The Annals of Applied Statistics
2020, Vol. 14, No. 4, 1604–1621
https://doi.org/10.1214/20-AOAS1365

MEASURING TIMELINESS OF ANNUAL REPORTS FILING BY JUMP ADDITIVE MODELS

BY YICHENG KANG

Department of Mathematical Sciences, Bentley University, ykang@bentley.edu

Foreign public issuers (FPIs) are required by the Securities and Exchanges Commission (SEC) to file Form 20-F as comprehensive annual reports. In an effort to increase the usefulness of 20-Fs, the SEC recently enacted a regulation to accelerate the deadline of 20-F filing from six months to four months after the fiscal year-end. The rationale is that the shortened reporting lag would improve the informational relevance of 20-Fs. In this work we propose a jump additive model to evaluate the SEC's rationale by investigating the relationship between the timeliness of 20-F filing and its decision usefulness using the market data. The proposed model extends the conventional additive models to allow possible discontinuities in the regression functions. We suggest a two-step jump-preserving estimation procedure and show that it is statistically consistent. By applying the procedure to the 20-F study, we find a moderate positive association between the magnitude of the market reaction and the filing timeliness when the acceleration is less than 17 days. We also find that the market considers the filings significantly more informative when the acceleration is more than 18 days and such reaction tapers off when the acceleration exceeds 40 days.

1. Introduction. Foreign public issuers (FPIs) are required by the Securities and Exchanges Commission (SEC) to file Form 20-F as comprehensive annual reports. Similar to Form 10-K of U.S. domestic firms, 20-Fs are regarded as the most valuable source of financial information available to equity investors of foreign firms. Both the 10-K and the 20-F give a comprehensive summary of a company's financial performance. Detailed information such as company cashflow, organizational structure, executive compensation and corporate governance are included. In the accounting literature there has been extensive evidence that 10-Ks contain useful information for investors' investment decision making (e.g., Griffin (2003); Asthana, Balsam and Sankaraguruswamy (2004); Callen, Livnat and Segal (2006); You and Zhang (2009); De Franco, Wong and Zhou (2011); Christensen, Heninger and Stice (2013)). However, the literature on the decision usefulness of 20-Fs provides rather mixed results (e.g., Meek (1983); Etter, Rees and Lukawitz (1999); Olibe (2001); Chen and Sami (2008); Chen and Sami (2013); Kim, Li and Li (2012)), possibly because 20-Fs differ from 10-Ks in several ways. For instance, the deadline for filing a 20-F is six months after the fiscal year-end, much longer than the two-month or three-month deadline for filing a 10-K. The delay may cause investors to search private information before the release of 20-Fs, rendering 20-Fs less informative. In an effort to improve the information environment for FPIs in the U.S., the SEC recently enacted a regulation to accelerate the deadline of 20-F filing from six months to four months after the fiscal year-end. Effective for the fiscal year ending on or after December 15, 2011, the SEC requires all the FPIs to comply with the new four-month deadline. The aim is to increase the usefulness of 20-Fs by providing investors access to financial information prepared by FPIs in a timelier fashion.

Received April 2020; revised June 2020.
Key words and phrases. Backfitting, decision usefulness, jump detection, jump-preserving estimation, market reaction, shortened reporting lag.

Motivated by the SEC regulation change, this study examines empirically whether timelier filing increases market responses and, more importantly, how much acceleration leads to the most significantly improved relevance of 20-F filing. To appropriately address this question, all the other control variables that might cause a difference in market reaction need to be properly taken into account by the empirical analysis. Particularly in our 20-F application, firm-specific variables (e.g., asset and liability) ought to be carefully controlled. It can be challenging because the functional relationships between these variables and the market response are hardly linear. In fact, it is difficult to even specify a parametric form for these functional relationships. Thus, a flexible nonparametric method would be desirable. In the statistics literature, nonadditive nonparametric models, such as local regression, are commonly used. Despite being flexible enough, these models become less useful in dimensions higher than two or three. This problem of dimensionality is often referred to as the curse of dimensionality (Hastie, Tibshirani and Friedman (2009)). Additive models (Breiman and Friedman (1985)) mitigate this issue by imposing an additive structure and are still flexible to use without making restrictive parametric assumptions. Thus, we adopt additive models as the framework for our empirical analysis. As for the relationship between the market response and the filing timeliness (i.e., the relationship of interest), in cases when a firm accelerates its 20-F filing significantly (e.g., a month), it might cause an abrupt shift in both the magnitude and the direction of the market response because investors may view such improvement of filing timeliness as a material change in terms of the information relevance of the 20-F. In cases when the acceleration is less significant (e.g., only a few days), the increase in market reaction would be less obvious. By this consideration there may exist jumps in the functional relationship (Qiu (2005)) between the filing timeliness and the market response. This is especially true when there is a shift in the direction of the market response. The jump locations indicate the amount of acceleration with which significant shifts in the market reaction are associated, and the jump sizes quantify the magnitudes of such shifts. Thus, it is important for our additive modeling framework to accommodate the existence of possible jumps in the regression functions.

There has been much discussion in the literature about estimation of jump regression curves (e.g., Hall and Titterington (1992), Kang (2020a), Kang et al. (2019), Ma and Yang (2011), Müller (2002), Qiu (2005), Wang (1995), Xia and Qiu (2015)). More recent research considers a similar problem of detecting structural changes in the context of time series analysis (e.g., Casini (2018), Casini (2019), Casini and Perron (2019), Wu and Zhao (2007)). These existing methods consider the problem of jump detection in cases where there is only one or two predictors. It is unclear whether these methods can be directly extended to cases where there are multiple predictors, such as in our 20-F application. There also has been some recent research on change point detection in higher dimensions (e.g., Cho and Fryzlewicz (2015), Avanesov and Buzun (2018)). However, jump regression is different from change point detection (Xia and Qiu (2015)), in that the mean response could be an arbitrary continuous curve between two consecutive jump points in the former, but is constant between two consecutive change points in the latter. Thus, change point detection methods are generally not applicable to our 20-F problem. In this article we propose a novel jump additive model (JAM) that detects jump locations and preserves jump sizes. We do not impose any parametric assumptions on the functional relationships besides the additive structure and allow multiple predictors in the model. We show that the estimated number of jumps, the estimated jump locations, the estimated jump sizes and the estimated regression functions converge to the truths as the sample size grows. Applying JAM to our 20-F timeliness study, we find a moderate positive association between the magnitude of the market reaction and the filing timeliness when the acceleration is less than 17 days. We also find that the market considers the filings significantly more informative when the acceleration is a little more than 18 days and such reaction

tapers off when the acceleration exceeds 40 days. JAM is implemented in our R-package *jam*, which is included in the Supplementary Material (Kang (2020)).

The remainder of this article is organized as follows. The research design for our 20-F study is set up in Section 2. The proposed estimation procedure of jump additive models is described in detail in Section 3. The empirical findings of our 20-F study are presented in Section 4. Some simulated examples are given in Section 5. Several concluding remarks are provided in Section 6. Asymptotic properties of the proposed procedure are discussed in the Appendix. Proofs of the theorems and some technical arguments are given in the Supplementary Material (Kang (2020)).

2. Research design. Since stock prices can react differently to the release of information (Beaver (1968); Kim and Verrecchia (1991)), a commonly used measure of market response is cumulative abnormal return (e.g., Doyle and Magilke (2013)). This measure is also adopted here. More specifically, let

$$rt_{j,t} = \alpha_j + \beta_j \times rt_{m,t} + \epsilon_{j,t},$$

where $rt_{j,t}$ is the daily stock return of the jth firm on day t and $rt_{m,t}$ is the return on the U.S. market index S&P 500 on day t. Let $t = 0$ denote the 20-F filing date. Estimate α_j and β_j using standard linear regression during the period $[-260, -11]$. Denote the estimates for α_j and β_j by $\widehat{\alpha}_j$ and $\widehat{\beta}_j$, respectively. Since we are concerned with both the direction and the magnitude of the market reaction, we calculate cumulative abnormal return for each 20-F filing date over event window $[T_1, T_2]$ as follows:

$$y_j = \sum_{t \in [T_1, T_2]} (rt_{j,t} - \widehat{rt}_{j,t}),$$

where $\widehat{rt}_{j,t} = \widehat{\alpha}_j + \widehat{\beta}_j rt_{m,t}$. $[T_1, T_2]$ is chosen to be $[-1, 4]$, because it takes into account potential information leakage and possible delays of 20-F releases. y_j is called the cumulative abnormal return around the jth firm's 20-F filing.

