi}$ for some $\varpi \in (0, 1/2)$. The truncation is used to conduct jump-robust estimation of the spot variances. The sample analogue of the residual U_t^* is given by $U_t \equiv Y_t - \hat{\boldsymbol{\beta}}_n^\top \mathbf{Z}_t$. For each $i \in \mathcal{J}_n$, we estimate the pre-jump and the post-jump spot variances respectively using

$$\begin{cases} \hat{\Sigma}_{n,i-} \equiv \dfrac{\sum_{j=0}^{m_n-1} \left|\Delta_{i-m_n+j}^n U\right|^2 1_{\left\{\left|\Delta_{i-m_n+j}^n U\right| \leq v_n\right\}}}{\Delta_n \sum_{j=0}^{m_n-1} 1_{\left\{\left|\Delta_{i-m_n+j}^n U\right| \leq v_n\right\}}}, \\ \hat{\Sigma}_{n,i+} \equiv \dfrac{\sum_{j=1}^{m_n} \left|\Delta_{i+j}^n U\right|^2 1_{\left\{\left|\Delta_{i+j}^n U\right| \leq v_n\right\}}}{\Delta_n \sum_{j=1}^{m_n} 1_{\left\{\left|\Delta_{i+j}^n U\right| \leq v_n\right\}}}. \end{cases} \tag{13}$$

The asymptotic distribution of $\hat{\mathbf{h}}_n = \Delta_n^{-1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)$ can be approximated via simulation as follows. Firstly, we draw a collection of mutually independent random variables $(\tilde{\kappa}_i, \tilde{\xi}_{i-}, \tilde{\xi}_{i+})_{i\in\mathcal{J}_n}$ such that $\tilde{\kappa}_i$ is uniformly distributed on the unit interval and $\tilde{\xi}_{i\pm}$ are standard normal. We set

$$\tilde{\varsigma}_{n,i} \equiv \left(\sqrt{\tilde{\kappa}_i \hat{\Sigma}_{n,i-}}\tilde{\xi}_{i-} + \sqrt{(1-\tilde{\kappa}_i)\,\hat{\Sigma}_{n,i+}}\tilde{\xi}_{i+}\right)$$

and compute $\tilde{\mathbf{h}}_n \equiv \operatorname{argmin}_{\mathbf{h}} \sum_{i\in\mathcal{J}_n} \rho(\tilde{\varsigma}_{n,i} - \mathbf{h}^\top \Delta_i^n \mathbf{Z})$. The Monte Carlo distribution of $\tilde{\mathbf{h}}_n$ is then used to approximate the asymptotic distribution of $\hat{\mathbf{h}}_n$. Theorem 2, below, provides the formal justification.

Theorem 2. *Under the conditions of Theorem 1, the $\mathcal{F}$-conditional law of $\tilde{\mathbf{h}}_n$ converges in probability to the $\mathcal{F}$-conditional law of $\hat{\mathbf{h}}$ under any metric for the weak convergence of probability measures.*

Confidence sets of $\boldsymbol{\beta}^*$ can be constructed using the simulated distribution of $\tilde{\mathbf{h}}_n$. For concreteness, we describe an example with $\boldsymbol{\beta}^*$ being a scalar, which can also be considered as a component of a vector. For $\alpha \in (0, 1)$, a two-sided $1-\alpha$ confidence interval (CI) of $\boldsymbol{\beta}^*$ can be constructed as $[\hat{\boldsymbol{\beta}}_n - \Delta_n^{1/2} z_{n,1-\alpha/2}, \hat{\boldsymbol{\beta}}_n - \Delta_n^{1/2} z_{n,\alpha/2}]$ where $z_{n,\alpha}$ denotes the α-quantile of $\tilde{\mathbf{h}}_n$ computed using the simulated sample.

3 The case with noise

3.1 The noisy setting

We now generalize the above setup to a setting in which the observations of $\mathbf{X}_{i\Delta_n}$ are contaminated with measurement errors. That is, instead of the the process $\mathbf{X}$, we observe a noisy process $\mathbf{X}'$ at discrete times given by

$$\mathbf{X}'_{i\Delta_n} = \mathbf{X}_{i\Delta_n} + \boldsymbol{\chi}'_{i\Delta_n}, \tag{14}$$

where $(\boldsymbol{\chi}'_{i\Delta_n})_{i\geq 0}$ denote the error terms. In financial settings, these error terms are often referred to as the microstructure noise and are attributed to market microstructure frictions such as the bid-ask bounce ([26]). Parallel to (2), the residual process in the noisy setting is given by $U_t'^{*} \equiv Y_t' - \boldsymbol{\beta}^{*\top}\mathbf{Z}_t'$. Since the sizes of the measurement errors remain constant even asymptotically, the baseline method in Section 2 is no longer valid. In the online supplement, we provide an analytical example for a precise illustration.

Below, we assume that the error terms $(\boldsymbol{\chi}'_{i\Delta_n})_{i\geq 0}$ are conditionally independent with zero mean given the $\mathbf{X}$ process, while allowing for essentially unrestricted heteroskedasticity; see Assumption 4 in the appendix.

