Adaptive Trend Filtering of High Frequency Price Processes with Exogenous Covariates

Jason Bohne[1], Paweł Polak[1], El Mehdi Ainasse [2], Gary Kazantsev [2]

[1]Stony Brook University [2] Bloomberg L.P.

Stony Brook University

Abstract

Trend filtering with total variation regularization is a statistical learning method for estimating the hidden signal with (non)-linear components. We propose an **adaptive and interpretable extension** of trend filtering specifically designed for high-frequency financial time series. This approach simultaneously estimates **global and local price fluctuations** by dynamically adapting to external factors such as volume, spread, news, and sentiment, resulting in more accurate learning of the price process. Under weak assumptions, the proposed estimator can be shown to be **Minimax optimal** over a general class of finite total variation functions. A direct application of this work is the systematic **annotation of changepoints** in the dynamics of the price process. Future work will explore the utility of estimates and annotations in designing **relative positional embeddings** for the financial domain.

LASSO and ℓ_1-Trend Filtering

Classical LASSO estimates are defined as the solution to the convex optimization problem of

$$\widehat{\beta}_{LASSO}(\lambda) = \arg\min_{\beta} \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1 \quad (1)$$

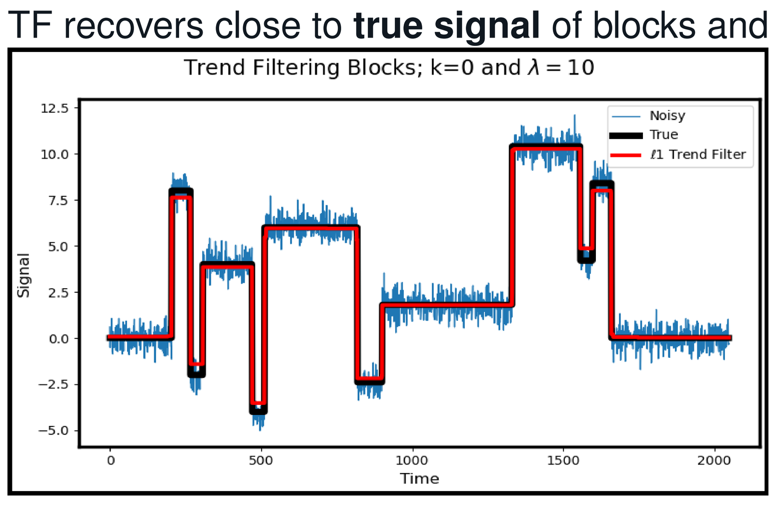
where $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$, with $\lambda \geq 0$ as a hyper parameter specifying the strength of regularization. A special case of the Lasso is the **Gaussian signal approximation** problem with estimates defined as the solution to equation 2 with penalty operator $J(\cdot)$ and hyperparameter $\lambda \geq 0$

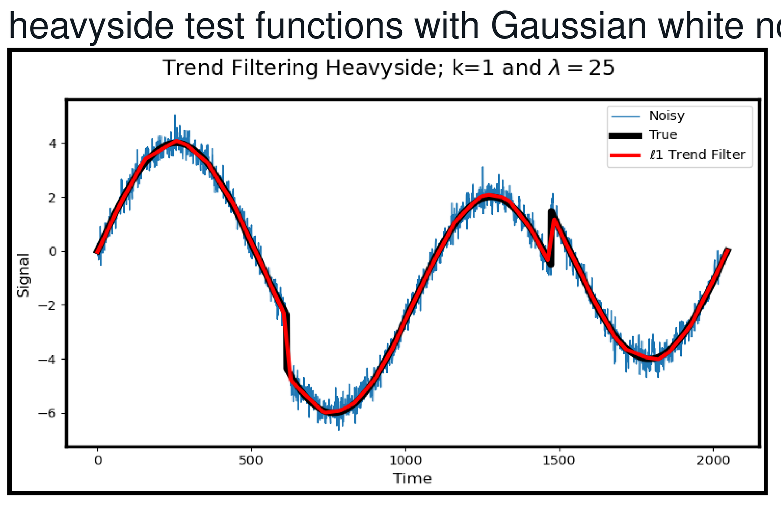
$$\widehat{\beta}_{GSA}(\lambda) = \arg\min_{\beta} \|\mathbf{Y} - \beta\|_2^2 + \lambda\|J(\beta)\|_1 \quad (2)$$