To investigate the relationship between the market response and the filing timeliness after controlling other factors, we propose the following empirical model:

$$\begin{aligned} y_j = {} & \phi_0 + \varphi_1(\mathit{Timeliness}) + \phi_2(\mathit{Assets}) + \phi_3(\mathit{Leverage}) \\ & + \phi_4(\mathit{Accruals}) + \phi_5(\mathit{ROA}) + \varepsilon_j, \end{aligned} \tag{1}$$

where ϕ_0 is the intercept, ε_j represents a random error and $\{\varphi_1, \phi_2, \dots, \phi_5\}$ are the unknown regression functions that describe the marginal relationships between the response and the corresponding explanatory variables (φ_1 is denoted differently because it might contain jumps while other ϕ_j's are assumed to be continuous). *Timeliness* is calculated as the day difference between the 20-F filing date of current year and that of the previous year. For instance, suppose that a company filed its 20-F on April 30 this year and its last filing was on May 30 the year before, then the value for *Timeliness* is -30. *Assets*, calculated as the logarithm of the total assets at the fiscal year-end, is a proxy for firm size. There has been evidence in the literature (e.g., Atiase (1985); Freeman (1987)) that firm size is often associated with the magnitude of the market reaction. *Leverage*, the ratio of the total liability to the total assets, controls for the firm risk and the future growth opportunity. *Accruals*, calculated as the income minus the operating cash flows scaled by the total assets, is an indicator of financial performance. *ROA*, the ratio of the net income to the total assets, is a proxy for profitability. As discussed in Section 1, a substantial improvement of filing timeliness may cause a significant shift in the market response. There could be jumps in φ_1. Therefore, properly estimating the jump locations and the unknown regression functions $\{\varphi_1, \phi_2, \dots, \phi_5\}$ is the primary task of our empirical analysis.

3. Jump additive models. Motivated by our study of timeliness of 20-F filing, we propose a jump additive model and a two-step estimation methodology. They are described in the subsequent four subsections. Jump additive models in their own right are of interest in methodological research. Thus, we use general mathematical terms throughout this section.

3.1. *Review of additive models and ACE.* First, we give a brief review of additive models (AM) and alternating conditional expectation (ACE) algorithm proposed in Breiman and Friedman (1985). Let Y, $X_1, \ldots, X_p$ be random variables with Y the response and $X_1, \ldots, X_p$ the predictors. Assume that

$$Y = \phi_0 + \phi_1(X_1) + \cdots + \phi_p(X_p) + \varepsilon, \tag{2}$$

where $\phi_1(\cdot), \ldots, \phi_p(\cdot)$ are arbitrary measurable functions of the corresponding random variables and ε is a mean-zero random error and independent of $(X_1, \ldots, X_p)$. Without further restrictions on (2), the solution is not unique, since we can add any constant to one ϕ_j and subtract the same constant from some other ϕ_j. The standard convention is to further assume that $E[\phi_1(X_1)] = \cdots = E[\phi_p(X_p)] = 0$. Without loss of generality, we can also assume that $\phi_0 = 0$. In practice, this can be achieved by centering the response variable. The AM solves (2) for $\phi_1, \ldots, \phi_p$ by minimizing

$$e^2(\phi_1, \phi_2, \ldots, \phi_p) = E\left[Y - \sum_{j=1}^{p} \phi_j(X_j)\right]^2$$

holding $E(\phi_1) = \cdots = E(\phi_p) = 0$. The solutions are called *optimal transformations*. The ACE algorithm solves the above minimization problem through a series of bivariate conditional expectations.

ACE ALGORITHM.

(i) Set $\phi_1(X_1), \ldots, \phi_p(X_p) = 0$.
(ii) Iterate until $e^2(\phi_1, \ldots, \phi_p)$ fails to decrease:
For $j = 1, \ldots, p$, do

$$\phi_j \leftarrow E\left[Y - \sum_{k \neq j} \phi_k(X_k) | X_j\right].$$

The following theorem gives a convergence result for the ACE algorithm.

THEOREM 1. *Define, for $j = 1, \ldots, p$, $\phi_j^{(m)}$ after m stages of iteration in the ACE algorithm and $\widetilde{\phi}_m = \sum_{j=1}^{p} \phi_j^{(m)}$. If* (A1)–(A2) *in the Supplementary Material hold, then $\widetilde{\phi}_m \xrightarrow{L_2} P_X Y$ as $m \to \infty$.*

A similar result was obtained in Breiman and Friedman (1985) which also involves estimating the optimal transformation of the response. Since transformation of the response is not needed in our method, we give in the Supplementary Material a proof tailored for our purpose. Theorem 1 states that the ACE algorithm converges to the optimal transformations. The ACE algorithm is developed in the context of known distributions, that is, it assumes that the joint distribution of $(Y, X_1, \ldots, X_p)$ is known. In practice, population distributions are rarely known. Instead, one has a data set $\{(y_i, x_{i1}, \ldots, x_{ip}), 1 \leq i \leq n\}$ that are realizations from Y, $X_1, \ldots, X_p$. The conditional expectations in the ACE algorithm need to be replaced by suitable univariate nonparametric estimates. Estimates that are commonly used with the ACE

algorithm (e.g., Buja, Hastie and Tibshirani (1989); Hastie and Tibshirani (1990); Hastie, Tibshirani and Friedman (2009)) include kernel smoothing (Cheng and Lin (1981); Gasser and Müller (1979); Nadaraya (1964); Priestley and Chao (1972); Watson (1964)), local polynomial regression (Fan and Gijbels (1996); Hastie and Loader (1993)) and splines (Anselone and Laurent (1968); Demmler and Reinsch (1975); Kimeldorf and Wahba (1970); Wahba (1990)). These estimation methods assume that the regression functions $\{\phi_j, 1 \le j \le p\}$ are smooth and thus can not preserve the jumps well in cases when certain ϕ_j is discontinuous. In the next two subsections, we propose a two-step procedure that is jump-preserving.

3.2. *Jump detection*. Assume that $\{(y_i, x_{i1}, \dots, x_{ip}), 1 \le i \le n\}$ is a random sample from (2),

$$y_i = \varphi_1(x_{i1}) + \phi_2(x_{i2}) + \cdots + \phi_p(x_{ip}) + \varepsilon_i, \tag{3}$$

where φ_1 has jumps at $s_1 < \cdots < s_T$, $\phi_2, \dots, \phi_p$ are continuous and $x_{ij} \in [a_j, b_j]$ for $i = 1, \dots, n$, $j = 1, \dots, p$. Both the number of jumps T and the jump locations $\{s_t, 1 \le t \le T\}$ are unknown. Here, φ_1 is denoted differently to indicate the different smoothness assumption.

For a given $x_1^* \in [a_1 + h_1, b_1 - h_1]$, consider the following locally weighted least square problems:

$$\min_{c_0, c_1} \sum_{i=1}^{n} \left[y_i - c_0 - c_1(x_{i1} - x_1^*)\right]^2 K_-\left(\frac{x_{i1} - x_1^*}{h_1}\right), \tag{4}$$

$$\min_{c_0, c_1} \sum_{i=1}^{n} \left[y_i - c_0 - c_1(x_{i1} - x_1^*)\right]^2 K_+\left(\frac{x_{i1} - x_1^*}{h_1}\right), \tag{5}$$

where $h_1 \in (0, (b_1 - a_1)/2)$ is a bandwidth parameter and $K_\pm$ are one-sided kernel functions with the support included in $[0, 1]$ and $[-1, 0]$, respectively. The bandwidth controls the size of the local neighborhood. The kernel functions control the weights and are often chosen such that observations closer to x_1^* receive more weights. The solutions to c_0 in (4) and (5) are one-sided estimates of the following conditional expectation:

$$\begin{aligned} P(x_1^*) &= E\left[\varphi_1(X_1) + \phi_2(X_2) + \cdots + \phi_p(X_p) | X_1 = x_1^*\right] \\ &= \varphi_1(x_1^*) + \sum_{j=2}^{p} E\left[\phi_j(X_j) | X_1 = x_1^*\right]. \end{aligned}$$

By algebraic manipulations the solutions to (4) and (5) are

$$\widehat{P}_-(x_1^*) = \sum_{i=1}^{n} y_i W_{-,i} \quad \text{and} \quad \widehat{P}_+(x_1^*) = \sum_{i=1}^{n} y_i W_{+,i},$$

where

$$W_{\pm,i} = K_\pm\left(\frac{x_{i1} - x_1^*}{h_1}\right) \frac{w_{\pm,2} - w_{\pm,1}(x_{i1} - x_1^*)}{w_{\pm,0} w_{\pm,2} - w_{\pm,1}^2},$$

$$w_{\pm,u} = \sum_{i=1}^{n} (x_{i1} - x_1^*)^u K_\pm\left(\frac{x_{i1} - x_1^*}{h_1}\right), \quad u = 0, 1, 2.$$

Both $\widehat{P}_-(x_1^*)$ and $\widehat{P}_+(x_1^*)$ should estimate $P(x_1^*)$ well if $[x_1^* - h_1, x_1^* + h_1]$ does not contain any true jump position. On the other hand, if x_1^* is a true jump position, then $\widehat{P}_-(x_1^*)$ is a good estimator of $P_-(x_1^*)$ and $\widehat{P}_+(x_1^*)$ is a good estimator of $P_+(x_1^*)$, where $P_\pm(x_1^*)$ denote the one-sided limits of P at point x_1^*. Thus, $\widehat{P}_+(x_1^*) - \widehat{P}_-(x_1^*)$ should be close to

$d = P_+(x_1^*) - P_-(x_1^*)$. Suppose that the joint distribution function of $(X_1, \dots, X_p)$ is smooth. Then, $P_+(x_1^*) - P_-(x_1^*) = \varphi_{1,+}(x_1^*) - \varphi_{1,-}(x_1^*)$. As a result, $\widehat{P}_+(x_1^*) - \widehat{P}_-(x_1^*)$ is useful for detecting jump locations and estimating jump sizes.