3.2 Pre-averaging jump regressions

We propose a pre-averaging method to address the noisy data: we first locally smooth the noisy returns and then conduct the jump regression. In this paper, a function $g : \mathbb{R} \mapsto \mathbb{R}_+$ is called a weight function if it is supported on $[0, 1]$, continuously differentiable with Lipschitz continuous derivative and is strictly positive on $(0, 1)$. We also consider an interger sequence k_n of smoothing bandwidth. Below, we denote $g_n(j) = g(j/k_n)$. The pre-averaged returns are weighted moving averages of the noisy returns given by

$$\bar{\mathbf{X}}'_{n,i} = \sum_{j=1}^{k_n-1} g_n(j)\,\Delta_{i+j}^n \mathbf{X}', \quad i \in \mathcal{I}'_n \equiv \{0, \ldots, \lfloor T/\Delta_n \rfloor - k_n + 1\}. \tag{15}$$

The notations $\bar{\mathbf{Z}}'_{n,i}$ and $\bar{Y}'_{n,i}$ are defined similarly.

To guide intuition, we note that $\bar{\mathbf{X}}'_{n,i}$ can be decomposed into the contributions from jumps, the diffusive component and the noise component. The latter two components can be shown to have order $\sqrt{k_n \Delta_n}$ and $1/\sqrt{k_n}$, respectively. As a result, the rate-optimal choice of k_n is

$$k_n = \lfloor \theta / \Delta_n^{1/2} \rfloor, \quad \text{for some } \theta \in (0, \infty). \tag{16}$$

With this choice, the diffusive and the noise components are balanced at order $\Delta_n^{1/4}$. Accordingly, we consider a truncation sequence $\mathbf{u}'_n$ that satisfies $u'_{j,n} \asymp \Delta_n^{\varpi'}$ for all $1 \leq j \leq d-1$ and some $\varpi' \in (0, 1/4)$ and select pre-averaged jump returns using $\mathcal{J}'_n \equiv \mathcal{I}'_n \setminus \left\{ i : -\mathbf{u}'_n \leq \bar{\mathbf{Z}}'_{n,i} \leq \mathbf{u}'_n \right\}$. The set $\mathcal{J}'_n$ plays the role of an approximation to

$$\mathcal{J}'^{*}_n \equiv \{ i : \tau \in (i\Delta_n, (i+k_n)\Delta_n], \tau \in \mathcal{T} \}, \tag{17}$$

which collects the indices of the overlapping pre-averaging windows that contain the jump times.

The noise-robust estimator of $\boldsymbol{\beta}^*$ can be adapted from (7) by using pre-averaged returns and is defined as

$$\hat{\boldsymbol{\beta}}'_n = \underset{\mathbf{b}}{\operatorname{argmin}} \; \frac{1}{k_n} \sum_{i \in \mathcal{J}'_n} \rho\left(\bar{Y}'_{n,i} - \mathbf{b}^\top \bar{\mathbf{Z}}'_{n,i} \right). \tag{18}$$

Here, the normalizing factor $1/k_n$ is naturally introduced because each jump time τ is associated with k_n consecutive elements in $\mathcal{J}'^{*}_n$.

3.3 Asymptotic properties of $\hat{\boldsymbol{\beta}}_n'$

We derive the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n'$ by using a similar strategy as in Section 2.4. We consider the reparametrization $\mathbf{h} = \Delta_n^{-1/4}(\mathbf{b} - \boldsymbol{\beta}^*)$. The associated objective function

$$M_n'(\mathbf{h}) = \frac{1}{k_n \Delta_n^{q/4}} \sum_{i \in \mathcal{J}_n'} \rho\left(\bar{Y}_{n,i}' - \left(\boldsymbol{\beta}^* + \Delta_n^{1/4}\mathbf{h}\right)^\top \bar{\mathbf{Z}}_{n,i}'\right) \tag{19}$$

is minimized by $\hat{\mathbf{h}}_n' = \Delta_n^{-1/4}(\hat{\boldsymbol{\beta}}_n' - \boldsymbol{\beta}^*)$. Similarly as in Theorem 1, we study the asymptotic distribution of $\hat{\mathbf{h}}_n'$ by first establishing the finite dimensional asymptotic distribution of $M_n'(\cdot)$ and then using a convexity argument.

The asymptotic distribution of $M_n'(\cdot)$ is more difficult to study than that of $M_n(\cdot)$. The key complication is that each jump is associated with k_n overlapping pre-averaged returns. These pre-averaged returns are correlated and their number grows asymptotically. Consequently, we consider $\mathbb{R}$-valued processes $(\zeta_p(s))_{s\in[0,1]}$ and $(\zeta_p'(s))_{s\in[0,1]}$ which, conditional on $\mathcal{F}$, are mutually independent centered Gaussian processes with covariance functions given by

$$\begin{cases} \mathbb{E}\left[\zeta_p(s)\zeta_p(t)|\mathcal{F}\right] = \theta\Sigma_{\tau_p-}\int_{-1}^{0} g(s+u)\,g(t+u)\,du + \theta\Sigma_{\tau_p}\int_{0}^{1} g(s+u)\,g(t+u)\,du, \\ \mathbb{E}\left[\zeta_p'(s)\zeta_p'(t)|\mathcal{F}\right] = \theta^{-1}A_{\tau_p-}\int_{-1}^{0} g'(s+u)\,g'(t+u)\,du + \theta^{-1}A_{\tau_p}\int_{0}^{1} g'(s+u)\,g'(t+u)\,du, \end{cases} \tag{20}$$

where the process A is given by $A_t \equiv (1, -\boldsymbol{\beta}^{*\top})\mathbf{a}_t\mathbf{a}_t^\top(1, -\boldsymbol{\beta}^{*\top})^\top$, and $\mathbf{a}_t$ is the noise volatility. Roughly speaking, the $\mathcal{F}$-conditional Gaussian processes $\zeta_p(\cdot)$ (resp. $\zeta_p'(\cdot)$) capture the joint asymptotic behavior of the pre-averaged diffusive component (resp. noise component) of the residual process $Y_t' - \boldsymbol{\beta}^{*\top}\mathbf{Z}_t'$ around the jump time τ_p. We then set

$$\varsigma_p(s) = \zeta_p(s) + \zeta_p'(s), \quad s \in [0,1].$$

The process $\varsigma_p(\cdot)$ plays a similar role as the variable ς_p in Theorem 1.

