sciendo

ACD Modeling High-Frequency FX and Market Microstructure

Jorge Esteban Hernández[*♣]

[♣]Independent Researcher

Submitted: November 11, 2023 • Accepted: May 3, 2024

ABSTRACT: This paper advances high-frequency foreign exchange (FX) market microstructure analysis by adapting Autoregressive Conditional Duration (ACD) models to study intervals between price updates. By treating these updates as random variables within a point process, the models adeptly capture the dynamic structure of conditional durations and retain key information in high-frequency series. These series display properties critical for understanding market behavior and liquidity dynamics. The findings challenge the belief that increased data frequency reduces microstructural relevance, showing it actually improves understanding of market dynamics. This study broadens econometric model applications and offers updated insights into FX market behavior, providing practical information for academics, practitioners, and policymakers. It contributes significantly to the literature and lays a foundation for future research.

JEL classification: C58, C51, G15, F31.

Keywords: High-Frequency, Market Microstructure, Stylized Facts, Point Processes, Autoregressive Conditional Duration Models, Price Durations.

1 Introduction

This paper applies Autoregressive Conditional Duration (ACD) models to high-frequency price data in the foreign exchange (FX) market, unveiling new insights into market dynam-

[*]Corresponding Author. E-mail: jorge.estebanh@alumni.uam.es

DOI: 10.33119/ERFIN.2023.8.2.2

ics and the price formation process. A key contribution of this research is demonstrating how ACD models, which treat observations (or arrival times) as random variables defining a point process, thereby preserving the dynamic structure of conditional durations, can be adapted from their traditional use in analyzing transaction durations in centralized markets to explore price-based high-frequency data. This adaptation is essential for decentralized markets such as the FX market, characterized by a fragmented structure that introduces complexities to trading, both financially and in terms of information flow (Osler, 2009; Evans and Rime, 2019).

The microstructure of the FX market focuses on the currency trading process and how information is disseminated among market participants, ultimately determining the exchange rates. This dissemination is implicitly captured within high-frequency price series, thereby underscoring the necessity of such analyses (Maechler, 2020). Technological advancements in electronic trading have profoundly transformed the FX market by significantly reducing processing times and increasing the frequency of events such as transactions and market price updates. These developments necessitate a re-evaluation of market microstructure theories, particularly in terms of how price updates and liquidity are interconnected at micro intervals. Contrary to the traditional belief that high-frequency data may diminish the significance of microstructure analysis, the findings suggest that higher frequency actually enhances the importance of microstructural considerations (Dacorogna et al., 2001; O'Hara, 2015).

By utilizing ACD models to study high-frequency price durations, the research provides substantial evidence that the dynamics of price formation are intricately linked to the microstructural conditions of the market. This analysis not only broadens the application of existing econometric models to new types of data, specifically price durations rather than transaction durations, but also updates the understanding of today's FX market behavior, noting trends such as unprecedented volatility levels that significantly impact the microstructure due to the market's dynamic nature. Moreover, this paper enriches the existing literature by providing detailed practical insights for market practitioners, academic theorists, and policymakers on applying ACD models to high-frequency price data series, such as emphasizing the importance of recognizing and utilizing the stochastic properties known as stylized facts, which are essential for aligning theoretical frameworks and empirical models with observed market behaviors, thereby laying a solid foundation for navigating and interpreting the complexities of modern financial markets.

2 High-frequency FX and microstructure fundamentals

2.1 The market anatomy

The FX market is a global market where currencies from different countries are exchanged, determining their relative values through exchange rates. It operates continuously, with a vast number of buyers and sellers participating in major financial hubs worldwide. This market stands out for its competitiveness and efficiency, as outlined by Kallianiotis (2013). FX trading volumes are exceptionally high, making it the largest and most liquid asset class globally. According to the Bank for International Settlements (2019), the average daily traded volume reaches 6.6 trillion USD.

The FX market covers three main geographical regions (APAC, EMEA, and Americas) on a follow-the-sun basis, with Tokyo, London, and New York serving as the primary hubs, respectively. Trading occurs round the clock on weekdays, starting from Monday at 7:00 Auckland time and ending on Friday at 17:00 New York time. Exchange rates are influenced by numerous factors, including information flows and expectations at national, supranational, and international levels. Compared to other markets, FX exhibits lower profit margins, which makes leverage widely accessible for achieving higher returns.

The market operates over-the-counter, where participants engage in direct transactions with each other. This enables the transfer of purchasing power, facilitates credit and trading instruments for commercial activities, investment opportunities, and currency hedging across various assets. While the exchange rate is determined by market players, central banks still retain some control through interventions.

The FX market, known for its exceptional liquidity and extensive trading, continues to expand. This growth can be attributed to the widespread use of electronic tools, leading to increased systematic trading. Algorithmic execution and a wide range of trading platforms contribute to heightened liquidity levels while reducing operating costs, attracting participants from diverse backgrounds. Notably, there has been an important increase in the involvement of retail investors, whose flows have become significant in today's observed trends.

2.1.1 Participants in the FX market

According to Kallianiotis (2013) and Osler (2009), the FX market differs from conventional markets in terms of fragmentation based on liquidity access levels. The interbank market represents the top tier, accounting for over half of daily traded volumes. In this segment, the world's largest investment and commercial banks directly interact, providing exclusive liquidity to each other at prices not commonly available in lower tiers. The second tier includes flows involving central and regional banks, non-bank financial institutions, non-financial corporations, and retail participants. It is important to note that these flows are not publicly disclosed, and only the involved parties are aware of their existence. Based on the classifications proposed by Carpenter and Wang (2007) and King et al. (2012), the market participants and their characteristics can be broken down as follows:

- Central banks play a crucial role in the foreign exchange market as they exert significant influence on long-term exchange rates through their monetary policies at the national level. Their main objective is to ensure short-term stability of their currency by utilizing substantial foreign currency reserves to influence market directions and meet policy objectives.