Next, we identify the jump locations. Define

$$\xi_k = a_1 + (b_1 - a_1)\frac{k}{m}, \quad k = 1, \dots, m-1,$$

where $m = [(b_1 - a_1)/(2h_1)] + 1$ and $[x]$ denotes the integer part of x. Then, $\{\xi_k, 1 \le k \le m-1\}$ is sequence of points equally spaced in $[a_1, b_1]$, and the distance between any two neighboring points in $\{\xi_k, 1 \le k \le m-1\}$ is $(b_1 - a_1)/m$ which is slightly smaller than $2h_1$. For a given point x_1^*, there must be at least one and no more than two points in $\{\xi_k, 1 \le k \le m-1\}$ located in $(x_1^* - h_1, x_1^* + h_1)$. So we suggest detecting jumps at the points $\{\xi_k, 1 \le k \le m-1\}$ only. It saves much computation without missing any jumps. Suppose that $\xi_{k_1} < \cdots < \xi_{k_r}$ are the points in $\{\xi_k, 1 \le k \le m-1\}$ satisfying

$$\left| \frac{\widehat{P}_+(\xi_{k_l}) - \widehat{P}_-(\xi_{k_l})}{\sqrt{\sum_{i=1}^n W_{+,i}^2 + \sum_{i=1}^n W_{-,i}^2}} \right| > q, \quad l = 1, \dots, r, \tag{6}$$

where q is a threshold parameter and $\sqrt{\sum_{i=1}^n W_{+,i}^2 + \sum_{i=1}^n W_{-,i}^2}$ is the normalizing constant. Then, the points $\{\xi_{k_l}, 1 \le l \le r\}$ are flagged as jump candidates. But if ξ_{k_l} is flagged, some of its neighboring points in $\{\xi_k, 1 \le k \le m-1\}$ are likely to be flagged as well, even if they are actually continuity points. Thus, we suggest the following modification procedure to cancel some of the jump candidates. If there are two integers $r_1 < r_2$ such that

$$\begin{cases} \xi_{k_{l+1}} - \xi_{k_l} = \dfrac{1}{m} & l = r_1, r_1 + 1, \dots, r_2 - 1, \\ \xi_{k_{r_1}} - \xi_{k_{r_1 - 1}} > \dfrac{1}{m}, & \\ \xi_{k_{r_2+1}} - \xi_{k_{r_2}} > \dfrac{1}{m}, & \end{cases}$$

then we say that all candidates $\{\xi_{k_l}, r_1 \le l \le r_2\}$ form a *tie*. For the jump candidates in a tie, we replace all of them by a new candidate defined as $(\xi_{r_1} + \xi_{r_2})/2$. After this modification the current candidates consist of two types of points: those in $\{\xi_{k_l}, 1 \le l \le r\}$, which do not belong to any tie, and the middle points of all ties. Denote the current jump candidates by $\eta_1 < \eta_2 < \cdots < \eta_\tau$. We define τ, $\{\eta_t, 1 \le t \le \tau\}$ and $\{\widehat{d}_t = \widehat{P}_+(\eta_t) - \widehat{P}_-(\eta_t), 1 \le t \le \tau\}$ as the estimator of the number of jumps T, the jump locations $\{s_t, 1 \le t \le T\}$ and the jump sizes $\{d_t, 1 \le t \le T\}$, respectively.

3.3. *Jump-preserving backfitting*. Write

$$\varphi_1(x_1) = \phi_1(x_1) + \sum_{t=1}^T d_t \mathbb{1}(x_1 > s_t), \quad x_1 \in [a_1, b_1],$$

where ϕ_1 is a continuous function in $[a_1, b_1]$ and $\mathbb{1}(\cdot)$ is the indicator function which takes value of 1 if the argument in the parenthesis is true and takes value of 0 otherwise. The function ϕ_1 is called the *continuity part* of φ_1, and the summation $J(x_1) = \sum_{t=1}^T d_t \mathbb{1}(x_1 > s_t)$ is called the *jump part* of φ_1. The jump part has been estimated in Section 3.2. We now estimate the continuous functions $\{\phi_1, \phi_2, \dots, \phi_p\}$ from the new data

$$\left\{ \left(y_i - \sum_{t=1}^\tau \widehat{d}_t \mathbb{1}(x_{i1} > \eta_t), x_{i1}, \dots, x_{ip} \right), 1 \le i \le n \right\}.$$

Since the discontinuities are largely removed in the new data, the *backfitting* algorithm (Friedman and Stuetzle (1981)) for fitting the conventional additive models can be applied here. It consists of estimating each continuous function holding all the others fixed, then cycling through this process. More specifically, if the current estimates are $\widehat{\phi}_1, \dots, \widehat{\phi}_p$, then each is updated by smoothing the *partial residuals* $\nu_{ij} = \widehat{y}_i^* - \sum_{j' \neq j} \widehat{\phi}_{j'}(x_{ij'})$ against x_{ij}, where $\widehat{y}_i^* = y_i - \widehat{J}(x_{i1})$ with $\widehat{J}(x_{i1}) = \sum_{t=1}^{\tau} \widehat{d}_t \mathbb{1}(x_{i1} > \eta_t)$. The backfitting procedure is summarized below.

BACKFITTING ALGORITHM.

(i) Initialize: $\widehat{\phi}_1 \equiv -\frac{1}{n}\sum_{i=1}^n \widehat{J}(x_{i1})$, $\widehat{\phi}_j \equiv 0$ for $j = 2, \dots, p$.

(ii) Iterate until the functions $\widehat{\phi}_j$ change less than a prespecified threshold.

For $j = 1, \dots, p$ do:

$$\widehat{\phi}_j \leftarrow \mathcal{S}_j\left[\left\{\widehat{y}_i^* - \sum_{j' \neq j} \widehat{\phi}_{j'}(x_{ij'})\right\}_1^n\right],$$

$$\widehat{\phi}_j \leftarrow \begin{cases} \widehat{\phi}_j - \dfrac{1}{n}\displaystyle\sum_{i=1}^n \big(\widehat{\phi}_j(x_{ij}) + \widehat{J}(x_{i1})\big) & \text{if } j = 1, \\ \widehat{\phi}_j - \dfrac{1}{n}\displaystyle\sum_{i=1}^n \widehat{\phi}_j(x_{ij}) & \text{otherwise;} \end{cases}$$

End For Loop;

End Iteration Loop.

The second step of the above For Loop in (ii) of the backfitting algorithm is to ensure that $1/n\sum_{i=1}^n \widehat{\varphi}_1(x_{i1}) = 1/n\sum_{i=1}^n \widehat{\phi}_2(x_{i2}) = \cdots = 1/n\sum_{i=1}^n \widehat{\phi}_p(x_{ip}) = 0$, since it is assumed in (3) that $E[\varphi_1(X_1)] = E[\phi_2(X_2)] = \cdots = E[\phi_p(X_p)] = 0$. Above, $\mathcal{S}_j$ in the backfitting algorithm can be any appropriate univariate smoother; here, we consider local polynomial regression (Fan and Gijbels (1996)). Specifically, for a given j and any given $x_j^* \in [a_j + h_2, b_j - h_2]$, solve the following minimization problems:

$$\min_{g_0, g_1} \sum_{i=1}^n \big[\nu_{ij} - g_0 - g_1(x_{ij} - x_j^*) - \cdots - g_\kappa(x_{ij} - x_j^*)^\kappa\big]^2 K\left(\frac{x_{ij} - x_j^*}{h_2}\right), \tag{7}$$

where κ is the order of the local polynomial regression, $h_2 \in (0, \min_j(b_j - a_j)/2)$ is the bandwidth parameter and K is the kernel function with the support in $[-1, 1]$. The solution to g_0 in (7) is defined to be the κth order local polynomial estimate at x_j^*. In particular, procedure (7) is often called local constant kernel (LCK) smoothing and local linear kernel (LLK) smoothing, when $\kappa = 0$ and $\kappa = 1$, respectively.