The stable convergence in law of $M_n'(\mathbf{h})$ and $\hat{\mathbf{h}}_n'$ are described by Theorem 3 below.

Theorem 3. *Suppose Assumptions 1–4. Then* $(M'_n(h_k))_{1\leq k\leq \bar{k}} \xrightarrow{\mathcal{L}\text{-}s} (M'(h_k))_{1\leq k\leq \bar{k}}$*, for any* $h_k \in \mathbb{R}$*,* $1 \leq k \leq \bar{k}$ *and* $\bar{k} \geq 1$*, where*

$$M'(\mathbf{h}) \equiv \sum_{p\in\mathcal{P}} \int_0^1 \rho\left(\varsigma_p(s) - \mathbf{h}^\top \Delta \mathbf{Z}_{\tau_p} g(s)\right) ds. \tag{21}$$

If $M(\cdot)$ *is uniquely minimized by some random variable* $\hat{\mathbf{h}}'$ *almost surely in restriction to* Ω_0*, then* $\hat{\mathbf{h}}'_n = \Delta_n^{-1/4}(\hat{\boldsymbol{\beta}}'_n - \boldsymbol{\beta}^*) \xrightarrow{\mathcal{L}\text{-}s} \hat{\mathbf{h}}'$*.*

An interesting special case of Theorem 3 is the least-squares estimator with $\rho(u) = u^2$, which extends prior results in [20] to the current setting with noise. In this case, $\hat{\boldsymbol{\beta}}'_n$ admits a closed-form solution as the least-squares estimator of $\bar{Y}'_{n,i}$ versus $\bar{\mathbf{Z}}'_{n,i}$ for $i \in \mathcal{J}'_n$. The limiting variable $\hat{\mathbf{h}}'$ in Theorem 3 can also be explicitly expressed as

$$\hat{\mathbf{h}}' = \left(\int_0^1 g(s)^2 ds \sum_{p\in\mathcal{P}} \Delta\mathbf{Z}_{\tau_p} \Delta\mathbf{Z}_{\tau_p}^\top\right)^{-1} \left(\sum_{p\in\mathcal{P}} \Delta\mathbf{Z}_{\tau_p} \int_0^1 g(s)\varsigma_p(s) ds\right).$$

Since the processes $\varsigma_p(\cdot)$, $p \geq 1$, are $\mathcal{F}$-conditionally Gaussian, $\hat{\mathbf{h}}'$ is also $\mathcal{F}$-conditionally Gaussian. Here, the $\mathcal{F}$-conditional Gaussianity is obtained under a setting where $\mathbf{Z}$ and $\boldsymbol{\sigma}$ may jump at the same time. In contrast, the least-squares estimator is not $\mathcal{F}$-conditionally Gaussian when there are co-jumps in the noise-free setting. Intuitively, the indeterminacy of the exact jump time within a Δ_n-interval has negligible effect within a pre-averaging window of length $k_n\Delta_n$, so the extra layer of mixing from the uniform varaibles κ_p (recall (10)) does not appear in the pre-averaging setting.

3.4 Feasible inference in the noisy setting

We now describe a feasible inference procedure for $\boldsymbol{\beta}^*$ based on Theorem 3. This procedure adapts that in Section 2.5 to the pre-averaging setting. Since each jump time is associated with many pre-averaged returns in $\mathcal{J}_n'$, the first step is to group these returns into clusters accordingly. We partition $\mathcal{J}_n'$ into disjoint subsets $(\mathcal{J}_{n,p}')_{p\in\mathcal{P}_n}$ such that, for $p, l \in \mathcal{P}_n$ with $p < l$, the elements in $\mathcal{J}_{n,p}'$ are less than those in $\mathcal{J}_{n,l}'$ by at least $k_n/4$. Each cluster is associated with a jump time. The underlying theoretical intuition is as follows. It can be shown that the pre-averaged returns that do not contain jumps are not selected by $\mathcal{J}_n'$ uniformly with probability approaching one. Therefore, the elements of $\mathcal{J}_n'$ are clustered around associated jump times within a window of length $k_n\Delta_n$. Since the jump times are separated by a fixed amount of time, these clusters are eventually separated by any time window with shrinking length. In practice, this grouping procedure works well because we are mainly interested in relatively large jumps that are naturally well-separated in time.

For cluster $p \in \mathcal{P}_n$, we estimate the associated jump size and the spot variances Σ_t and A_t as follows. The jump size is estimated by

$$\Delta\hat{\mathbf{Z}}_{n,p} = \frac{\sum_{i\in\mathcal{J}_{n,p}'} \bar{\mathbf{Z}}_{n,i}'}{\sum_{j=\lfloor (k_n-|\mathcal{J}_{n,p}'|)/2\rfloor}^{\lfloor (k_n-|\mathcal{J}_{n,p}'|)/2\rfloor+|\mathcal{J}_{n,p}'|-1} g\left(j/k_n\right)}. \tag{22}$$

The denominator in (22) could be replaced by $\sum_{j=1}^{k_n} g\left(j/k_n\right)$ or $k_n\int_0^1 g\left(u\right)du$ without affecting the asymptotics. That being said, the current version of $\Delta\hat{\mathbf{Z}}_{n,p}$ makes a simple finite-sample adjustment that accounts for the fact that, when a jump occurs near the boundary of a pre-averaging window, the associated pre-averaged return may not be selected by $\mathcal{J}_n'$.

