Real-Time Data Processing and Retailer's Performance Under Demand Uncertainty

Ruiqing Cao
Stockholm School of Economics
Catherine (Sijia) Ma
Tilburg University

Cite as:
Cao Ruiqing, Ma Catherine (Sijia) (2024), Real-Time Data Processing and Retailer's Performance Under Demand Uncertainty. *Proceedings of the European Marketing Academy*, 52nd, (119360)

Paper from the 53rd Annual EMAC Conference, Bucharest, Romania, May 28-31, 2024

Real-Time Data Processing and Retailer's Performance Under Demand Uncertainty

Abstract: Retail firms are increasingly adopting various technologies to collect real-time information on consumers, products, and channels. Stream processing is the up-to-date technology for retailers to extract values from the real-time information. However, harnessing the effectiveness of stream processing and developing dynamic capabilities therein, especially under demand uncertainty, has become an emerging challenge for traditional retailers. This study focuses on investigating the impact of stream processing on firm performance under demand uncertainty. We found that stream processing can significantly impact firms' profitability, especially in periods with volatile demand shocks. More importantly, we emphasize the key role of complementary data architecture in the effect of stream processing as a potential mechanism. This study contributes to the literature by providing deeper insights into the economic value of stream processing and the importance of architectural support as a key contingency. It also provides managerial implications for firms seeking to develop dynamic capabilities under demand uncertainty.

Keywords: Big data; Stream processing; Demand uncertainty; Data architecture.

Track: Retailing & Omni-Channel Management

1. Introduction

Accurate forecasting of sales and consumption is particularly important for marketing in firms under demand uncertainty because firms need to adjust marketing budget allocations and overall marketing strategies based on market information (Liu et al. 2016). Demand uncertainty refers to the variability in customer populations and preferences (Grewal and Tansuhaj 2001). To deal with the demand uncertainty, previous studies have devised several solutions for firms, such as various forecasting models (Ren et al. 2017), inventory management (Chuang et al. 2019), and pricing strategies (Wen et al. 2019). This paper adds to this literature by uncovering the novel impact of stream processing adoption on firm performance under demand uncertainty.

Stream processing refers to a type of big data technology that focuses on the real-time processing of continuous streams of data in motion. Applications like Kafka and Hadoop are typical examples that implement stream processing with real-time data in business organizations (Kreps et al. 2011). With the increasing digitization of marketing channels and consumerization of Information technologies, there is a tremendous amount of real-time data available in the retail industry, mainly pertaining to dimensions like customers, products, time, locations, and marketing channels (Bradlow et al. 2017). Nevertheless, the impact of stream processing on firm performance under demand uncertainty is unclear and undocumented in the literature. On the one hand, stream processing digests the large amount of real-time information available from the consumers' side into demand predictions, which enhances retailers' dynamic capability in volatile markets. On the other hand, the technological investment of implementing stream processing may be cumbersome for retailers. traditional incumbents often find it much harder to maneuver big data for driving competencies in servicing customer needs (Henderson & Clark, 1990; Iansiti & Lakhani, 2020). Therefore, it is important to empirically analyze the economic impact of stream processing on firm performance under demand uncertainty.

This study constructs a unique panel dataset for establishments in the U.S. clothing retail sector to answer this research question. We use COVID-19 as an exogenous shock and adopted three sets of measures for demand shock. The analysis results suggest that stream processing adoption mitigates the negative impact of demand shock and heightened uncertainty on establishment revenue. We show that these moderation effects exist primarily for establishments with complementary infrastructural technologies (e.g., microservices architecture and complex databases). Our findings contribute to the literature on the value of big data in retailing (Bradlow et al., 2017; Choi et al., 2018), and focus specifically on the role of enabling technologies and organizational capital in helping firms withstand a macroeconomic shock (e.g., economic crisis or global pandemic) that require swift adaptation and responses driven by accurate information in real-time that depends on a changing external environment.