- Investment and commercial banks primarily manage client funds, conduct investment transactions, and offer execution services. Their aim is to generate profits through risk-based operations in the market.
- Non-bank financial institutions encompass asset management firms, hedge funds, pension funds, insurance companies, and possess significant influence on short-term dynamics due to their information processing capabilities, research and analysis resources, and technological advancements. These institutions are primarily involved in speculative flows and hedging across asset classes. Broking institutions provide market access (liquidity and credit) to smaller players or those wishing to operate anonymously. Brokers do not take positions themselves and generate profits based on transaction volumes through commissions or mark-ups.
- Non-financial corporations utilize the FX market for non-speculative purposes, such as converting cash for goods and services in foreign currencies. While their transactions may have minimal or short-lived impacts on the market, international trade flows can define longer-term trends for certain currency pairs. However, this class of participants is experiencing a decline as more companies prefer to operate through the execution desks of their banking counterparts, which provide better liquidity access and efficient execution algorithms.
- Retail investors consist of professional or semi-professional traders seeking additional income, often as a secondary source. This group has witnessed exponential growth in recent years and typically operates through broking firms, which offer leverage.

2.1.2 Traded instruments

In the FX market, trading occurs through currency pairs, where one currency is bought (or sold) against another at a specific exchange rate. There are two types of orders: passive orders, which are placed in order books, and aggressive orders, which follow the market price for immediate execution. These orders can take various forms, each with its own characteristics and used to fulfill different investor needs. According to the Bank for International Settlements (2019), FX instruments can be broadly classified into four categories.

- Spot transactions involve the exchange of two currencies at a specific rate, typically settling within two business days (except for CAD, TRY, and RUB, which settle within one day). Spot transactions account for nearly one-third of the FX market.
- Outright forwards or currency futures are contracts that allow the exchange of two currencies at a predetermined rate on a specified future date, which is typically

further away from the spot date. However, within the category of forwards, there are also those known as inside forwards (or pre-spot), which settle within the same (overnight - ON) or the next (tomorrow-next - TN) business day. Currency futures are forward contracts standardized and listed on an exchange, with details such as contract size and maturity being predetermined.

- Swaps enable the exchange of two currencies at a specific rate, with the near leg settling on a given date and the far leg reversing the exchange at a different rate on a later date. While even swaps do not involve spot risk, uneven swaps do, as the notional amounts on each leg differ. Swaps account for almost half of the traded volumes in the market.

- Over-the-counter (OTC) options are contracts that grant the right to exchange two currencies at a predetermined rate over a defined period. This category encompasses various derivative products, including currency swaptions (which add optionality to a swap) and currency warrants.

2.2 High-frequency data and its unique characteristics

The FX market operates in a decentralized way, which means that only the participants involved in each transaction possess knowledge of the trading flows. Zhou (1996) argues that market data flows demonstrate similar characteristics as prices are not obtained from a centralized entity. Instead, market makers on the selling side generate distinct prices at different time points. The concept of a market price is achieved by aggregating multiple liquidity layers from various sources and selecting the highest bid and lowest ask prices, commonly referred to as the top of the book. The market price serves as a reference in primary venues (ECNs) where numerous participants and flows converge.

Accordingly, the price data employed in the ensuing sections is derived from an aggregation of various price feeds sourced from market makers, notably including leading investment banks. The liquidity harnessed is thus attributable to the top tier, or the interbank market, as per Kallianiotis (2013) and Osler (2009) classification. The collection of price ticks contains the top of the book spot prices for the EUR/USD exchange rate. The considerable volume of ticks contained in this series (encompassing 1,342,595 price updates) stems from the fact that the EUR/USD constitutes the most liquid currency pair, representing approximately 28% of the trading volume.

The temporal scope designated for this study corresponds to the 11th trading week of 2020, specifically from the trade date of March 9th, 2020, to the market closure on the trade date of March 13th, 2020. This particular timeframe is of interest due to the heightened market volatility observed, which, in turn, furnishes an abundance of data, thereby enhancing the quality and robustness of the analysis and resultant findings.

2.2.1 High-frequency time series

High-frequency data refers to time series consisting of sequential ticks that represent observations related to market activity, typically market prices. Each tick generally includes a timestamp (in UTC time for standardization across time zones), bid and ask prices, and may also include the sizes associated with each price.

The key characteristic of high-frequency data is that timestamps are irregularly distributed over time, as they correspond to the moments when price updates occur rather than following a predetermined time scale. Analyzing high-frequency data provides valuable insights into market behavior, dynamics, and microstructure.

High-frequency series are characterized by their large size due to the substantial number of logged observations, which provides a significant amount of implicit information. On average, each major currency pair generates around 500,000 ticks within a 24-hour period, and the number of ticks is not capped and varies depending on market conditions.

The difference between the bid and ask prices is known as the spread, which represents the cost of simultaneously buying and selling the instrument. In high-frequency markets, price fluctuations are generally equal to or smaller than the quoted spread. Therefore, spreads widen during periods of reduced liquidity or high volatility and narrow during opposite market conditions.