The ℓ_1 **Trend Filter (TF)** arises when the penalty operator in the signal approximation problem is the k-th order differences. For $k=0$ the estimates are **piecewise constant**, $k=1$ **piecewise linear**,. . .. The resulting objective is of the following form where D^{k+1} is a discrete difference linear operator.

$$\widehat{\beta}_{TF}(\lambda) = \arg\min_{\beta} \|\mathbf{Y} - \beta\|_2^2 + \lambda\left\|D^{(k+1)}\beta\right\|_1 \quad (3)$$

TF recovers close to **true signal** of blocks and heavyside test functions with Gaussian white noise





Adaptive Trend Filtering

Adaptive Trend Filtering (ATF) estimates are defined as the solution to the optimization problem of

$$\widehat{\beta}_{ATF}(\lambda) = \arg\min_{\beta} \|\mathbf{Y} - \beta\|_2^2 + \lambda\left\|D^{w,k+1}\beta\right\|_1 \quad (4)$$

where $D^{w,k+1}$ is a **weighted discrete difference operator** with weights $w_1, \ldots, w_n$ that are incrementally learned in the estimation procedure from the observed signal and chosen covariates. **Minimax optimal rates** can be shown when the true function f_0 belongs to the class of finite total variation functions

$$F_k(C) = f : [0,1] \to \mathbb{R} : f \text{ is k times weakly differentiable and } TV(f^k) \leq C \quad (5)$$

In the case where $k=1$ recall the resulting estimates are **piecewise linear**. Then under assumptions of sub-Gaussian errors and a tuning parameter on the order of $\lambda = \mathcal{O}(n^{\frac{1}{5}}C_n^{-\frac{3}{5}})$, with conditions similar to assumptions on design point spacing within [3] of

$$max_{i=1,\ldots,n}(w_i^{-1}) = \mathcal{O}(n^{-\frac{2}{5}}C_n^{-\frac{4}{5}}) \quad (6)$$

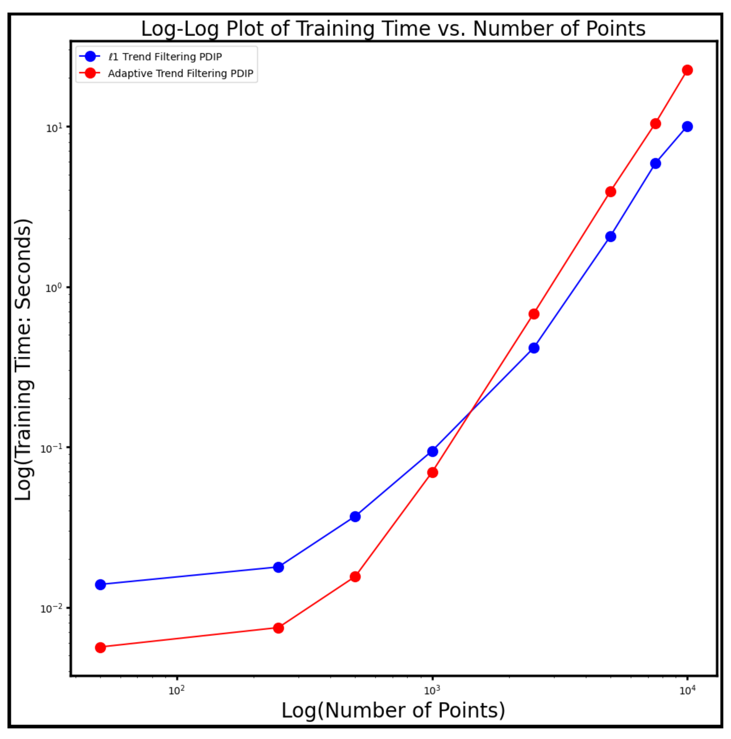
then the resulting ATF estimates will converge to f_0 in probability at the **Minimax optimal rate** of

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_i - f_0(x_i))^2 = \mathcal{O}_{\mathbb{P}}(n^{-\frac{4}{5}}C_n^{\frac{2}{5}}) \quad (7)$$