We observe that Σ_t and A_t are the spot variances of the diffusive and the noise components of the residual process U'^*, respectively. We approximate this residual process by $U_t' = Y_t' - \hat{\boldsymbol{\beta}}_n'^{\top}\mathbf{Z}_t'$ and then apply the spot variance estimators in [2]. We denote $g_n'(j) \equiv g_n(j) - g_n(j-1)$ and $\hat{U}_{n,i}' = \sum_{j=1}^{k_n} g_n'(j)^2\left(\Delta_{i+j}^n U'\right)^2$. We take a sequence of truncation threshold $v_n' \asymp \Delta_n^{\varpi'}$, $\varpi \in (0, 1/4)$, for constructing jump-robust spot variance estimators. For $i \geq 0$, we set

$$\hat{\Sigma}'_{n,i} = \frac{\sum_{j=1}^{k'_n} \left(\bar{U}'^2_{n,i+j} - \frac{1}{2} \hat{U}'_{n,i+j} \right) 1_{\left\{ \left| \bar{U}'_{n,i+j} \right| \leq v'_n \right\}}}{\Delta_n \sum_{j=1}^{k'_n} 1_{\left\{ \left| \bar{U}'_{n,i+j} \right| \leq v'_n \right\}} \sum_{j=1}^{k_n} g_n \left(j \right)^2},$$

$$\hat{A}_{n,i} = \frac{\sum_{j=1}^{k'_n} \hat{U}'_{n,i+j} 1_{\left\{ \left| \bar{U}'_{n,i+j} \right| \leq v'_n \right\}}}{2 \sum_{j=1}^{k'_n} 1_{\left\{ \left| \bar{U}'_{n,i+j} \right| \leq v'_n \right\}} \sum_{j=1}^{k_n} g'_n \left(j \right)^2}.$$

We use $\hat{\Sigma}'_{n,\min \mathcal{J}'_{n,p} - k'_n - k_n}$ and $\hat{\Sigma}'_{n,\max \mathcal{J}'_{n,p} + k_n - 1}$ to estimate Σ_t before and after the jump associated with cluster p. Similarly, the pre- and post-jump estimators of A_t are given by $\hat{A}_{n,\min \mathcal{J}'_{n,p} - k'_n - k_n}$ and $\hat{A}_{n,\max \mathcal{J}'_{n,p} + k_n - 1}$.

Algorithm 1, below, describes a simulation-based method for approximating the asymptotic distribution of $\hat{\mathbf{h}}'_n$ described in Theorem 3. Theorem 4 shows its first-order validity. Computer code for implementing this algorithm is available in the online supplement to this paper.

Algorithm 1.

Step 1. For cluster p, simulate random variables $(\tilde{r}'_{n,p,i})_{\{i:|i| \leq k_n - 1\}}$ given by

$$\tilde{r}'_{n,p,i} \equiv \tilde{r}_{n,p,i} + (\chi'_{n,p,i} - \chi'_{n,p,i-1}),$$

where $(\tilde{r}_{n,p,i}, \chi'_{n,p,i})$ are $\mathcal{F}$-conditionally independent such that $\tilde{r}_{n,p,i}$ is centered Gaussian with conditional variance $\Delta_n \hat{\Sigma}'_{n,\min \mathcal{J}'_{n,p} - k'_n - k_n}$ (resp. $\Delta_n \hat{\Sigma}'_{n,\max \mathcal{J}'_{n,p} + k_n - 1}$) when $i < 0$ (resp. $i \geq 0$) and $\chi'_{n,p,i}$ is centered Gaussian with variance $\hat{A}_{n,\min \mathcal{J}_{n,p} - k'_n - k_n}$ (resp. $\hat{A}_{n,\max \mathcal{J}'_{n,p} + k_n - 1}$) when $i < 0$ (resp. $i \geq 0$).

Step 2. Compute $\tilde{\mathbf{h}}'_n$ as the minimizer of

$$\tilde{M}'_n(\mathbf{h}) \equiv \frac{1}{k_n} \sum_{p \in \mathcal{P}_n} \sum_{i=0}^{k_n - 1} \rho \left(\left(\Delta_n^{-1/4} \sum_{j=1}^{k_n - 1} g_n \left(j \right) \tilde{r}'_{n,p,j-i} \right) - g_n \left(i \right) \mathbf{h}^\top \Delta \hat{\mathbf{Z}}_{n,p} \right).$$

Step 3. Approximate the $\mathcal{F}$-conditional asymptotic distribution of $\hat{\mathbf{h}}'_n$ using that of $\tilde{\mathbf{h}}'_n$, which can be formed by repeating Steps 1 and 2 in a large number of simulations. $\square$

Theorem 4. *Under the conditions of Theorem 3, the $\mathcal{F}$-conditional law of $\tilde{\mathbf{h}}'_n$ converges in probability to the $\mathcal{F}$-conditional law of $\hat{\mathbf{h}}'$ under any metric for the weak convergence of probability measures.*

4 Monte Carlo study

We now examine the asymptotic theory above in simulation scenarios that mimic our empirical setting in Section 5.