While ticks provide information regarding dynamics and market conditions, the bid-ask bounce, which refers to the continuous oscillation of bid and ask prices within a narrow range when there is market activity but no price action, introduces a systematic bias in the data. This bias negatively affects the quality of analysis.

High-frequency data does not conform to normal or log-normal distributions, contrary to many models that assume log-normal price distributions for valuation methods such as the Black-Scholes model for options pricing. The irregularly spaced observations in high-frequency data series require the use of sampling techniques to homogenize observations over time, often achieved by linearly interpolating between observations or extrapolating from the previous value until the next anticipated time distribution mark is reached (e.g., hours, minutes, seconds).

Although the choice between these techniques has a negligible impact on results given the abundance of data, utilizing these techniques discards the implicit information contained in the irregularity, which can provide valuable insights into volatility, liquidity, seasonality, and other aspects of market dynamics. For a more detailed exploration of high-frequency time series, refer to Aldridge (2013).

2.3 Statistical properties of high-frequency data: a set of stylised facts

Dacorogna et al. (2001) were pioneers in proposing a set of stylized facts based on their empirical research of high-frequency data, which laid the groundwork for its analysis. Sun (2007) later revisited these facts.

2.3.1 The mid-price

The mid-price is commonly used to assign a single value to each tick and is calculated as:

$$\mathrm{GMP}_t = \sqrt{P_{\mathrm{ask},t} \cdot P_{\mathrm{bid},t}} \tag{1}$$

In some cases, the logarithm of the geometric mean is also used as the mid-price:

$$\mathrm{LMP}_t = \log \sqrt{P_{\mathrm{ask},t} \cdot P_{\mathrm{bid},t}} = \frac{\log P_{\mathrm{ask},t} + \log P_{\mathrm{bid},t}}{2} \tag{2}$$

The price of a currency pair is asymmetric, meaning that the exchange rate of the first currency for the second currency is not equal to the rate in the opposite direction (its inverse). However, using the logarithmic mid-price makes the exchange rate anti-symmetric or opposite in sign in both directions, enabling comparisons of returns among multiple currency pairs, independent of the unit of price.

2.3.2 Returns: distribution and autocorrelation

Returns are calculated as the difference between one price and the previous price. Specifically, the return at time t with a time increment Δt is calculated as follows:

$$R\left(t, \Delta t\right) = \mathrm{LMP}\left(t + \Delta t\right) - \mathrm{LMP}\left(t\right) = \log \frac{\mathrm{GMP}\left(t + \Delta t\right)}{\mathrm{GMP}\left(t\right)} \tag{3}$$

Returns provide improved information about the behavior of the series compared to analyzing prices alone. Market makers often use returns as a measure of performance because their distribution is more stable over time, nearly stationary, and exhibits a more symmetrical shape than price distributions. Returns are also employed in estimating volatility, particularly when measured at different frequencies. The distribution of returns tends to be leptokurtic, with a relatively small skewness coefficient and a high degree of symmetry as shown on Figure 1.

Measuring returns at different frequencies reveals that the tails widen as the frequency increases, as illustrated in Figure 2. The returns measured at 10- and 90-minute intervals are compared to the normal distribution using a Q-Q plot in Figure 3, confirming deviations from normality. The abundance of data in each tail allows for a detailed assessment.

Another characteristic of returns is the presence of negative first-order autocorrelation, which diminishes as the price formation process is completed. Figure 4 illustrates the bid and ask prices of EUR/USD during market opening, a time when thin liquidity delays the price formation process, making the negative first-order autocorrelation evident. The bid-ask bounce predominantly explains this negative autocorrelation, where the probability of a trade being executed on the bid followed by a trade on the ask is higher than the probability of the second trade also occurring on the bid. The correlogram in Figure 5 demonstrates the negative first-order autocorrelation in EUR/USD returns during the market open.