2. Theory Development

2.1 Value of big data for retailers and real-time data processing

Data and digital technologies are valuable resources of firms (Bharadwaj, 2000) that enable firms to outcompete in the consumer market. Retailers collect enormous amounts of data through customer transactions, collecting and processing increasingly high volume, variety, and velocity of data (hence "big data"). The proliferation of larger quantities and varieties of data sources include not only traditional enterprise applications (e.g., ERP and CRM), but also non-traditional data sources such as IoT devices, wireless sensor networks, social media, and third-party data vendors (Bradlow et al., 2017).

For retailers, big data capabilities are particularly demanding for information processing systems, due to the need to incorporate not only historical data, but also ***real-time data*** from multiple sources such as inventory levels, transactions, customer feedback, competitor pricing and macroeconomic trends, in order to take actions promptly in response to fluctuations in market demand. In contrast to traditional batch processing which ingests data in discrete chunks at scheduled intervals, stream processing entails continuous handling of data streams. In recent years, Apache Kafka has become an increasingly popular architecture choice for implementing real-time stream processing capability. Real-time processing capability has many use cases across different business and operational domains, such as demand nowcasting and inventory and supply chain management (Bradlow et al., 2017). Research has shown that the timeliness of data sources are critical to demand forecasting (Liu et al., 2016). This requires firms to incorporate new information on sales and inventory and data updates in real time. Other use cases of real-time data processing include dynamic pricing and customer services and automated recommendation (Kopalle et al., 2023).

These applications of real-time capability are particularly useful during times of economic crises, which leads to unusally high market uncertainty where traditional demand forecasting (using historical sales data) fail due to large short-run fluctuations and unpredictability in consumer demand. In such an environment, firms need accurate and up-to-date information to make prompt decisions in a limited frame of time. Efficient processing of high-velocity information flows is hence crucial to responding to changing market environment to optimize decisions in real-time (Aloysius et al., 2018). The ability to incorporate real-time information allows firms to tune into changes in the market and external environment and become resilient to negative shocks. Hence, we propose:

H1: Real-time data stream processing capability mitigates retailers' revenue loss due to sector-wide demand contraction and uncertainty induced by macroeconomic shock.

2.2 Architectural support

Whether firms can successfully deploy analytics and other big data applications depend on other organizational factors and enabling technologies. Big data systems are not standalone technologies that can be plugged into the organization and readily work without reconfiguring to fit with existing systems. IT architecture facilitates the exchange of data and information within the organization to meet business needs. Design of big data architecture is critical to success of big data application in driving firm value (e.g., Chen & Zhang, 2014; Bradlow et al., 2018). In order for stream processing technology to effectively drive value creation through enabling real-time data processing, firms must have appropriate data architecture configured to support such capability. We propose:

H2: Data architecture support is crucial to effectiveness of real-time data stream processing capability in mitigating the negative effect of demand contraction and uncertainty on revenue.

3 Empirical Setting: COVID-19 and Contraction of the Clothing Retail Sector

The empirical setting entails establishments of large corporations in the clothing retail sector in the United States. We choose to focus on clothing retailers for two reasons. First, clothing is *a non-durable good* where the value declines relatively quickly over time, and hence having up-to-date information about products and users are very useful to managing the inventory and forecasting demand, in order to maintain revenue performance. Second, clothing retail is the only non-durable goods sector that *experienced large sector-wide demand contraction during COVID-19*, while non-durable goods sectors, namely NAICS 445 (food and grocery) and NAICS 446 (health and personal care) experienced slight increase in total demand at the sector level. In addition, the sample exclusively consists of brick-and-mortar retailers (and thus exclude e-commerce or non-store retailers such as Amazon, eBay, Wayfair, or Etsy).

The onset of the COVID-19 pandemic in early 2020 led to ***large demand declines*** and business closures around the United States. While e-commerce sales soared as potential customers moved to purchase products online, traditional brick-and-mortar retailers suffered performance declines. The clothing retail sector made up close to 40% of total retail sector employment loss from 2019 to 2020. Retailers selling non-essential products are particularly hurt, because they provide primarily products that are not necessary to daily living and can be delayed. The COVID-19 pandemic also led to ***heightened demand uncertainty*** among large clothing retailers. Corporations' quarterly sales data revealed much larger demand uncertainty in 2019 relative to 2020, measured as total deviation from predicted quarterly sales following prior literature (e.g., Gaur, Fisher, & Raman, 2005, see Section 4).