4.1 Setting

We consider two types of jump regression estimators. One is the least-squares estimator. The other is L_1-type estimators computed using $\rho(u) \equiv u\left(q - 1_{\{u<0\}}\right)$, $q \in (0,1)$. We refer to the latter as the quantile jump regression estimators because they resemble the classical regression quantiles ([18], [17]). We conduct experiments in the general setting with noise. The sample span is $T = 1$ year, containing 250 trading days. Each day contains $m = 4680$ high-frequency returns sampled at every five seconds. The returns are expressed in annualized percentage terms. There are 1000 Monte Carlo trials.

We adopt a data generating process that accommodates features such as leverage effect, price-volatility co-jumps, and heteroskedasticity in noise and jump sizes. Let W_1, W_2, B_1

and B_2 be independent Brownian motions. We generate the efficient prices according to

$$\begin{cases} d\log\left(V_{1,t}\right) = -\lambda_N \mu_V dt + \tilde{\sigma} dB_{1,t} + J_{V,t} dN_t, \quad V_{1,0} = \bar{V}_1, \\ \log\left(V_{2,t}\right) = \log\left(\bar{V}_2 - \beta_C^2 \bar{V}_1\right) + B_{2,t}, \\ dZ_t = \sqrt{V_{1,t}}\left(\rho dB_{1,t} + \sqrt{1-\rho^2} dW_{1,t}\right) + \varphi_{Z,t} dN_t, \\ dY_t = \beta_C \sqrt{V_{1,t}}\left(\rho dB_{1,t} + \sqrt{1-\rho^2} dW_{1,t}\right) + \sqrt{V_{2,t}} dW_{2,t} + \beta^* \varphi_{Z,t} dN_t, \end{cases} \tag{23}$$

where the parameter of interest is $\beta^* = 1$ and other components are given by

$$\begin{cases} \bar{V}_1 = 18^2, \quad \bar{V}_2 = 26^2, \quad \rho = -0.7, \quad \tilde{\sigma} = 0.5, \quad \beta_C = 0.89, \\ J_{V,t} \overset{i.i.d.}{\sim} \text{Exponential}\left(\mu_V\right), \quad \mu_V = 0.1, \\ \varphi_{Z,t} | V_{1,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \phi^2 V_{1,t}\right), \quad \phi = 0.055, \\ N_t \text{ is a Poisson process with intensity } \lambda_N = 20. \end{cases} \tag{24}$$

We generate the noise terms for Y and Z respectively as $a_{Y,t}\chi_{Y,t}$ and $a_{Z,t}\chi_{Z,t}$, where $(\chi_{Y,t}, \chi_{Z,t})_{t\geq 0}$ are drawn independently from the standard normal distribution and the volatility processes of the noise are given by $a_{Y,t} = \bar{a}\sqrt{\beta_C^2 V_{1,t} + V_{2,t}}$ and $a_{Z,t} = \bar{a}\sqrt{V_{1,t}}$. We set $\bar{a} = 0.0028$ so that the magnitude of the noise is three times the local standard deviation of the diffusive returns. In other words, the contribution of the noise in the realized variance computed using 5-second returns is 18 times the contribution of the diffusive component. The simulated returns are therefore fairly noisy.

We implement the estimation procedures with two pre-averaging windows, $k_n \in \{36, 60\}$, for checking robustness. We fix $k_n' = 720$, while noting that results for $k_n' = 960$ are very similar so they are omitted for brevity. The weight function is $g(x) = g_0(|2x-1|)1_{\{0\leq x\leq 1\}}$, where $g_0(x) = 1 - 3x^2 + 2x^3$. For each trading day, the truncation threshold is chosen adaptively as $u_n' = 7\sqrt{BV\left(Z'\right)}$, where $BV\left(Z'\right)$ is the average of $(\pi/2)\,|\bar{Z}'_{n,ik_n}||\bar{Z}'_{n,(i+1)k_n}|$ over all i such that the pre-averaging windows associated with $\bar{Z}'_{n,ik_n}$ and $\bar{Z}'_{n,(i+1)k_n}$ are within the same day. The statistic $BV(Z')$ is a jump-robust proxy for the standard deviation of the pre-averaged returns, formed using the bipower construction of [5] and [23]. We set $v_n' = 4BV(U')$.

4.2 Results

Table 1 reports the simulation results. Panels A and B present results for $k_n = 36$ and $k_n = 60$, respectively. For each estimator, we report its bias, mean absolute deviation (MAD) and root mean squared error (RMSE). We also report the coverage rates of CIs at nominal levels 90%, 95% and 99%. Here, a level $1-\alpha$ CI is given by $[\hat{\boldsymbol{\beta}}_n' - \Delta_n^{1/4} z_{n,1-\alpha/2}, \hat{\boldsymbol{\beta}}_n' - \Delta_n^{1/4} z_{n,\alpha/2}]$, where $z_{n,\alpha}$ denotes the α-quantile of $\tilde{\boldsymbol{h}}_n'$ given by Algorithm 1.

From Table 1, we see that the proposed estimators have very small biases and are fairly accurate. The least-squares estimator is more accurate than the quantile regression estimators, indicating some tradeoff between efficiency and robustness. However, we note that the accuracy of the LAD estimator (i.e., $q = 0.5$) is similar to that of the least-squares estimator. In addition, we observe that the coverage rates of the CIs are very close to the associated nominal levels. Overall, the simulation evidence supports the asymptotic theory.

5 Empirical application

We now apply the robust jump regression method to study the sensitivity of the stock price of Microsoft (NASDAQ: MSFT) to market jumps. The S&P 500 ETF is used as a proxy for the market portfolio. The asset prices are sampled at every five seconds from January 3, 2007 to September 30, 2014. We discard incomplete trading days and, for now, also discard two well-known days with major "Flash Crashes" (May 6, 2010 and April 23, 2013). The resultant sample contains 1931 trading days. We apply the noise-robust method developed in Section 3, for which tuning parameters are set similarly as in the simulations.