Figure 1: Histogram of EUR/USD returns

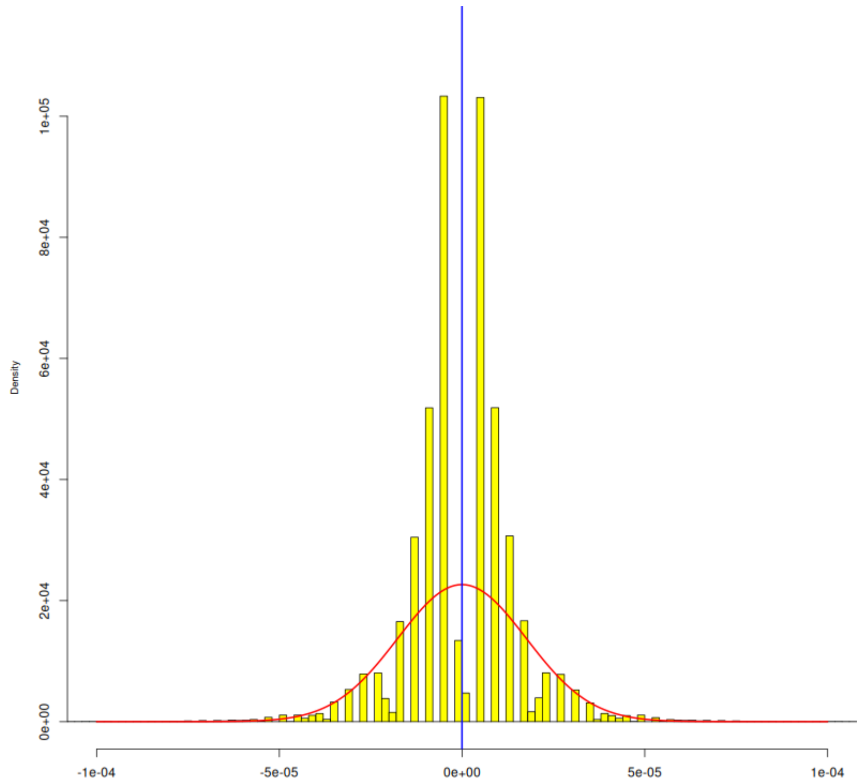
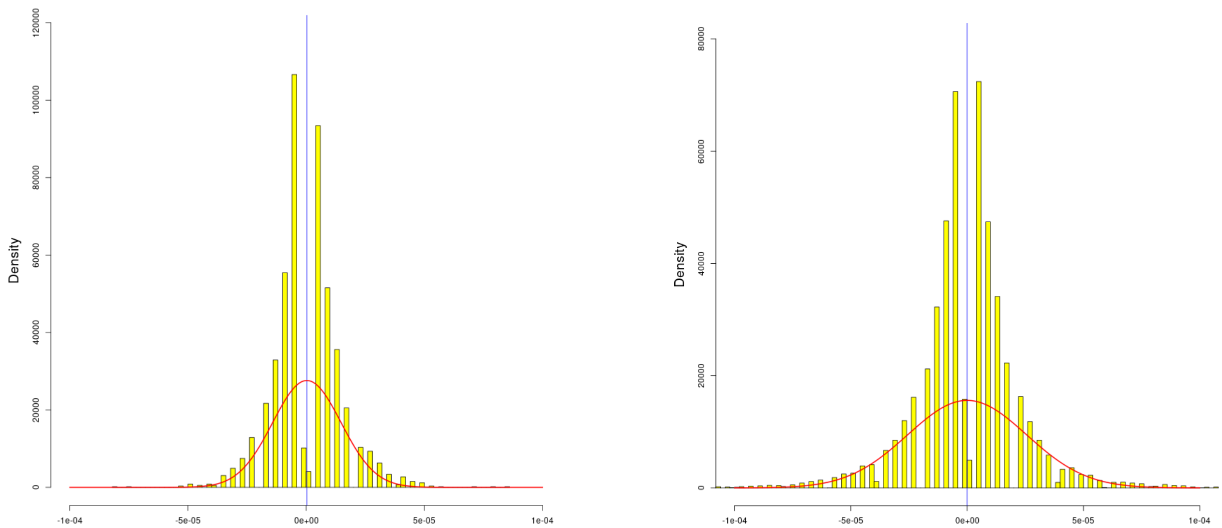


Figure 2: Histograms of EUR/USD returns at 10- (left) and 90-minute (right) intervals



2.3.3 Seasonality, clustering and memory

Ito and Hashimoto (2006) demonstrated that high-frequency data exhibits seasonality through periodic fluctuations over time, not only in prices but also in spreads and volatility patterns. Empirically, tick frequency, trading volumes, and volatility returns move procyclically, while spreads are countercyclical with respect to these factors (tightening during the most liquid times). Additionally, daily U-shaped patterns can be observed in

Figure 3: Distribution of EUR/USD returns at 10- (red) and 90-minute (blue) intervals