4 Data and Measures

We combine four data sources to conduct empirical analyses: (1) the Aberdeen Compute Intelligence Technology Database (CITDB) in 2019 and 2020, (2) Census Bureaus' Monthly State Retail Sales, to measure state-level demand change, (3) Compustat Quarterly Fundamentals data, to measure corporation-level demand uncertainty, and (4) Keystone-Microsoft survey, to measure corporation-wide data architecture. The key variables are summarized as follows.

Establishment Revenue. Measured by CITDB (2019 and 2020) at the establishment level.

Real-Time Data Stream Processing. Stream processing architecture is typically implemented by Apache Kafka, a distributed publish-subscribe messaging system designed to handle large data streams based on committed logs. Kafka acts as a central hub for ingesting, storing and processing big data in real time. Kafka allows message exchanging across multiple consumers and producers in a fault-tolerant manner. We measure stream processing capability at the establishment level by the presence of open-source stream processing frameworks (Apache

Kafka and Apache Flink) and commercial data platforms for running stream processing frameworks (Google Dataproc, Hortonworks, and Confluent) in 2020.

Demand Change and Demand Uncertainty. Demand Change is measured by the year-on-year change in total retail sales by sector and state, available in the Census Bureau's experimental data product Monthly State Retail Sales featuring retail sales information at the state level. Demand Uncertainty (e.g., Gaur, Fisher, & Raman, 2005; Rajagopalan, 2013; Chuang, Oliva, & Heim, 2019) is measured at the corporation-year level, using the Compustat Quarterly Fundamentals data, as the sum of absolute quarterly percentage deviation in revenue from fitted value, estimated by the sum of a linear trend and seasonality (quarter fixed effects) when estimating the linear model that predicts quarterly corporation revenue.

Microservices Architecture. In 2020, Keystone Strategy and Microsoft designed a survey, based on deep domain expertise in data and cloud technologies. The survey data was collected by interviewing senior-level technology executives (e.g., CIOs, CDOs) at large corporations who oversee the entire organization's IT systems and digitalization efforts. Microservices is coded as 1 (0) is the answer to the question: *"Is your data fabric or platform built on a micro-services architecture?"* is *Yes (No).*

Other Data Architecture Variables. Using CITDB data in 2020, we construct the following variables for each establishment: number of non-relational (noSQL) database products typically storing unstructured data, and batch processing indicator (prsence of Apache Spark, Apache Hadoop, Apache Oozie or commercial data platforms for running these frameworks such as Google Dataproc and Hortonworks).

5 Empirical Methods

We use Ordinary Least Squares regressions to estimate the effects of COVID-19 induced demand uncertainty on establishment revenue. The data sample consists of balanced panels at the establishment level across two years (2019 and 2020). The most stringent regression specification controls for establishment fixed effects and corporation-state-year fixed effects.

To assess the role of real-time (data stream) processing capability in establishments' ability to withstand the COVID-19 induced demand shock, we estimate the moderation effects of the ***stream processing indicator*** on log establishment revenue in Model 1. The coefficient $\boldsymbol{a}$ estimates the extent to which streaming processing improves the establishment revenue when the main regressor (demand shock/decline/uncertainty due to COVID-19) has a larger magnitude.

$$Ln(Revenue)\_it = a*SP\_i \times Shock\_it + FE(i) + FE(c,s,t) + e\_it \quad (1)$$

There are three different "Shock" variables: (1) the COVID-19 indicator equal to 1 for all observations where year is 2020 (*Post*), which does not account for differences across states and across corporations in demand decline and volatility, (2) state-level demand decline in clothing retail (*DDemand*), and (3) corporation-level demand uncertainty (*DUnc*). We evaluate model robustness by estimating the moderation effects after controlling for different levels of fixed effects, e.g., year, corporation-year, and corporation-state-year. We also conduct split-sample analyses to evaluate boundary conditions of the main effects related to stream processing capability.