After the estimates $\widehat{\phi}_1, \dots, \widehat{\phi}_p$ are obtained by the backfitting algorithm, the estimated regression function φ_1 is defined by

$$\widehat{\varphi}_1(x_1) = \widehat{\phi}_1(x_1) + \sum_{t=1}^{\tau} \widehat{d}_t \mathbb{1}(x_1 > \eta_t), \quad x_1 \in (a_1, b_1). \tag{8}$$

3.4. *Data-driven parameter selection.* There are three parameters to choose in our proposed procedure: h_1, q and h_2. We select them sequentially. Let us first consider the choice for h_2, assuming that the jump locations and jump sizes have been properly estimated. The purpose of h_2 is for estimating the continuous functions $\{\phi_1, \dots, \phi_p\}$ from data where discontinuities have been largely removed. Hence, h_2 plays the same role as a bandwidth parameter

does in fitting continuous additive models. In the literature on bandwidth selection in nonparametric regression, plug-in procedures (e.g., Gasser, Kneip and Köhler (1991); Loader (1999); Ruppert, Sheather and Wand (1995)) are easy to compute and often enjoy nice theoretical properties. Particularly for fitting conventional additive models using local polynomial smoothing, Opsomer and Ruppert (1998) proposed a direct plug-in (DPI) bandwidth selection method with the following formula:

$$h_{2,\mathtt{DPI}} = \left(\frac{(\kappa+1)(\kappa!)^2 R(K_{(\kappa)})\sigma^2 \sum_{j=1}^p (b_j - a_j)}{2n\mu_{\kappa+1}(K_{(\kappa)})^2 \sum_{j=1}^p \theta_{jj}(\kappa+1,\kappa+1)}\right)^{1/(2\kappa+3)}, \tag{9}$$

where

$$\theta_{jj}(\kappa_1,\kappa_2) = \frac{1}{n}\sum_{i=1}^n \left(\phi_j^{(\kappa_1)}(x_{ij}) - \bar{\phi}_j^{(\kappa_1)}\right) \times \left(\phi_j^{(\kappa_2)}(x_{ij}) - \bar{\phi}_j^{(\kappa_2)}\right)$$

with $\bar{\phi}_j^{(\kappa_r)} = (1/n)\sum_{i=1}^n \phi_j^{(\kappa_r)}(x_{ij})$ for any positive integer κ_r, $r = 1, 2$ and $j = 1, \dots, p$, and $R(K_{(\kappa)})$ and $\mu_{\kappa+1}(K_{(\kappa)})$ are quantities involving the kernel function K only. Their exact expressions are provided in the Supplementary Material. Opsomer and Ruppert (1998) showed that $h_{2,\mathtt{DPI}}$ achieves the optimal rate of convergence in the mean average squared error (MASE), defined by

$$\mathrm{MASE} = \frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[\left(\sum_{j=1}^p \widehat{\phi}_j(x_{ij}) - \sum_{j=1}^p \phi_j(x_{ij})\right)^2\right].$$

Because of this optimality, we adopt the DPI method for selecting h_2. For simplicity, we focus on the special case of (9) when LLK smoothing is used (i.e., $\kappa = 1$). The expression for $h_{2,\mathtt{DPI}}$ simplifies to

$$h_{2,\mathtt{DPI}} = \left(\frac{\sigma^2 R(K)\sum_{j=1}^p (b_j - a_j)}{n\mu_2(K)^2 \sum_{j=1}^p \theta_{jj}(2,2)}\right)^{1/5}, \tag{10}$$

where $R(K) = \int K(u)^2\,du$ and $\mu_2(K) = \int u^2 K(u)\,du$. There are several unknown quantities (i.e., $\theta_{jj}(2,2)$ and σ^2) in (10) to be estimated before it can be used directly in practice. $\theta_{jj}(2,2)$ involves the second order derivatives of ϕ_j. Following Opsomer and Ruppert (1998), we obtain pilot estimates $\{\widehat{\phi}_j^{(2)}(x_{ij})\}$ by applying the backfitting algorithm with local cubic smoothing to data $\{(\widehat{y}_i^*, x_{i1}, \dots, x_{ip}), i = 1, \dots, n\}$, in which an initial bandwidth $h_{2,0}$ is used. The performance of the DPI method is relatively insensitive to the choice of $h_{2,0}$. Our numerical experience suggests that $h_{2,0} = 0.75\max_j(b_j - a_j)$ works reasonably well. As a byproduct of fitting the backfitting algorithm to $\{(\widehat{y}_i^*, x_{i1}, \dots, x_{ip}), i = 1, \dots, n\}$ with local cubic smoothing, we can estimate σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left[\widehat{y}_i^* - \sum_{j=1}^p \widehat{\phi}_j(x_{ij})\right]^2.$$

Next, we consider the choice for q. Let $\phi_{j|1}(x_{i1}) = \mathrm{E}[\phi_j(X_j)|X_1 = x_{i1}]$ and $\delta_{ij} = \phi_j(x_{ij}) - \phi_{j|1}(x_{i1})$ for $j = 2, \dots, p$. Consequently, δ_{ij} and $\phi_j(X_j) - \mathrm{E}[\phi_j(X_j)|X_1]$ have the same distribution and $\{\delta_{ij}, i = 1, \dots, n\}$ are i.i.d. with mean 0. Denote the variance of δ_{ij}

by τ_j^2. Now, write

$$\begin{aligned}\widehat{P}_+(x_1^*) &= \sum_{i=1}^n \varphi_1(x_{i1})W_{+,i} + \sum_{j=2}^p\sum_{i=1}^n \phi_{j|1}(x_{i1})W_{+,i} \\ &\quad + \sum_{j=2}^p\sum_{i=1}^n \delta_{ij}W_{+,i} + \sum_{i=1}^n \varepsilon_i W_{+,i}. \\ &= P_+(x_1^*) + O(h_1^2) + \mathbf{1}'\sum_{i=1}^n W_{+i}\begin{pmatrix}\varepsilon_i \\ \boldsymbol{\delta}_i\end{pmatrix},\end{aligned} \tag{11}$$

where $\mathbf{1}$ is a vector of length p with all elements equal to 1 and $\boldsymbol{\delta}_i = (\delta_{i2}, \dots, \delta_{ip})'$. Under the null hypothesis that x_1^* is not a jump point, we have

$$\frac{\widehat{P}_+(x_1^*) - \widehat{P}_-(x_1^*)}{\sqrt{\sum_{i=1}^n (W_{+i}^2 + W_{-i}^2)}} \xrightarrow{d} N(0, \mathbf{1}'\Sigma\mathbf{1}), \tag{12}$$

where $\xrightarrow{d}$ denotes convergence in distribution and Σ is the covariance matrix of $(\varepsilon_i, \boldsymbol{\delta}_i')'$. The technical arguments for deriving (11) – (12) are provided in the Supplementary Material. Since ε_i is independent with $(X_{i1}, \dots, X_{ip})$, Σ has the following form:

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \tau_2^2 & \tau_{23} & \cdots & \tau_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \tau_{2p} & \tau_{3p} & \cdots & \tau_p^2 \end{pmatrix},$$

where $\tau_{j_1 j_2} = \mathrm{E}[\delta_{ij_1}\delta_{ij_2}]$ for $j_1, j_2 = 2, \dots, p$. Based on (12), q can be chosen to be $z_{1-\alpha/2}\sqrt{\mathbf{1}'\widehat{\Sigma}\mathbf{1}}$, where $\widehat{\Sigma}$ is an estimate for Σ, α is the significance level and $z_{1-\alpha/2}$ denotes the $100(1-\alpha/2)$ percentile of $N(0, 1)$. Commonly used α values include 0.05, 0.01, 0.001, etc. In our theoretical justifications in the Supplementary Material, the asymptotic condition (B5) is stated in terms of q for convenience. It can be equivalently stated in terms of α. That is, (B5) quantifies the rate of the type I error (i.e., α) approaching to zero as the sample size grows.