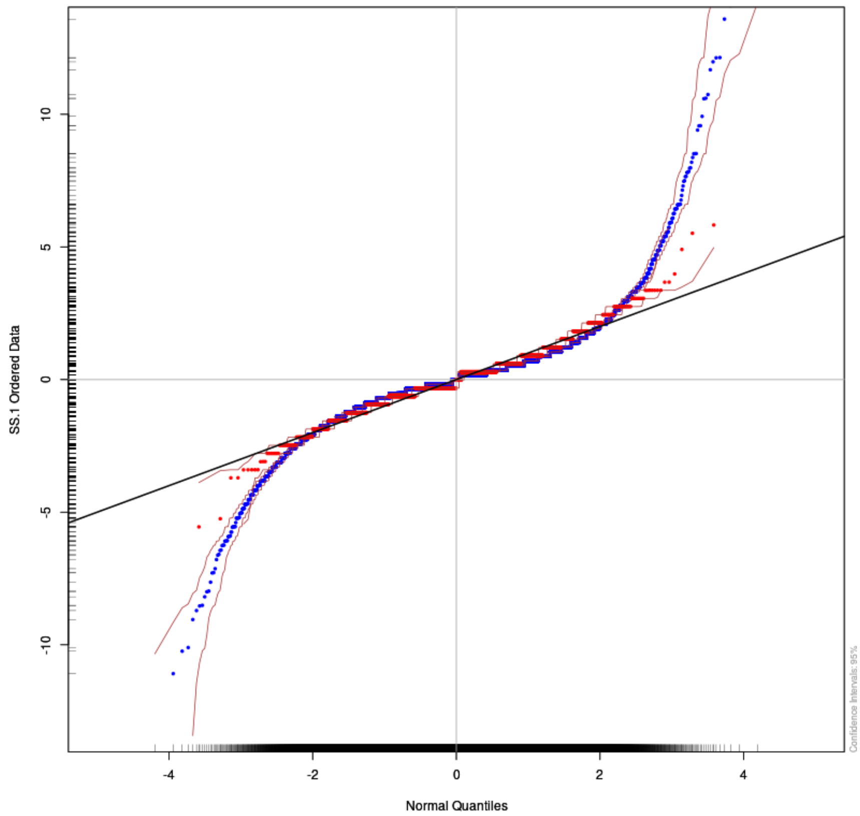
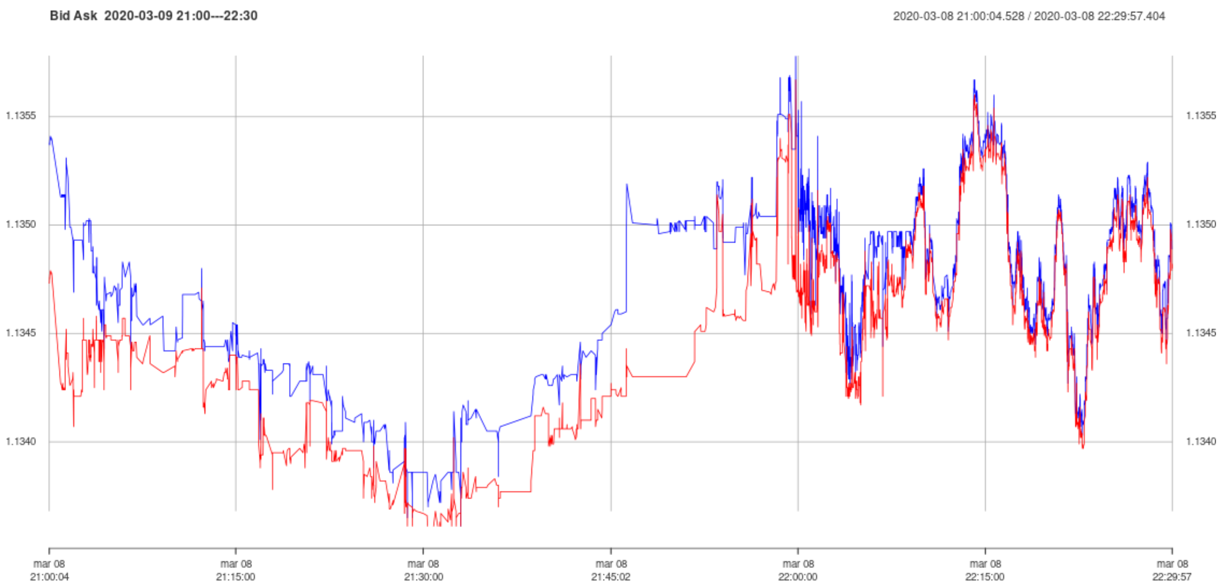


Figure 4: EUR/USD bid (red) and ask (blue) prices over the market open



market activity during the Tokyo and London sessions (excluding New York due to its overlapping with London). Figure 6, generated using exponential moving averages of intensity as per Dacorogna et al. (2001), clearly illustrates the recurrence of certain intraday patterns throughout the trading week for EUR/USD.

Figure 5: ACF correlogram of EUR/USD returns over the market open

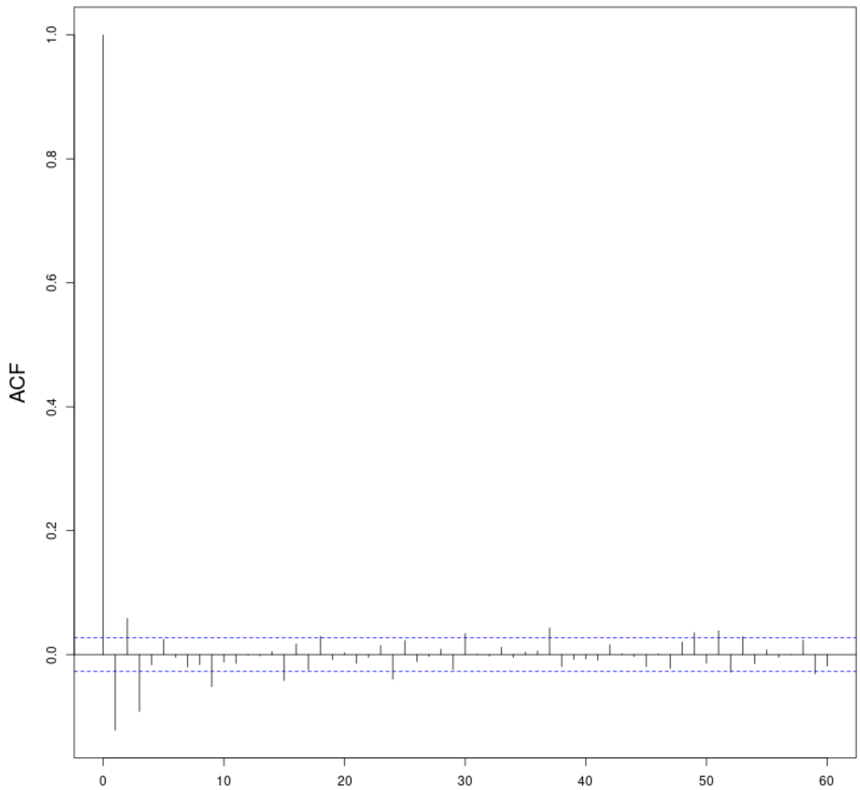
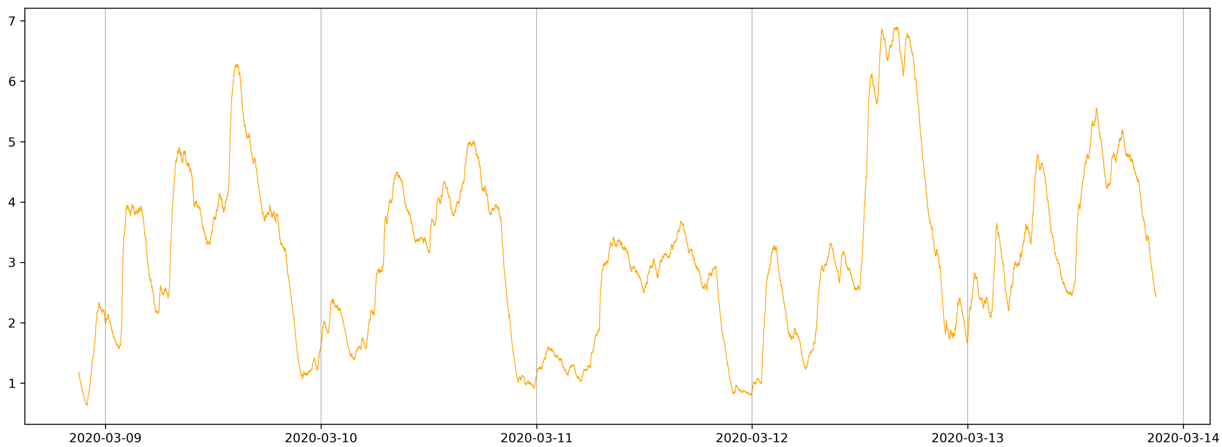