Algorithms

- Requires **estimation of the unknown parameters**, $\beta_1, \ldots \beta_n$ in the ℓ_1/ adaptive trend filter defined in 3 and 4
- Solve the **convex quadratic program** composed of the primal objective in 8 and dual objective in 9 with an interior point method
- Discrete difference operator is Toeplitz and banded with bandwidth $k+2$, allowing the linear system in each iteration to be solved in **linear time complexity**
- Specialized alternating direction method of multipliers routines have been developed in [2] that allow for **distributed learning in Hadoop and Apache Spark**
- ATF computational time is competitive to TF across various problem sizes as displayed in the log-log plot in Figure1

Figure 1: Log-Log of Computational Time vs. Problem Size



The **primal** objective of ATF

$$\|\mathbf{Y} - \beta\|_2^2 + \lambda\|z\|_1 \quad s.t. \quad z = D^{(w,k+1)}\beta \quad (8)$$

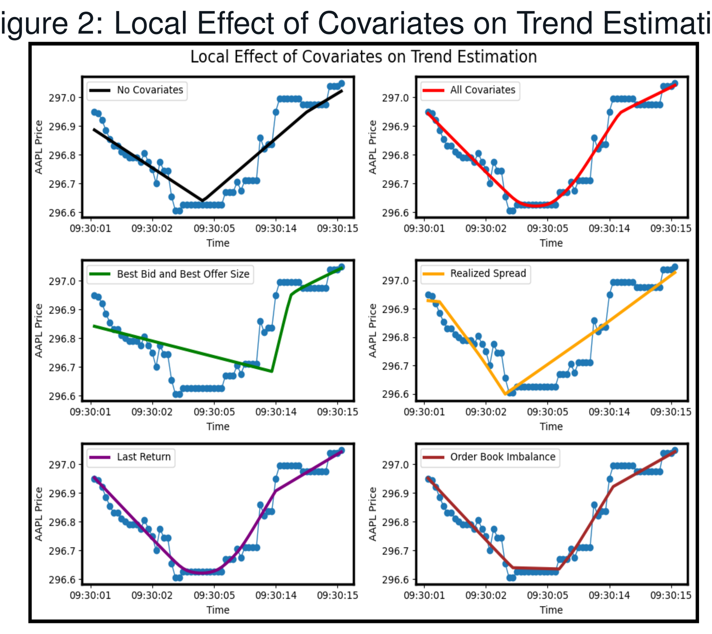
And **Lagrangian dual** objective is of the form