To obtain an estimate $\widehat{\Sigma}$, we suggest the following two-step procedure. First, obtain pilot estimates $\{\tilde{\varphi}_1, \tilde{\phi}_j, j = 2, \dots, p\}$ by applying the backfitting algorithm with local cubic smoothing to the original data $\{(y_i, x_{i1}, \dots, x_{ip}), i = 1, \dots, n\}$, where an initial bandwidth $h_{1,0}$ is used. Similar to $h_{2,0}$, the choice of $h_{1,0}$ is not critical, and we find that an initial bandwidth that covers about 75% data points works well. To mitigate the effect of possible jumps in φ_1, we suggest estimating σ^2 using a robust statistic, such as median absolute deviation (MAD). We adopt the MAD estimator $\tilde{\sigma} = 1.4826 \cdot \mathrm{med}_i\{|\tilde{\varepsilon}_i - \mathrm{med}_k\{\tilde{\varepsilon}_k\}|\}$ (Rousseeuw and Croux (1993)), where med denotes the sample median and $\{\tilde{\varepsilon}_i\}$ are the residuals from the pilot estimates, that is, $\tilde{\varepsilon}_i = y_i - \tilde{\varphi}_1(x_{i1}) - \sum_{j=2}^p \tilde{\phi}_j(x_{ij})$. Second, obtain residuals $\{\widehat{\delta}_{ij}\}$ by applying LLK smoothing to $\{(x_{i1}, \tilde{\phi}_j(x_{ij})), i = 1, \dots, n\}$ for each $j = 2, \dots, p$, and then estimate

$$\widehat{\tau}_{j_1 j_2} = \frac{1}{n}\sum_{i=1}^n (\widehat{\delta}_{ij_1} - \overline{\widehat{\delta}}_{\cdot j_1})(\widehat{\delta}_{ij_2} - \overline{\widehat{\delta}}_{\cdot j_2}),$$

where $\overline{\widehat{\delta}}_{\cdot j_r} = (1/n)\sum_{i=1}^n \widehat{\delta}_{ij_r}$ for $r = 1, 2$.

Next, we consider the choice of h_1 which is used for jump detection. Let $S = \{s_t, t = 1, \dots, T\}$ denote the set of the true jump locations. Recall that, for a given h_1, we only

examine (6) at locations $\{\xi_k = a_1 + (b_1 - a_1)k/m : k = 1, \dots, m-1\}$. Let $\tilde{S} = \{\xi_k : \xi_k = \operatorname{argmin}_{l=1,\dots,m-1}|s_t - \xi_l|, t = 1, \dots, T\}$. Using the method described in Section 3.2, we obtain an estimate $\widehat{S}$. The Hausdorff distance between $\tilde{S}$ and $\widehat{S}$, $d_H(\tilde{S}, \widehat{S})$ (see the Appendix for the formal definition) takes discrete values in $\{(b_1 - a_1)k/m : k = 0, \pm 1, \pm 2, \dots\}$. Ideally, $d_H(\tilde{S}, \widehat{S}) = 0$ with high probability. In the literature, Gijbels and Goderniaux (2004) proposes a bootstrap procedure to estimate $P(d_H(\tilde{S}, \widehat{S}) = 0)$ and to select the bandwidth for which this estimated probability is maximal. Following Gijbels and Goderniaux (2004), we select h_1 as follows:

- Step 1: *Computation of residuals*. For a given h_1, with q and h_2 calculated by the aforementioned procedures, construct the estimates $\widehat{S}$ and $\{\widehat{\varphi}_1, \widehat{\phi}_2, \dots, \widehat{\phi}_p\}$. Define $\widehat{\varepsilon}_i = y_i - \widehat{\varphi}_1(x_{i1}) - \sum_{j=2}^{p} \widehat{\phi}_j(x_{ij})$.
- Step 2: *Bootstrap simulation*. Let $\varepsilon_1^{(1)}, \dots, \varepsilon_n^{(1)}$ be a resample randomly drawn with replacement from the set $\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_n$. Define

$$y_i^{(1)} = \widehat{\varphi}_1(x_{i1}) + \sum_{j=2}^{p} \widehat{\phi}_j(x_{ij}) + \varepsilon_i^{(1)}.$$

  Then, $\{(y_i^{(1)}, x_{i1}, \dots, x_{ip}), i = 1, \dots, n\}$ is the first bootstrap sample.
- Step 3: *Determination of the bootstrap probability*. Compute the analog $\widehat{S}^{(1)}$ for the first bootstrap sample. From B bootstrap replications, obtain B values of $\{\widehat{S}^{(b)} : 1 \le b \le B\}$, and estimate the discrete probability $P(d_H(\tilde{S}, \widehat{S}) = 0)$ via

$$\widehat{P}\big(d_H(\tilde{S}, \widehat{S}) = 0\big) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{d_H(\widehat{S}, \widehat{S}^{(b)}) = 0\}.$$

- Step 4: *Determination of* h_1. For a grid of potential values for h_1, choose the value for which $\widehat{P}(d_H(\tilde{S}, \widehat{S}) = 0)$ is maximum.

The theoretical underpinning for the bootstrap procedure is established in Gijbels, Hall and Kneip (2004). More specifically, by the same arguments in Theorem 2 in Gijbels, Hall and Kneip (2004), for all the discrete probabilities $\widehat{P}(d_H(\tilde{S}, \widehat{S}) = (b_1 - a_1)k/m)$, $k = 0, \pm 1, \pm 2, \dots$, we have

$$\sup_{k=0,\pm 1,\dots} \left| \widehat{P}\left(d_H(\tilde{S}, \widehat{S}) = \frac{(b_1 - a_1)k}{m}\right) - P\left(d_H(\tilde{S}, \widehat{S}) = \frac{(b_1 - a_1)k}{m}\right) \right| \to 0$$

in probability.

The bootstrap procedure is computationally intensive since it requires repeating our proposed procedure many times. To reduce running time, we employ parallel computing in the implementation of the companion R package *jam* which is included in the Supplementary Material (Kang (2020)).

4. Empirical analysis. In this section we apply our proposed procedure to analyzing timeliness of the 20-F filings.

4.1. *Data*. Our sample period covers 2008–2011 because the SEC required all the FPIs to follow the new four-month filing deadline starting from the fiscal year ending on or after December 15, 2008. The SEC also provided a transition period of three years which allowed FPIs to comply with this new rule effective for the fiscal year ending on or after December 15, 2011. Following the data selection method in Doyle and Magilke (2013) and Liu (2016), we obtained 20-Fs and 20-F filing dates from the SEC EDGAR system, daily stock returns and

index data from the CRSP U.S. stock database, firm financial data from the COMPUSTAT database and firm major operation country, filer status and accounting standard from the Audit Analytics database. Sample firms were initially obtained from the SEC's official annual summary of "International Registered and Reporting Companies." Then, we deleted the firms not listed on either the NASDAQ or the NYSE, the Canadian firms (because of the special filing requirements for Canadian firms), over-the-counter firms, debt issuing firms and the firms that lack complete data during the test period. We identified 272 filings in the final sample.

4.2. *Parameter values*. Since the predictors are of different units (e.g., *Assets* are denominated in U.S. dollars and *Leverage* ratio is a fraction), each predictor is centered and standardized before our procedure is applied. Using the method described in Section 3.4, the following grid of values is specified for selecting $h_1, : \{k/\mathrm{SE}_1 : k = 2, 3, \dots, 12\}$, where SE_1 denotes the sample standard deviation of *Timeliness*. h_1 is chosen to be $6/\mathrm{SE}_1 = 0.22$ by our bootstrap procedure, where $B = 1000$. The asymptotic variance in (12) is estimated to be $\mathbf{1}'\widehat{\Sigma}\mathbf{1} = 0.0035$, and the plug-in $h_{2,\mathtt{DPI}}$ is calculated to be 0.81.

4.3. *Results*. The jump additive model (1), estimated by the proposed procedure, is shown in Figure 1. Jump detection criterion (6) is plotted in Figure 1(a) which suggests a possible jump location around where *Timeliness* is -18. From Figure 1(b) it can be seen that the magnitude of market reaction (i.e., the absolute value of cumulative abnormal return) increases moderately as firms improve their filing timeliness by less than 17 days (the more negative the *Timeliness* value, the timelier the filing). When firms accelerate their filing by 18 days, there is an abrupt shift in the direction of market response, indicating that the market changes its directional view. This is possibly because market investors consider such filings significantly more relevant and the financial information disclosed in these 20-Fs is then quickly priced in by the market, triggering the change of view. Nonetheless, when firms further accelerate their filings by more than 40 days, we find little market reaction. One explanation is that timelier filing reduces the preparation time and results in more errors in financial information. The market may view filings with too much acceleration not as reliable. Hence, there is a trade-off between timeliness and reliability. Taken together, the market reaction is the most significant when the filing is accelerated by a little more than 18 days. As for the control variables, it can be seen from Figure 1(c)–(f) that firms with more assets (larger firms), healthy leverage ratio (good growth potential but not deeply in debt) and higher ROA (more profitable) tend to associate with more positive market reaction. This is consistent with our intuition. Confidence bands have also been added to Figure 1(b)–(f). There are several ways to obtain confidence bands after the jump part has been removed. One way is to follow the F-distribution approximation approach suggested by Hastie and Tibshirani (1990). It needs to assume that the errors follow a normal distribution. An alternative approach is to create bootstrap residuals as in Section 3.4 and compute the backfitting estimates of the bootstrap response on the predictors. The bootstrap approach is adopted here for producing the confidence bands in Figure 1, where the number of bootstrap samples is chosen to be 1000.