	Bias	MAD	RMSE	CI Coverage		
				90%	95%	99%
			Panel A. $k_n = 36$			
Least-squares	-0.003	0.018	0.024	0.896	0.945	0.985
$q = 0.10$	-0.002	0.028	0.039	0.890	0.941	0.981
$q = 0.25$	-0.002	0.022	0.031	0.886	0.939	0.987
$q = 0.50$	-0.002	0.019	0.026	0.891	0.935	0.985
$q = 0.75$	-0.003	0.023	0.031	0.888	0.941	0.989
$q = 0.90$	-0.004	0.029	0.041	0.883	0.929	0.989
			Panel B. $k_n = 60$			
Least-squares	-0.002	0.022	0.032	0.919	0.956	0.986
$q = 0.10$	-0.002	0.034	0.049	0.892	0.94	0.987
$q = 0.25$	-0.003	0.028	0.041	0.895	0.946	0.99
$q = 0.50$	-0.003	0.024	0.035	0.903	0.951	0.985
$q = 0.75$	-0.003	0.028	0.04	0.893	0.942	0.985
$q = 0.90$	-0.003	0.035	0.05	0.894	0.944	0.986

Table 1: Summary of simulation results. We report biases, mean absolute deviations (MAD), root mean squared errors (RMSE) and coverage rates of confidence intervals (CI) for the least-squares and the q-quantile jump regression procedure. Panels A and B report results for $k_n = 36$ and 60, respectively. There are 1000 Monte Carlo trials.

We perform an additional sensitivity check regarding the choice of the truncation threshold u'_n: we set $u'_n = \bar{u}\sqrt{BV(Z')}$ and vary $\bar{u}$ from 6 to 7.5. As in prior work, the truncation threshold is also scaled to account for the deterministic diurnal volatility pattern, but the details are omitted for brevity.

Table 2 reports the point estimates and 95% CIs from the least-squares and the LAD procedures implemented using various tuning schemes. We see that the least-squares and the LAD estimates are generally similar and have good statistical precision. These estimates appear reasonably insensitive to various changes in the tuning parameters.

k_n	$\bar{u}$	Least Squares		LAD	
		$\hat{\beta}'_n$	95% CI	$\hat{\beta}'_n$	95% CI
36	6.0	0.877	[0.841; 0.911]	0.897	[0.874; 0.921]
36	6.5	0.885	[0.847; 0.924]	0.905	[0.877; 0.934]
36	7.0	0.898	[0.855; 0.939]	0.899	[0.869; 0.931]
36	7.5	0.897	[0.852; 0.945]	0.901	[0.866; 0.934]
60	6.0	0.894	[0.849; 0.941]	0.915	[0.884; 0.945]
60	6.5	0.895	[0.843; 0.951]	0.915	[0.877; 0.948]
60	7.0	0.885	[0.823; 0.944]	0.899	[0.858; 0.939]
60	7.5	0.890	[0.838; 0.939]	0.878	[0.833; 0.923]

Table 2: Pre-averaging jump beta estimates for MSFT. Confidence intervals (CI) are computed using 1000 Monte Carlo repetitions from Algorithm 1.

Figure 1a shows a scatter plot for the estimated jump sizes $\Delta\hat{Z}_n$ and $\Delta\hat{Y}_n$ along with fitted regression lines. This figure suggests that the linear model indeed provides a reasonable fit for the central scatter of the jump pairs. We further compute quantile jump regression estimators at quantiles $q \in \{0.1, 0.2, \ldots, 0.9\}$. Figure 1b plots these estimates (dashed line) with associated 95% CIs. Note that the simulation-based CIs are not necessarily centered around the point estimates. For this reason, we also plot a centered version of the beta estimate (solid line) that is defined as the 50% confidence bound for the jump beta. Figure 1b suggests a modest increase in the quantile beta estimates across quantiles. By way of economic interpretation, the residuals of the linear model are the hedging errors from a portfolio using a proportion, or hedge ratio (beta), of the market to hedge aggregate jump risk, and the statistical objective function measures total loss from un-hedged jump variation. Figure 1b indicates that an investor who weights more heavily negative losses should use a somewhat smaller hedge ratio.

Finally, we examine the robustness of the least-squares and the LAD estimators against outliers. While this type of comparison can be easily made via artificial numerical experiments, here we aim to demonstrate the robustness of the LAD estimator in a real-data setting. We do so by including the two aforementioned Flash Crashes into our sample. The idea here is to use these Flash Crashes as extreme, but realistic, examples to "stress test" the robustness properties of the proposed estimators.

Table 3 reports the least-squares and the LAD estimates for samples with or without the two Flash Crash days. Results from various tuning schemes are presented. We find that these outlying observations indeed induce substantial downward biases in the least-squares estimates. The bias is most pronounced when the truncation threshold is high. In contrast, the LAD estimator is remarkably robust against these outliers. This finding reaffirms the relevance of our initial motivation for developing jump regressions with general loss functions.