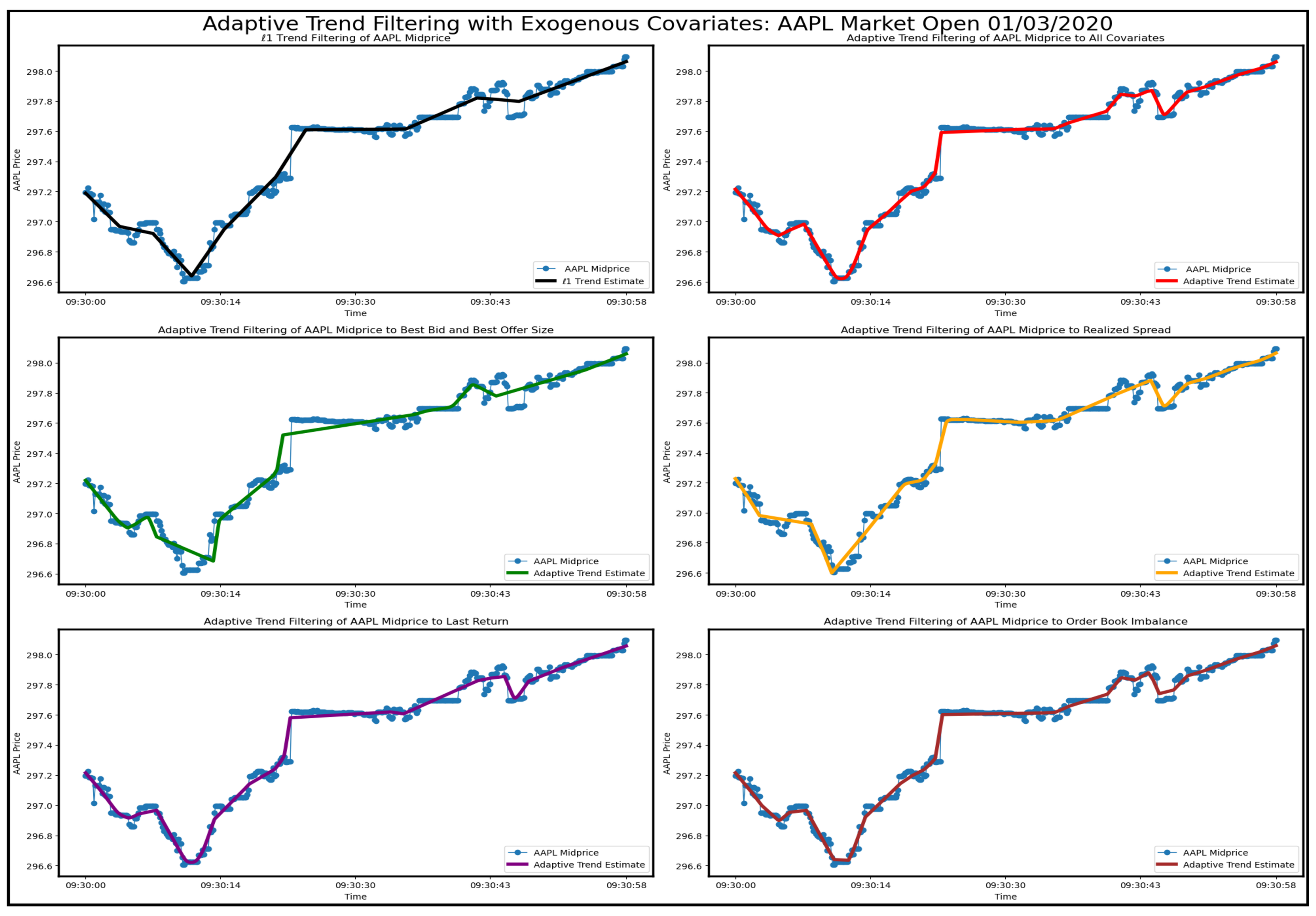
$$\frac{1}{2}\left\|\nu^T D^{(w,k+1)} D^{(w,k+1)T}\nu\right\|_2^2 \quad s.t. \quad \lambda 1 \leq \nu \leq \lambda 1 \quad (9)$$

Empirical Analysis

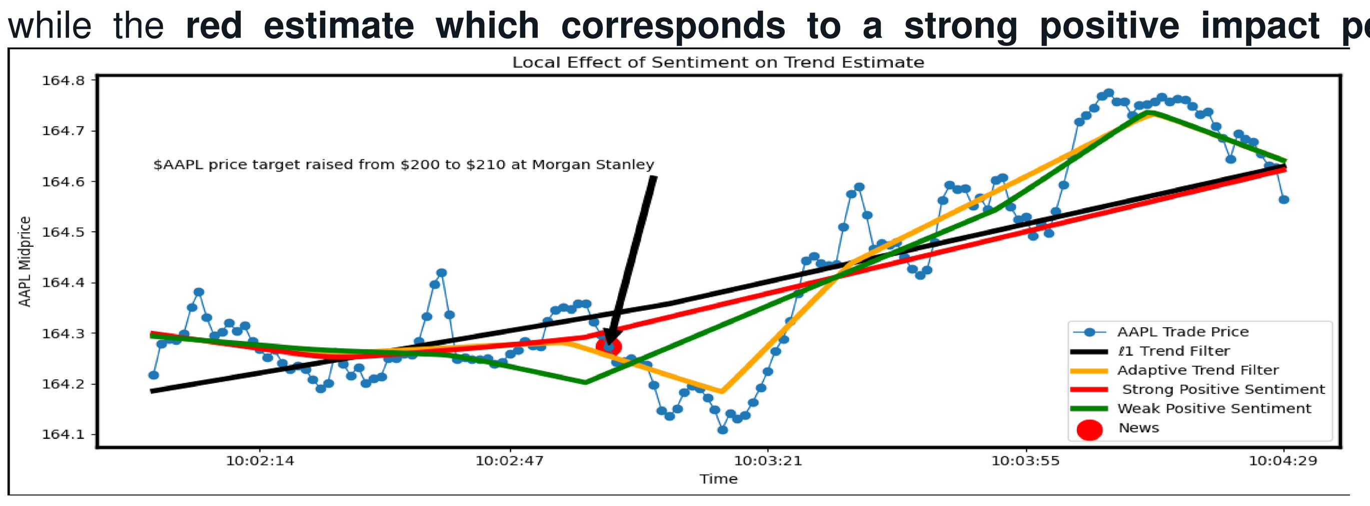
- Applied to **high-frequency trade and quote data** from New York Stock Exchange
- Lagged **market microstructure features** as exogenous covariates; Spread, Volume, Last Return, Order Book Imbalance
- Learns underlying hidden signal by locally **adapting price process estimates** to smoothness information of the covariates
- **Smoothness, location, and magnitude** of trend estimates vary locally, see Figure 2 for a comparison of AAPL at market open