5. Simulation study. We present some simulation results regarding the numerical performance of the jump detection procedure, the backfitting procedure and the parameter selection procedure. Throughout this section the kernel function K in (7) is chosen to be the Epanechnikov kernel $K(x) = 0.75(1 - x^2)\mathbb{1}(|x| < 1)$. The one-sided kernels $K_\pm$ in (4)–(5) are chosen to be $K_-(x) = 2K(x)\mathbb{1}(x < 0)$ and $K_+(x) = 2K(x)\mathbb{1}(x > 0)$. In the local kernel smoothing literature the Epanechnikov kernel is often used because of its good theoretical

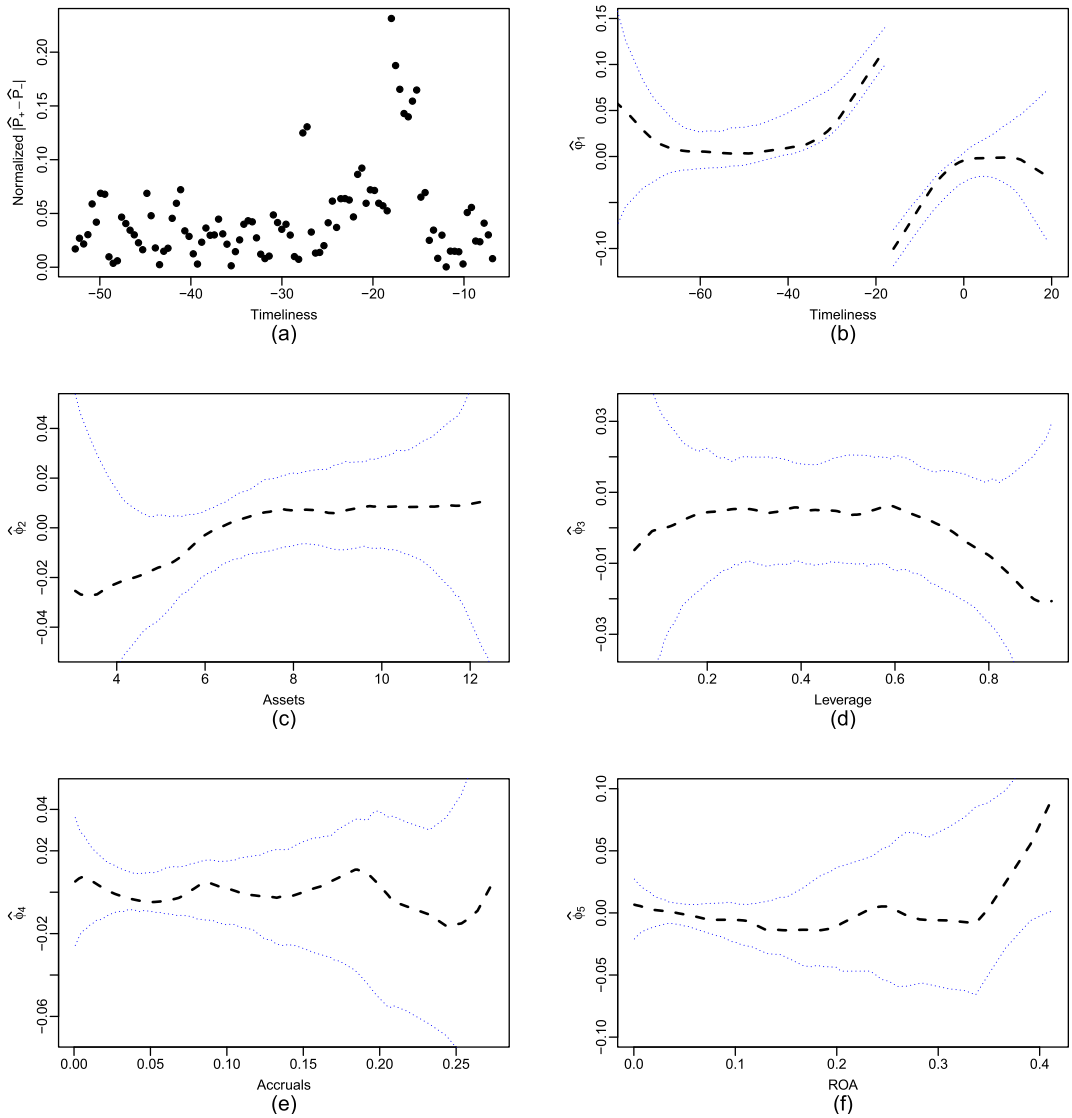

FIG. 1. *The jump additive model* (1) *estimated by the proposed procedure.* (*a*): *Plot of the jump detection criterion.* (*b*)–(*f*): *The estimated regression function* (*dashed line*) *and* 99% *confidence bands* (*dotted line*) *of* $\varphi_1, \phi_2, \dots, \phi_5$, *respectively.*

properties (Fan and Gijbels (1996)). In our simulation the data is generated from the following model:

$$Y = 0.2 + \varphi_1(X_1) + \phi_2(X_2) + \phi_3(X_3) + \varepsilon,$$

$$\varphi_1(x_1) = e^{-x_1} - 0.2 \times \mathbb{1}(x_1 > 0.5) - c_1,$$

$$\phi_2(x_2) = 0.5x_2^3 - c_2, \qquad \phi_3(x_3) = 0.5x_3^3 - c_3,$$

$$X_1 \stackrel{d}{=} U, \qquad X_2 \stackrel{d}{=} \rho U + (1-\rho)V_2, \qquad X_3 \stackrel{d}{=} \rho U + (1-\rho)V_3,$$

where $\stackrel{d}{=}$ denotes equality in distribution, U, V_2 and V_3 are independent random variable with identical distribution uniform[0, 1], c_1, c_2 and c_3 are constants such that $E[\varphi_1(X_1)] = E[\phi_2(X_2)] = E[\phi_2(X_2)] = 0$, $\rho = 0.3$ (it implies that $\mathrm{corr}(X_1, X_2) = \mathrm{corr}(X_1, X_3) \approx 0.4$)

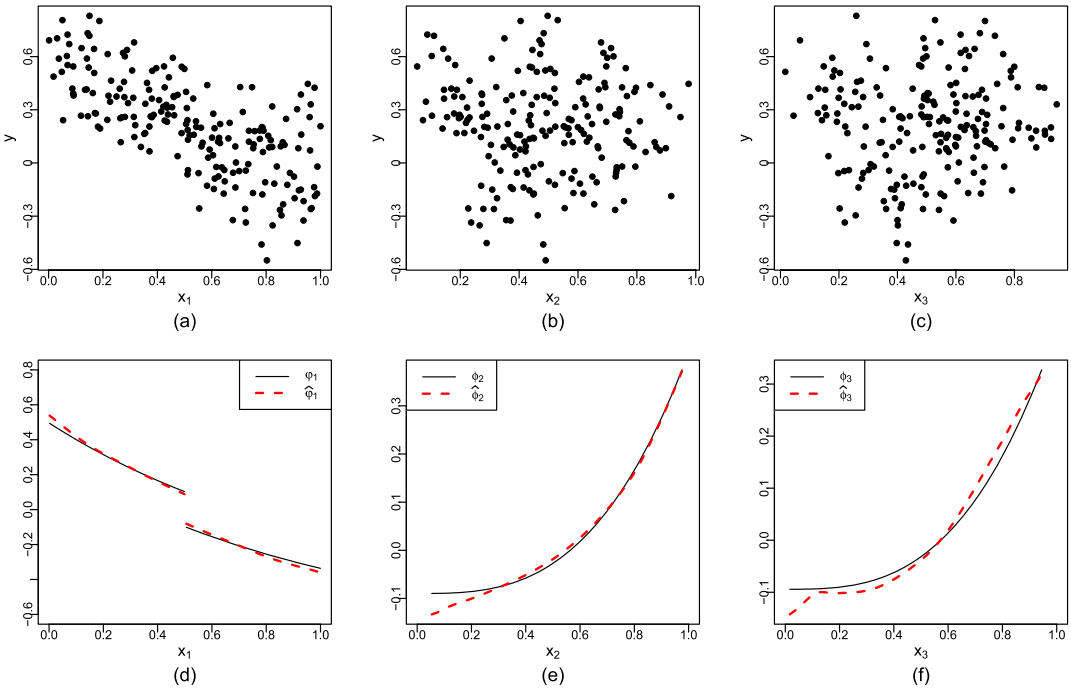

FIG. 2. (*a*)–(*c*): *Scatterplot of* Y *vs.* X_1, X_2 *and* X_3, *respectively.* (*d*)–(*f*): *Plot of* φ_1 *and* $\widehat{\varphi}_1$, ϕ_2 *and* $\widehat{\phi}_2$ *and* ϕ_3 *and* $\widehat{\phi}_3$, *respectively.*