6 Results

6.1 Descriptive Statistics

The final data sample consists of a balanced establishment-year panel of 11,438 establishments in 2 years (2019 – 2020). These establishments belong to 22 large retail corporations (each with hundreds or thousands of U.S.-based branches that conduct business in the NAICS 445 sector, which are included as sample data points). To construct the data sample, we start with the list of large corporations with total revenue above 2 billion USD in the NAICS 44 sector. We then match these corporations to clothing retail establishments (NAICS code 448) in the CI Technology Database (CITDB). The median establishment has an annual revenue of 2 million USD, 2 non-relational (noSQL) database products, and 4 public cloud technologies. Around 45% of the establishments have data stream processing capability.

6.2 Stream processing capability lessens the impact of external demand shock on revenue loss among clothing retail establishments

Table 1 presents results generally consistent with Hypothesis 1, as shown by the coefficients and p-values on the interaction between stream processing capability and various measures of COVID-19 induced demand variation and uncertainty. To evaluate the boundary conditions for the role of stream processing capability, especially depending on complementary architectural support, we conduct split-sample analyses.

Table 2 shows that among corporations that build their data fabric on the microservices architecture, establishments with stream processing experience 64% less negative impact of COVID-19 on average. For these corporations, the negative effect of a 1% decline in sector-level demand on revenue is mitigated by 4% among establishments with stream processing capability relative to those without it. Similarly, the negative effect of a 1-standard-deviation increase in corporation-level demand fluctuation on revenue is 1.1% smaller among establishments with stream processing capability relative to those without it. In contrast, when corporations do not build their data fabric on the microservices architecture, stream processing capability does not significantly contribute to the differential impact of aggregate demand fluctuation and uncertainty on revenue performance.

Table 3 shows the regression results in subsamples split by levels of complementary infrastructural technologies, specifically non-relational (i.e., noSQL) databases typically used for storing large amounts of unstructured data. Among those with above-median non-relational database variety, establishments that also have stream processing capability experience 39% better performance post-COVID relative to those without stream processing capability. Establishments' revenue performance response to one standard deviation (8 percentage points) larger decline in sector-state level aggregate demand has a 20% smaller magnitude when the establishment has stream processing capability, relative to when it does not have stream processing capability. In contrast, among establishments with below-median non-relational database variety, stream processing capability does not seem to introduce significant differences in response to COVID-19 demand shock. We also conducted robustness checks showing that these results are not driven by big data processing in general (e.g., batch processing).

References

Aloysius, J. A., Hoehle, H., Goodarzi, S., & Venkatesh, V. (2018). Big data initiatives in retail environments: Linking service process perceptions to shopping outcomes. *Annals of operations research*, 270, 25-51.

Bradlow, E.T., Gangwar, M., Kopalle, P. & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of retailing*, 93(1), 79-95.