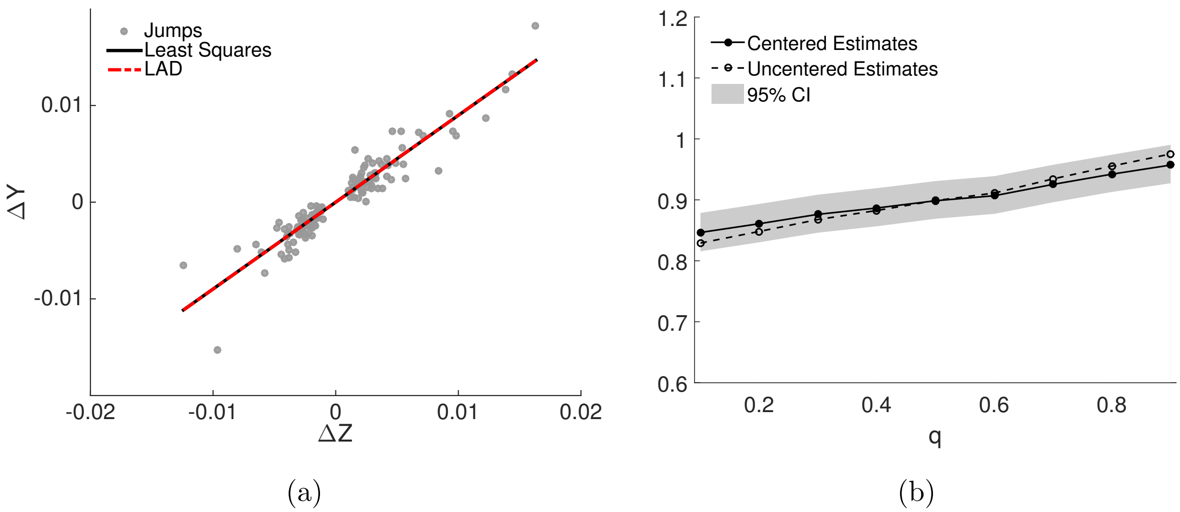

Figure 1: Illustration of the pre-averaging jump regressions with $k_n = 36$, $k_n' = 720$ and $\bar{u} = 7$. (a) Scatter plot of the jump size estimates $\Delta\hat{Y}_n$ and $\Delta\hat{Z}_n$ with fitted regression lines using the least-squares and the LAD estimates. (b) Quantile jump regression estimates at quantile $q \in \{0.1, 0.2, \ldots, 0.9\}$. The centered estimate is defined as the 50% confidence bound. The uncentered estimate is given by eq. (18). Confidence intervals (CI) are computed using 1000 Monte Carlo repetitions from Algorithm 1.

6 Conclusion

In this paper we propose robust inference techniques for studying linear dependence between the jumps of discretely-observed processes, e.g., financial prices. The data for the inference consist of high-frequency observations of the processes on a fixed time interval with asymptotically shrinking length between observations. The jumps are hidden in the "big" increments of the process and the difference between the two drives the asymptotic behavior of our robust jump regression estimators. Our inference is based on minimizing the residual from the model-implied linear relation between the detected jumps in the data.

We allow for non-smooth loss functions so as to accommodate leading robust regression methods. Unlike the classical robust regression, in the current setting the limit of the objective function continues to be non-smooth as the asymptotics is driven by a finite number of jumps on the given interval, along with local price increments around these jump times. To further robustify the analysis against the presence of measurement error at the observation times, we locally smooth (pre-average) the discrete observations of the processes around the detected jump times. We provide easy-to-implement simulation methods for conducting feasible inference and illustrate their good finite sample behavior in a Monte Carlo study. In an empirical application, we illustrate the gains from the robust regression by analyzing the stability of the jump regressions during periods which include potential market disruptions.

7 Appendix

In this appendix, we discuss in detail the technical regularity conditions used in our asymptotic theory.

We assume that $\mathbf{X}$ is a d-dimensional Itô semimartingale of the form (8). The drift process $\mathbf{b}$ and the volatility process $\boldsymbol{\sigma}$ are càdlàg adapted. The jump component of $\boldsymbol{X}$ can be written as $\mathbf{J}_t = \int_0^t \int_{\mathbb{R}} \boldsymbol{\delta}\,(s,u)\,\mu\,(ds,du)$, where $\boldsymbol{\delta}(\cdot) : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^d$ is a predictable function and μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with its compensator $\nu\,(dt,du) = dt \otimes \lambda\,(du)$ for some measure $\lambda(\cdot)$ on $\mathbb{R}$. We assume the following condition.

Assumption 2. *(a) The process* $\mathbf{b}$ *is locally bounded; (b)* $\mathbf{c}_t$ *is nonsingular for* $t \in [0,T]$ *almost surely; (c)* $\nu\,([0,T] \times \mathbb{R}) < \infty$*.*

The only nontrivial restriction in Assumption 2 is the assumption of finite-activity jumps in $\mathbf{X}$. This assumption is mainly used to simplify our technical exposition because the empirical focus of jump regressions is the big jumps. Technically speaking, this means that we can drop Assumption 2(c) and focus on jumps with size bounded away from zero without changing the results in the main text.

The following condition is sufficient for the identification of the jump beta in the multivariate setting. Recall that $(\tau_p)_{p\geq 1}$ denote the successive jump times of $\mathbf{Z}$.

Assumption 3. *Suppose* $\mathbb{P}(\Omega_0) > 0$ *and, in restriction to* Ω_0*, the joint distribution of* $\left(\Delta \mathbf{Z}_{\tau_p}\right)_{p\geq 1}$ *is absolutely continuous with respect to the Lebesgue measure.*

The conditions on the measurement errors are given by the following.