and the random noise ε are generated from normal distribution $N(0, \sigma^2)$. We consider the cases when the sample size $n = 200$, 500 and the noise level $\sigma = 0.1$, 0.15, 0.2. Throughout this section the grid of values for h_1 is specified to be $\{0.1 + 0.01 \cdot k : k = 1, \dots, 49\}$. We find the grid with step size of 0.01 fine enough for our example, and the largest permitted should not exceed 0.5 which is half the length of the design interval for X_1. In addition, the number of bootstrap samples $B = 100$, and the significance level $\alpha = 0.05$. Let

$$h_{1,\mathtt{opt}} = \operatorname{argmax} P\big(d_H(\tilde{S}, \widehat{S}) = 0\big),$$

where the probability is evaluated based on 100 replicated simulations. Similarly, let $h_{2,\mathtt{opt}}$ denote the *optimal* h_2 for which the MASE is minimized. Furthermore, let $h_{1,\mathtt{bt}}$ and $h_{2,\mathtt{DPI}}$ denote the bandwidths selected by the bootstrap procedure and the plug-in procedure, respectively. The realizations of $\{(Y, X_1, X_2, X_3)\}$ when $n = 200$ and $\sigma = 0.15$ are shown in Figure 2(a)–(c). Plots of φ_1, ϕ_2 and ϕ_3 along with their estimates are shown, respectively, in Figure 2(d)–(f), where $h_{1,\mathtt{opt}}$ and $h_{2,\mathtt{DPI}}$ were used. It can be seen that: (i) the data is quite noisy that the jump structure can hardly been visually detected in the scatterplots, and (ii) the proposed procedure is able to preserve the jump well and largely eliminate the noise. It should also be pointed out that the nominal noise level σ is not the effective noise level for our jump detector (6). As shown in (12), the effective noise level is $\sqrt{\mathbf{1}'\Sigma\mathbf{1}}$ which is greater than σ. It explains why the jump structure can hardly be seen in Figure 2(a), where the jump size is 1.33 times the magnitude of the nominal noise level.

Next, we evaluate the numerical performance of the proposed procedure quantitatively. Table 1 summarizes the simulation results of the estimation procedure and the parameter selection. The columns denoted by *optimal* are associated with the cases when the optimal parameters are used (i.e., $h_{1,\mathtt{opt}}$ and $h_{2,\mathtt{opt}}$). The columns denoted by *data-driven* are associated with the cases when the parameters are chosen by the data-driven procedure (i.e., $h_{1,\mathtt{bt}}$ and $h_{2,\mathtt{DPI}}$). Pr denotes $P(d_H(\tilde{S}, \widehat{S}) = 0)$, and sePr denotes the standard error of

TABLE 1

Numerical performance of the proposed procedure and the parameter selection. The columns denoted by `optimal` *are associated with the cases when the optimal parameters are used. The columns denoted by* `data-driven` *are associated with the cases when the parameters are chosen by the data-driven procedure. Pr denotes* $P(d_H(\tilde{S}, \widehat{S}) = 0)$ *and sePr denotes the standard error of* $P(d_H(\tilde{S}, \widehat{S}) = 0)$. *seMASE denotes the standard error of the MASE. The numbers in rows of MASE and seMASE are in the unit of* 10^{-4}. *The results are based on* 100 *replicated simulations*

n		$\sigma = 0.1$		$\sigma = 0.15$		$\sigma = 0.2$	
		optimal	data-driven	optimal	data-driven	optimal	data-driven
200	h_1	0.48	0.43	0.46	0.44	0.50	0.45
	Pr	0.93	0.90	0.79	0.77	0.68	0.67
	sePr	0.026	0.030	0.041	0.042	0.047	0.047
	h_2	0.40	0.23	0.44	0.23	0.47	0.23
	MASE	14.43	15.44	25.29	28.01	39.02	44.05
	seMASE	4.23	4.49	5.49	5.91	7.55	8.21
500	h_1	0.29	0.34	0.41	0.40	0.40	0.37
	Pr	1.00	1.00	0.98	0.97	0.95	0.92
	sePr	0.00	0.00	0.014	0.017	0.022	0.027
	h_2	0.30	0.22	0.34	0.23	0.37	0.24
	MASE	3.70	3.92	6.70	7.08	10.09	10.99
	seMASE	0.43	0.46	1.99	2.05	2.45	2.56

$P(d_H(\tilde{S}, \widehat{S}) = 0)$. seMASE denotes the standard error of the MASE. The numbers in rows of MASE and seMASE are in the unit of 10^{-4}. The results are based on 100 replicated simulations. It can be seen from Table 1 that both MASE and seMASE decrease as n increases and as σ decreases. Similarly, $P(d_H(\tilde{S}, \widehat{S}) = 0)$, which measures the accuracy of our jump detection procedure, increases as n increases and as σ decreases. This is intuitively reasonable and consistent with our theoretical results included in the Appendix. In addition, across various cases of sample size and noise level, our data-driven parameter selection works well in the sense that both $P(d_H(\tilde{S}, \widehat{S}) = 0)$ and MASE, calculated using $h_{1,bt}$ and $h_{2,\mathrm{DPI}}$, are not significantly different from those calculated using the optimal parameters. However, it is worth noting that, in cases when the sample size is moderate (i.e., $n = 200$), our procedure performs reasonably well when the nominal noise level σ is half the jump size, but its performance starts to deteriorate as σ is of the same magnitude as the jump size. Thus, we consider our procedure better at detecting large jumps.

6. Conclusions. In this study we investigate the relationship between the filing timeliness of 20-F forms and the market reaction. We propose a novel jump additive model that accounts for possible discontinuities in the relationship which includes jumps in the magnitude and shifts in the direction of market reaction. We also propose a two-step jump-preserving estimation procedure and show that it consistently estimates the number of jump positions, the jump locations, the jump sizes and the regression functions. A major feature of our proposed procedure is that it does not impose restrictive parametric assumptions, so it is flexible to use. A data-driven parameter selection procedure is suggested as well. Therefore, the proposed procedure can be fully automated. By applying the jump additive model to our 20-F study, we find that the magnitude of the market reaction increases moderately when the filing timeliness is improved by less than 17 days. Market investors consider the 20-Fs significantly more relevant when the filings are accelerated by a little more than 18 days. However, as firms further accelerate their filings by more than 40 days, the market reaction tapers of,f possibly

because of concerns that filings with too much acceleration are prone to errors and thus less reliable.

There is still room for further improvement. Methodologically, our estimation procedure assumes that there is only one regression function in the model with possible discontinuities. Removing such restriction is certainly an interesting direction to pursue. In addition, we have assumed that random errors ε_i are i.i.d. Modifying our method to accommodate heteroscedasticity in errors is a useful extension. To further investigate the costs and benefits of shortened 20-F filing, it requires to examine the relationship between the filing timeliness and the reporting reliability and to determine whether such a relationship plays a role in diminishing market reactions to filings with large accelerations.

APPENDIX: ASYMPTOTIC PROPERTIES

We first establish the almost sure consistency of the jump detection procedure. Define D to be a set $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ of n points in p-dimensional space, that is, $\boldsymbol{x}_k = (x_{k1}, \dots, x_{kp})$. Let $\mathscr{D}_n$ be the collection of all such D. We have the following theorem.

THEOREM 2. *Assume that* (B1)–(B6) *in the Supplementary Material* (*Kang* (2020)) *hold. Then, for any given* $D \in \mathscr{D}_n$:

1. $d_H(S, \widehat{S}) = O(h_1)$ *a.s., where* $d_H(A, B)$ *denotes the Hausdorff distance between two point sets* A *and* B, *as defined by*

$$d_H(A, B) = \max\Big\{\sup_{a\in A}\inf_{b\in B} d_E(a, b), \sup_{b\in B}\inf_{a\in A} d_E(b, a)\Big\},$$

   with $d_E(\cdot, \cdot)$ *being the Euclidean distance. Furthermore, when* n *is sufficiently large, there is exactly one detected jump in each* $(s_t - h_1, s_t + h_1)$, $t = 1, \dots, T$.
2. $|\widehat{d}_t - d_t| = o(1)$ *a.s. for* $t = 1, \dots, T$.

Theorem 2 states that τ is an almost surely consistent estimator of T. Additionally, the estimators of jump locations and jump magnitudes are all almost surely consistent. Next, we give the consistency result for the backfitting algorithm.