Figure 6: EUR/USD activity during week 11 of 2020



Clustering is observed during periods of volatility, where large returns are followed by larger returns, and during periods of low volatility, where smaller returns are followed by smaller returns. Clustering is also observed in durations and other variables. Figures 7 and 8 present the clustering phenomenon in returns and price durations, respectively, for EUR/USD during week 11 of 2020.

Finally, long-term dependence or memory refers to the ability of a time series to exhibit certain persistent behavior in its structure over time. This property becomes more significant as the frequency of returns increases. Time series with long-term dependence

Figure 7: EUR/USD returns during week 11 of 2020

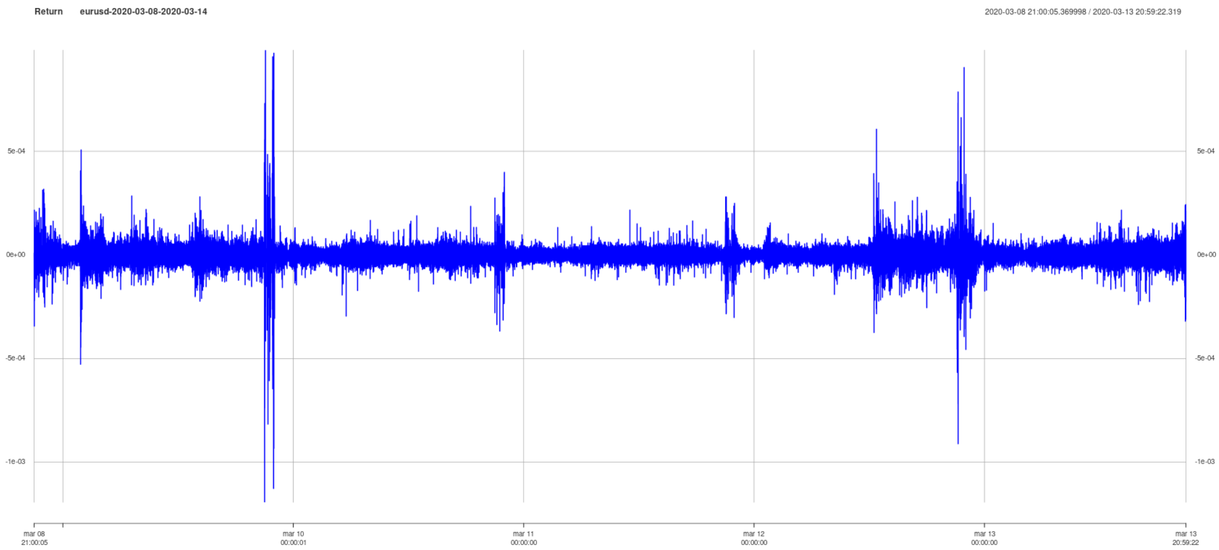
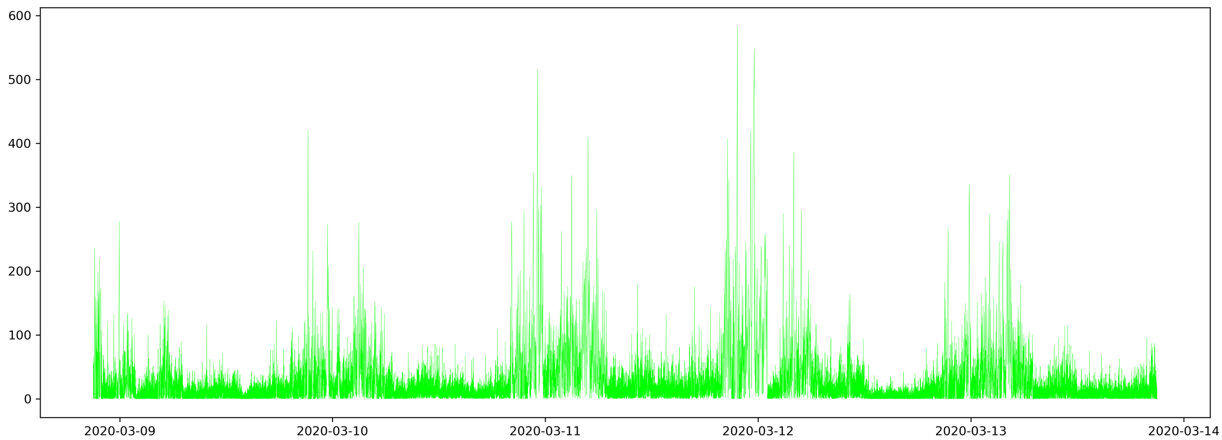


Figure 8: EUR/USD price durations during week 11 of 2020



are valuable because they allow for predicting future long-term behavior based on observed shorter-term behavior through extrapolation of time scales.