Assumption 4. *We have* $\boldsymbol{\chi}'_{i\Delta_n} = \mathbf{a}_{i\Delta_n}\boldsymbol{\chi}_{i\Delta_n}$ *such that (i) the* $\mathbb{R}^{d\times d}$*-valued process* $(\mathbf{a}_t)_{t\geq 0}$ *is càdlàg adapted and locally bounded; (ii) the variables* $(\boldsymbol{\chi}_{i\Delta_n})_{i\geq 0}$ *are mutually independent and independent of* $\mathcal{F}$ *such that* $\mathbb{E}\left[\boldsymbol{\chi}_{i\Delta_n}\right] = \mathbf{0}$, $\mathbb{E}\left[\boldsymbol{\chi}_{i\Delta_n}\boldsymbol{\chi}_{i\Delta_n}^{\top}\right] = \mathbf{I}_d$ *and* $\mathbb{E}\left[\|\boldsymbol{\chi}_{i\Delta_n}\|^v\right]$ *is finite for all* $v \geq 1$.

The essential part of Assumption 4 is that the noise terms $(\boldsymbol{\chi}'_{i\Delta_n})_{i\geq 0}$ are $\mathcal{F}$-conditionally independent with zero mean. For the results in the main text, we only need $\boldsymbol{\chi}_{i\Delta_n}$ to have finite moments up to a certain order; assuming finite moments for all orders is merely for technical convenience. Finally, we note that the noise terms are allowed to be heteroskedastic and serially dependent through the volatility process $(\mathbf{a}_t)_{t\geq 0}$.

		Least Squares			LAD		
		Flash Crashes?			Flash Crashes?		
k_n	$\bar{u}$	No	Yes	Difference	No	Yes	Difference
36	6.0	0.877	0.706	0.171	0.897	0.849	0.048
36	6.5	0.885	0.706	0.179	0.905	0.847	0.058
36	7.0	0.898	0.709	0.189	0.899	0.841	0.057
36	7.5	0.897	0.706	0.191	0.901	0.825	0.076
60	6.0	0.894	0.714	0.179	0.915	0.898	0.017
60	6.5	0.895	0.703	0.193	0.915	0.901	0.013
60	7.0	0.885	0.668	0.217	0.899	0.878	0.021
60	7.5	0.890	0.654	0.236	0.878	0.857	0.020

Table 3: Robustness assessment of the pre-averaging jump beta estimators. Note: We report the pre-averaging least-squares and LAD estimates for samples excluding (resp. including) the two days with major Flash Crashes (May 6, 2010 and April 23, 2013) under the column headed "No (resp. Yes)." The difference of the estimates using these two samples are reported in the column headed "Difference."

References

[1] Yacine Aït-Sahalia, Jianqing Fan, and Dacheng Xiu. High-frequency covariance estimates with noisy and asynchronous data. *Journal of the American Statistical Association*, 105:1504–1517, 2010.

[2] Yacine Aït-Sahalia, Jean Jacod, and Jia Li. Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168:207–222, 2012.

[3] O.E. Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3):C1–C32, 2009.

[4] Ole E. Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76:1481–1536, 2008.

[5] Ole E. Barndorff-Nielsen and N. Shephard. Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, 2:1–37, 2004.

[6] T. Bollerslev, S. Li, and V. Todorov. Roughing up beta: Continuous vs. discontinuous betas, and the cross-section of expected stock returns. *Journal of Financial Economics, forthcoming*, 2015.

[7] Rama Cont and Peter Tankov. *Financial Modeling with Jump Processes*. Chapman and Hall/CRC, 2003.

[8] Richard A. Davis, Keith Knight, and Jian Liu. M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications*, 40:145–180, 1992.

[9] C. J. Geyer. On the asymptotics of convex stochastic optimization. Technical report, Unpublished manuscript, 1996.

[10] Peter R. Hansen and Asger Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24:127–161, 2006.

[11] Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. Wiley, 2009.

[12] I. Ibragimov and R. Has'minskii. *Statistical Estimation: Asymptotic Theory*. Springer, Berlin, 1981.

[13] Jean Jacod, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications*, 119:2249–2276, 2009.

[14] Jean Jacod, Mark Podolskij, and Mathias Vetter. Limit theorems for moving averages of discretized processes plus noise. *Annals of Statistics*, 38:1478–1545, 2010.

[15] Jean Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, second edition, 2003.

[16] Keith Knight. Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 17(3):pp. 261–278, 1989.

[17] Roger Koenker. *Quantile Regression*. Cambridge University Press, 2005.

[18] Roger Koenker and Jr. Bassett, Gilbert. Regression quantiles. *Econometrica*, 46(1):pp. 33–50, 1978.

[19] Jia Li. Robust estimation and inference for jumps in noisy high frequency data: A local-to-continuity theory for the pre-averaging method. *Econometrica*, 81:1673–1693, 2013.

[20] Jia Li, Viktor Todorov, and George Tauchen. Jump regressions. Technical report, Duke Univeristy, 2014.

[21] Cecilia Mancini. Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, LXIV:19–47, 2001.

[22] Robert C. Merton. Theory of finance from the perspective of continuous time. *Journal of Financial and Quantitative Analysis*, 10(4):659 – 674, 1975.

[23] Mark Podolskij and M. Vetter. Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and Their Applications*, 119:2803–2831, 2009.

[24] David Pollard. Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):pp. 186–199, 1991.

[25] B.L.S. Prakasa Rao. *Semimartingales and their Statistical Inference*. Chapman and Hall/CRC, 1999.

[26] Richard Roll. A simple model of the implicit bid-ask spread in an efficient market. *The Journal of Finance*, 39:1127–1139, 1984.

[27] Pavel Savor and Mungo Wilson. Asset pricing: A tale of two days. *Journal of Financial Economics*, 113:171–201, 2014.

[28] Lan Zhang, Per A. Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100:1394–1411, 2005.