THEOREM 3. *Let* $\widehat{\phi}_j(\cdot; m)$, $j = 1 \dots, p$ *denote the functions after repeating the backfitting loops* m *times with the LCK smoother. Repeat the ACE loop* m *times, getting functions denoted by* $\phi_j(\cdot; m)$, $j = 1, \dots, p$. *Under the same conditions as those in Theorem 2, for any small number* $\epsilon > 0$, *as* $n \to \infty$,

$$E\big[\|\widehat{\phi}_j(\cdot; m) - \phi_j(\cdot; m)\|^2_{[a_j, b_j], n} | X_1 \notin S_\epsilon, D\big] \to 0, \quad j = 1, \dots, p,$$

where $S_\epsilon = \bigcup_{t=1}^T [s_t - \epsilon, s_t + \epsilon]$, *and* $\|\widehat{\phi}_j(\cdot; m) - \phi_j(\cdot; m)\|^2_{I,n} = 1/n \sum_{x_{i,j}\in I} |\widehat{\phi}_j(x_{i,j}; m) - \phi_j(x_{i,j}; m)|^2$ *for* $j = 1, \dots, p$ *with* I *being an interval.*

The proofs of Theorem 2 and Theorem 3 are given in the Supplementary Material (Kang (2020)).

Acknowledgments. The author would like to thank the Editor, an Associate Editor and two referees for many constructive comments and suggestions which substantially improved the quality of the paper.

SUPPLEMENTARY MATERIAL

Proofs of the theoretical results (DOI: 10.1214/20-AOAS1365SUPPA; .pdf). This supplement includes proofs of the theoretical results.

The companion R package *jam* (DOI: 10.1214/20-AOAS1365SUPPB; .zip). The R package implements the jump additive models proposed in this article. More specifically, the function *jam.fit*() estimates the regression functions and the function *jam.par*() selects the model parameters by the data-driven approach. It also includes the data set used in the simulation study in Section 5.

REFERENCES

ANSELONE, P. M. and LAURENT, P. J. (1968). A general method for the construction of interpolating or smoothing spline-functions. *Numer. Math.* **12** 66–82. MR0249904 https://doi.org/10.1007/BF02170998

ASTHANA, S., BALSAM, S. and SANKARAGURUSWAMY, S. (2004). Differential response of small versus large investors to 10-K filings on EDGAR. *Account. Rev.* **79** 571–589.

ATIASE, R. K. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *J. Acc. Res.* 21–36.

AVANESOV, V. and BUZUN, N. (2018). Change-point detection in high-dimensional covariance structure. *Electron. J. Stat.* **12** 3254–3294. MR3861282 https://doi.org/10.1214/18-EJS1484

BEAVER, W. H. (1968). The information content of annual earnings announcements. *J. Acc. Res.* 67–92.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–619. MR0803258

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115

CALLEN, J. L., LIVNAT, J. and SEGAL, D. (2006). The information content of SEC filings and information environment: A variance decomposition analysis. *Account. Rev.* **81** 1017–1043.

CASINI, A. (2018). Tests for forecast instability and forecast failure under a continuous record asymptotic framework. Technical report. Available at arXiv:1803.10883.

CASINI, A. (2019). Theory of Evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Technical report, Dept. Economics and Finance, Univ. Rome Tor Vergata.

CASINI, A. and PERRON, P. (2019). Structural breaks in time series. *Oxford Research Encyclopedia of Economics and Finance*. https://doi.org/10.1093/acrefore/9780190625979.013.179

CHEN, L. H. and SAMI, H. (2008). Trading volume reaction to the earnings reconciliation from IAS to US GAAP. *Contemp. Account. Res.* **25** 15–53.

CHEN, L. H. and SAMI, H. (2013). The impact of firm characteristics on trading volume reaction to the earnings reconciliation from IFRS to US GAAP. *Contemp. Account. Res.* **30** 697–718.

CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **57** 223–233. MR0626817 https://doi.org/10.1007/BF00535491

CHO, H. and FRYZLEWICZ, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 https://doi.org/10.1111/rssb.12079

CHRISTENSEN, T. E., HENINGER, W. G. and STICE, E. K. (2013). Factors associated with price reactions and analysts' forecast revisions around SEC filings. *Res. Account. Regul.* **25** 133–148.

DEMMLER, A. and REINSCH, C. (1975). Oscillation matrices with spline smoothing. *Numer. Math.* **24** 375–382. MR0395161 https://doi.org/10.1007/BF01437406

DE FRANCO, G., WONG, M. F. and ZHOU, Y. (2011). Accounting adjustments and the valuation of financial statement note information in 10-k filings. *Account. Rev.* **86** 1577–1604.

DOYLE, J. T. and MAGILKE, M. J. (2013). Decision usefulness and accelerated filing deadlines. *J. Acc. Res.* **51** 549–581.

ETTER, E. R., REES, L. and LUKAWITZ, J. M. (1999). The usefulness to individual and institutional investors of annual earnings announcements and SEC filings by non-US companies. *J. Int. Account. Audit. Tax.* **8** 109–131.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

FREEMAN, R. N. (1987). The association between accounting earnings and security returns for large and small firms. *J. Account. Econ.* **9** 195–228.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. MR0650892

GASSER, T., KNEIP, A. and KÖHLER, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* **86** 643–652. MR1147088

GASSER, T. and MÜLLER, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (*Proc. Workshop*, *Heidelberg*, 1979). *Lecture Notes in Math.* **757** 23–68. Springer, Berlin. MR0564251

GIJBELS, I. and GODERNIAUX, A.-C. (2004). Bandwidth selection for changepoint estimation in nonparametric regression. *Technometrics* **46** 76–86. MR2043389 https://doi.org/10.1198/004017004000000130

GIJBELS, I., HALL, P. and KNEIP, A. (2004). Interval and band estimation for curves with jumps *J. Appl. Probab.* **41A** 65–79. MR2057566 https://doi.org/10.1239/jap/1082552191

GRIFFIN, P. A. (2003). Got information? Investor response to form 10-K and form 10-Q EDGAR filings. *Rev. Acc. Stud.* **8** 433–460.

HALL, P. and TITTERINGTON, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34** 429–440. MR1190262 https://doi.org/10.2307/1268942

HASTIE, T. and LOADER, C. (1993). Local regression: Automatic kernel carpentry. *Statist. Sci.* **8** 120–129.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

KANG, Y. (2020a). Consistent blind image deblurring using jump-preserving extrapolation. *J. Comput. Graph. Statist.* **29** 372–382. MR4116049 https://doi.org/10.1080/10618600.2019.1665536

KANG, Y. (2020). Supplement to "Measuring timeliness of annual reports filing by jump additive models." https://doi.org/10.1214/20-AOAS1365SUPPA, https://doi.org/10.1214/20-AOAS1365SUPPB

KANG, Y., GONG, X., GAO, J. and QIU, P. (2019). Errors-in-variables jump regression using local clustering. *Stat. Med.* **38** 3642–3655. MR3999222 https://doi.org/10.1002/sim.8205

KIM, Y., LI, H. and LI, S. (2012). Does eliminating the form 20-F reconciliation from IFRS to US GAAP have capital market consequences? *J. Account. Econ.* **53** 249–270.

KIM, O. and VERRECCHIA, R. E. (1991). Trading volume and price reactions to public announcements. *J. Acc. Res.* 302–321.

KIMELDORF, G. S. and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat.* **41** 495–502. MR0254999 https://doi.org/10.1214/aoms/1177697089

LIU, Z. (2016). Effect of Shortened Reporting Lag on the Usefulness of Form 20-F Ph.D. thesis. Florida International Univ.

LOADER, C. R. (1999). Bandwidth selection: Classical or plug-in? *Ann. Statist.* **27** 415–438. MR1714723 https://doi.org/10.1214/aos/1018031201

MA, S. and YANG, L. (2011). A jump-detecting procedure based on spline estimation. *J. Nonparametr. Stat.* **23** 67–81. MR2780816 https://doi.org/10.1080/10485250903571978

MEEK, G. K. (1983). US securities market responses to alternate earnings disclosures of non-US multinational corporations. *Account. Rev.* 394–402.

MÜLLER, C. H. (2002). Robust estimators for estimating discontinuous functions *Metrika* **55** 99–109. MR1903286 https://doi.org/10.1007/s001840200190

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

OLIBE, K. O. (2001). Assessing the usefulness of SEC form 20-F disclosures using return and volume metrics: The case of UK firms. *J. Econ. Finance* **25** 343–357.

OPSOMER, J. D. and RUPPERT, D. (1998). A fully automated bandwidth selection method for fitting additive models. *J. Amer. Statist. Assoc.* **93** 605–619. MR1631333 https://doi.org/10.2307/2670112

PRIESTLEY, M. B. and CHAO, M. T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392. MR0331616

QIU, P. (2005). *Image Processing and Jump Regression Analysis. Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR2111430 https://doi.org/10.1002/0471733156

ROUSSEEUW, P. J. and CROUX, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88** 1273–1283. MR1245360

RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270. MR1379468

WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397. MR1354236 https://doi.org/10.1093/biomet/82.2.385

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372. MR0185765

WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

XIA, Z. and QIU, P. (2015). Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika* **102** 397–408. MR3371012 https://doi.org/10.1093/biomet/asv018

YOU, H. and ZHANG, X.-J. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Rev. Acc. Stud.* **14** 559–586.