2.3.4 Exogenous impacts

Various monetary systems globally lead to different behaviors among currencies, which can be observed at longer-term horizons. Central banks can directly intervene in the market through press conferences, signals, etc., or indirectly through spontaneous arbitrary interventions. Examples of exogenous impacts in the FX market include the discontinuation of the Swiss National Bank's peg with the Euro in January 2015 and the release of macroeconomic data such as Non-Farm Payrolls (NFP) or Federal Open Market Committee (FOMC) announcements among others. These events can cause significant market moves based on the disparity between expectations or consensus and the actual outcomes, influenced by different approaches or subjective interpretations. Other exogenous impacts include WM/Reuters fixings, where bids and offers are aggregated and cleared at a given rate or options expiry, where hedges on options are unwound spot to close out positions. These events contribute to increased market activity.

3 Econometric analysis of high-frequency time series

3.1 Examination of simple point processes and price durations

To ensure the preservation of all original information when standardizing high-frequency data series, the complete set of original data is utilized for constructing durations. According to Engle and Russell (1998) and Hautsch (2012), the time interval between consecutive prices, known as price duration or tick time duration, is closely linked to volatility.

Price durations can be constructed in a more sophisticated manner, such as recursively extracting the first price from the successive prices that exceeds a predefined threshold compared to the immediately preceding one. Along with the extracted prices, their corresponding timestamps and the differences between them, referred to as durations, are recorded.

These differences can be evaluated in terms of absolute value and direction to obtain directional price variations. The resulting series also maintains an irregular time distribution, thus preserving the relevant dynamic information of the original series.

3.1.1 Models for durations

Simple point processes, as described by Hautsch (2012), are stochastic processes consisting of ordered arrival times, excluding the possibility of coincident moments. When these timestamps are accompanied by a corresponding observation, such as a price, they are referred to as marked point processes. In the context of econometric analysis of financial series, the study of point processes can be categorized into four major model families based on the various characteristics or information provided by these processes: intensity models, hazard models, duration models, and count models. The subsequent section focuses on duration models.

As mentioned earlier, durations in high-frequency data series contain information about intraday market activity and the microstructure of the market, including clustering episodes and seasonality. Duration models are employed for the analysis of data with irregular intervals, where the time intervals or durations between observations are considered as a stochastic process. Based on the foundations of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, Engle and Russell (1998) developed the Autoregressive Conditional Duration (ACD) model to characterize the dynamic structure of conditional durations, specifically the conditional intensity of observing events based on previous time intervals.

Let t_i denote the arrival time, and let $d_i = t_i - t_{i-1}$ be the interval between two consecutive arrival times, which will be called the duration. Let $\psi_i = E\left(d_i|F_{t_{i-1}}\right)$ represent the conditional expectation based on all preceding values of the process, i.e., all events forming the filtration $F_{t_{i-1}}$ up to time t_{i-1}. The ACD model of order (p, q) is based on the relationship:

$$d_i = \psi_i \varepsilon_i \tag{4}$$

where $\{\varepsilon_i\}$ is a process comprised of independent, identically distributed positive random variables. To ensure $E(\varepsilon_i) = 1$ the simplest and most appropriate choice is to assume an exponential distribution for ε_i resulting in a model known as $EACD(p, q)$. Additionally, the $ACD(p, q)$ model establishes the following dynamics for the conditional expectations:

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \, d_{i-j} + \sum_{j=1}^{q} \beta_j \, \psi_{i-j} \tag{5}$$

Similarly to GARCH models, the process $\eta_i = d_i - \psi_i$ is considered a martingale, meaning that the conditional expectation $E\left(\eta_i|F_{t_{i-1}}\right) = 0$. The $ACD(p, q)$ model, in terms of the prediction error η_i, can be represented as follows:

$$d_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \, d_{i-j} - \sum_{j=1}^{q} \beta_j \, \eta_{i-j} + \eta_j \tag{6}$$

which can be recognized as an Autoregressive Moving Average (ARMA) process of order $ARMA(\max(p, q), q)$. The stationarity conditions of the ACD model can be determined based on the stationarity conditions of this ARMA model, specifically when the roots of the characteristic polynomials $[1 - \alpha(L) - \beta(L)]$ and $[1 - \beta(L)]$ are outside the unit circle. Here, L represents the lag operator. Once the ARMA model is calibrated, it is used for duration prediction. Consequently, the expected value of durations can be calculated directly as:

$$E(d_i) = \frac{\omega}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j} \tag{7}$$

assuming that $\omega > 0$ and $1 > \sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j$, as expected durations are positive.