Figure 2: Local Effect of Covariates on Trend Estimation





Often **recent news** of an asset, such as an earnings announcement or an economic data release, will **influence its trading activity**. Below aggregated one-second quote data of AAPL near market-open is plotted the day after an earnings announcement. Using public Twitter data, a tweet corresponding to a **price target increase from Morgan Stanley** is extracted and, using FinBert [1], has an estimated positive sentiment score of 0.71. Depending on the impact decay specified, sentiment covariates can be engineered and used in ATF. Note the **green estimate corresponds to a weak positive sentiment and rebounds quickly** after the tweet while the **red estimate which corresponds to a strong positive impact persists longer**.



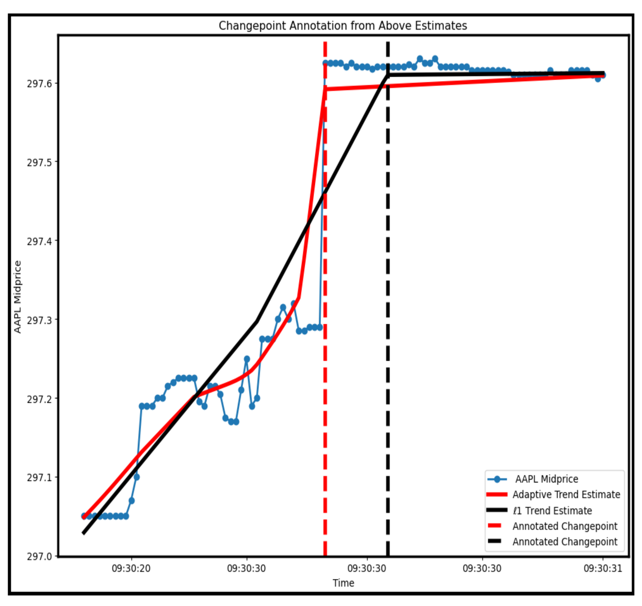
Changepoint Detection

- Multiple Changepoint Detection in the mean has been formulated as a **special case of the LASSO** and can be solved as an (A)TF

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^n} \|y - x\|_2^2 + \lambda\left\|D^2 x\right\|_1 \quad (10)$$

- Indices in (A)TF estimates of significant k-th order differences are selected to **identify candidate changepoints in price trends**
- Incorporating covariates in ATF allows for nonlinearity in trend estimation and changepoint annotation, see Figure 3 where **ATF captures the nonlinear jump and changepoint** quicker than TF

Figure 3: ATF Capturing Nonlinear Jump and Changepoint



Conclusions and Extensions

- Adaptive Trend Filtering is an extension of ℓ_1 Trend Filtering designed to **capture nonlinearities** within high-frequency time series. ATF learns the underlying signal while **locally adapting to external factors** such as volume, spread, and sentiment for more accurate learning. We discussed the direct use of resulting estimates for **interpretable changepoint annotation**
- We will explore if our results can be applied to designing **relative embeddings of local price dynamics** to utilize within attention-based architectures specific to the financial domain

References:

[1] Dogu Araci. *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.* 2019. arXiv: `1908.10063 [cs.CL]`.

[2] Aaditya Ramdas and Ryan J. Tibshirani. "Fast and Flexible ADMM Algorithms for Trend Filtering". In: *Journal of Computational and Graphical Statistics* 25.3 (July 2016), pp. 839–858. DOI: `10.1080/10618600.2015.1054033`. URL: `https://doi.org/10.1080%2F10618600.2015.1054033`.

[3] Ryan J. Tibshirani. "Adaptive piecewise polynomial estimation via trend filtering". In: *The Annals of Statistics* 42.1 (2014), pp. 285–323. DOI: `10.1214/13-AOS1189`. URL: `https://doi.org/10.1214/13-AOS1189`.