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS Quarterly*, 169-196.
Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.
Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and Operations Management*, 27(10), 1868-1883.
Chuang, H. H. C., Oliva, R., & Heim, G. R. (2019). Examining the link between retailer inventory leanness and operational efficiency: Moderating roles of firm size and demand uncertainty. *Production and Operations Management*, *28*(9), 2338–2364.
Dorfman, D. (2022). Retail trade employment: Before, during, and after the pandemic. *Beyond the Numbers: Employment & Unemployment*, *11*(4). https://www.bls.gov/opub/btn/volume-11/retail-trade-employment-before-during-and-after-the-pandemic.htm
Gaur, V., Fisher, M. L., & Raman, A. (2005). An econometric analysis of inventory turnover performance in retail services. *Management Science*, *51*(2), 181–194.
Grewal, R. & Tansuhaj, P. (2001). Building organizational capabilities for managing economic crisis: The role of market orientation and strategic flexibility. *Journal of Marketing*, 65(2), 67-80.
Kopalle, P. K., Pauwels, K., Akella, L. Y., & Gangwar, M. (2023). Dynamic pricing: Definition, implications for managers, and future research directions. *Journal of Retailing*.
Kreps, J., Narkhede, N. & Rao, J. (2011). Kafka: A distributed messaging system for log processing. In Proceedings of the *NetDB* 11, 1-7.
Liu, X., Singh, P.V. & Srinivasan, K. (2016). A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Science*, 35(3), 363-388.
Rajagopalan, S. (2013). Impact of variety and distribution system characteristics on inventory levels at US retailers. *Manufacturing & Service Operations Management*, *15*(2), 191–204.
Ren, S., Chan, H.L. & Ram, P. (2017). A comparative study on fashion demand forecasting models with multiple sources of uncertainty. *Annals of Operations Research*, 257, 335-355.
Wen, X., Choi, T.M. & Chung, S.H. (2019). Fashion retail supply chain management: A review of operational models. *International Journal of Production Economics*, 207, 34-55.
Yohn, D. L. (2020). The pandemic is rewriting the rules of retail. *Harvard Business Review*, 2–6.

Table 1: Main effects. *∗∗∗p < 0.001; ∗∗p < 0.01; ∗p < 0.05.*

Dep. Var.	Ln(Revenue)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post × SP	0.150***	0.244**	0.261***						
	(0.040)	(0.076)	(0.076)						
DDemand				0.001	0.004	0.000			
				(0.004)	(0.004)	(.)			
DDemand × SP				-0.009***	-0.013**	-0.017***			
				(0.003)	(0.005)	(0.005)			
DUnc							-0.562***	0.000	0.000
							(0.122)	(.)	(.)
DUnc × SP							0.358*	0.268	0.314*
							(0.141)	(0.140)	(0.136)
Branch FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Year FE	Y			Y			Y		
Firm-Year FEs		Y			Y			Y	
Firm-State-Year FEs			Y			Y			Y
Obs.	22,574	22,574	22,574	22,574	22,574	22,574	22,328	22,328	22,328
R-Squared	0.739	0.746	0.767	0.739	0.746	0.767	0.738	0.744	0.766

Table 2: Complementary infrastructural technologies: unstructured database and multi-cloud environment. *∗∗∗p < 0.001; ∗∗p < 0.01; ∗p < 0.05.*

Dep. Var.	Ln(Revenue)					
	(1)	(2)	(3)	(4)	(5)	(6)
Sample	MS=1	MS=0	MS=1	MS=0	MS=1	MS=0
Post × SP	0.638***	0.109				
	(0.134)	(0.075)				
DDemand × SP			-0.041***	-0.006		
			(0.009)	(0.004)		
DUnc × SP					7.209***	0.125
					(1.513)	(0.122)
Branch FE	Y	Y	Y	Y	Y	Y
Firm-State-Year FEs	Y	Y	Y	Y	Y	Y
Obs.	6710	5424	6710	5424	6710	5424
R-Squared	0.768	0.806	0.768	0.806	0.768	0.806

Table 3: Analysis of unstructured data base. *∗∗∗p < 0.001; ∗∗p < 0.01; ∗p < 0.05.*

Dep. Var.	Ln(Revenue)					
	(1)	(2)	(3)	(4)	(5)	(6)
Sample	ComplexDB >=P50	ComplexDB <P50	ComplexDB >=P50	ComplexDB <P50	ComplexDB >=P50	ComplexDB <P50
Post × SP	0.385***	0.062				
	(0.087)	(0.083)				
DDemand × SP			-0.025***	-0.002		
			(0.006)	(0.006)		
DUnc × SP					0.370**	0.788
					(0.129)	(0.815)
Branch FE	Y	Y	Y	Y	Y	Y
Firm-State-Year FEs	Y	Y	Y	Y	Y	Y
Obs.	13924	8378	13924	8378	13924	8132
R-Squared	0.783	0.760	0.783	0.760	0.782	0.758