However, the $EACD(p, q)$ models have certain limitations in modeling data series. To address this, Zhang et al. (2001) introduced the $GACD(p, q)$ model, which replaces the exponential distribution of ε_i with the generalized gamma distribution. The generalized gamma distribution offers a better fit to real data and greater flexibility. The probability density function of the generalized gamma distribution is defined as:

$$f(x; \kappa, \gamma, \lambda) = \frac{\gamma}{\lambda^{\kappa\gamma}\Gamma(\kappa)} x^{\kappa\gamma - 1} \mathrm{e}^{-(x/\lambda)^{\gamma}} \tag{8}$$

including special cases such as the Weibull distribution when $\kappa = 1$ and the gamma distribution when $\gamma = 1$. The estimation of the parameter θ in the model is performed by maximizing the quasi log-likelihood function:

$$L(\theta) = -\sum_{i=1}^{N(T)} \left[\frac{d_i}{\psi_i} + \log \psi_i\right] f(x;\kappa,\gamma,\lambda) = \frac{\gamma}{\lambda^{\kappa\gamma}\Gamma(\kappa)} x^{\kappa\gamma-1} \mathrm{e}^{-(x/\lambda)^{\gamma}} \tag{9}$$

with proper definitions of the conditional expectations of the process ψ_i. The estimators for the $GACD(p,q)$ model have been shown to be unbiased by Engle and Russell (1998) and Drost and Werker (2004).

Hautsch (2012) provides a series of tests for evaluating the adequacy of the model in the sample, focusing on the dynamics and distributional properties of the estimated residuals $\hat{\varepsilon}_i = d_i/\hat{\psi}_i$ as indicators of a good fit to the underlying series. The residuals in ACD models should be independent and identically distributed to eliminate potential intertemporal dependencies or autocorrelations. To assess these assumptions, portmanteau tests such as the Ljung-Box test and the Box-Pierce test are employed. In this case, only the Ljung-Box test is conducted due to its great accuracy with large sample sizes. The distribution of the estimated residuals $\hat{\varepsilon}_i$ should align with the distribution used for modeling the original series, which can be visually assessed using a Q-Q plot and analyzing its moments. Moreover, for a generalized gamma distribution, the mean should be equal to 1, as stated by Engle and Russell (1998).

3.1.2 Analysis of durations in the spot FX market

The analysis of durations in the spot foreign exchange market can be conducted using high-frequency time series data. The data corresponds to March 10, 2020, a day marked by unprecedented volatility due to the global pandemic of 2020. We select ticks between 13:30 UTC and 14:30 UTC, a time period known for its high concentration of daily market activity during the overlap of the European and American trading sessions, thereby minimizing potential intraday seasonality effects. To focus on actual price movements and eliminate transient values in the series, we consider only observations where the mid-price (to avoid issues related to bid-ask bounce effect) differs by at least one pip from the previous observation. A pip represents the smallest unit of price variation in the FX market, equivalent to one-hundredth of a cent or 10^{-4} of the price. This technique of selecting relevant data within the context of point processes is referred to as thinning.

The durations of price changes are depicted in Figure 9, and their basic statistical measures are presented in Table 1. Out of a total of 394 observations, the average duration between real price changes is 9.1 seconds. The median duration is 6.0 seconds, indicating a significant right-skewed asymmetry (also reflected in the values of the first and third quartiles, 2.9 and 11.3 seconds, respectively). This skewness is further supported by a relatively high positive skewness measure of 2.75.

Figure 9: Histogram showing the distribution of durations

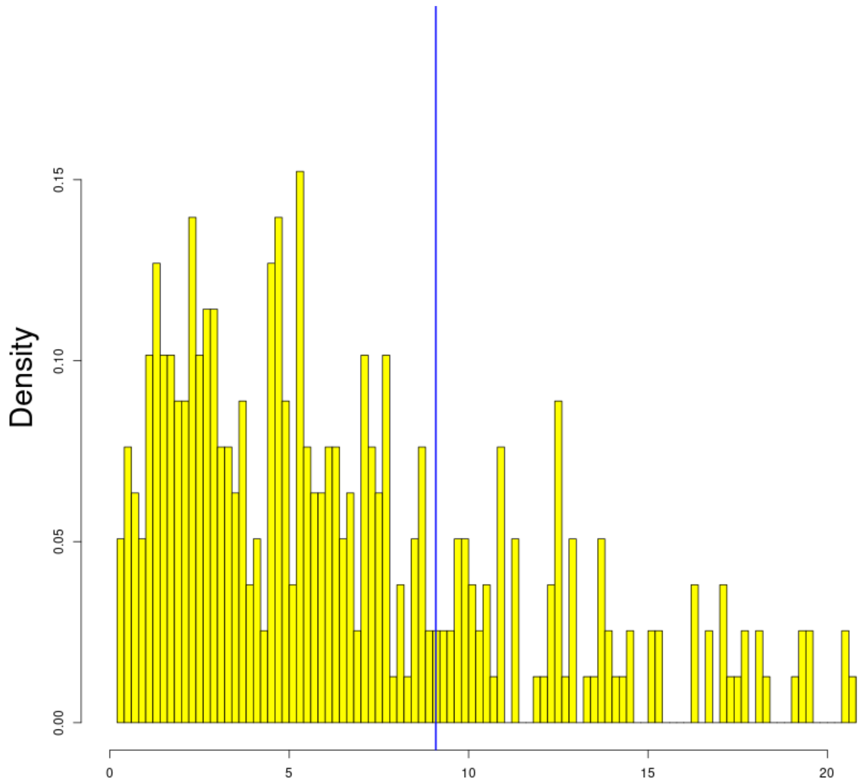


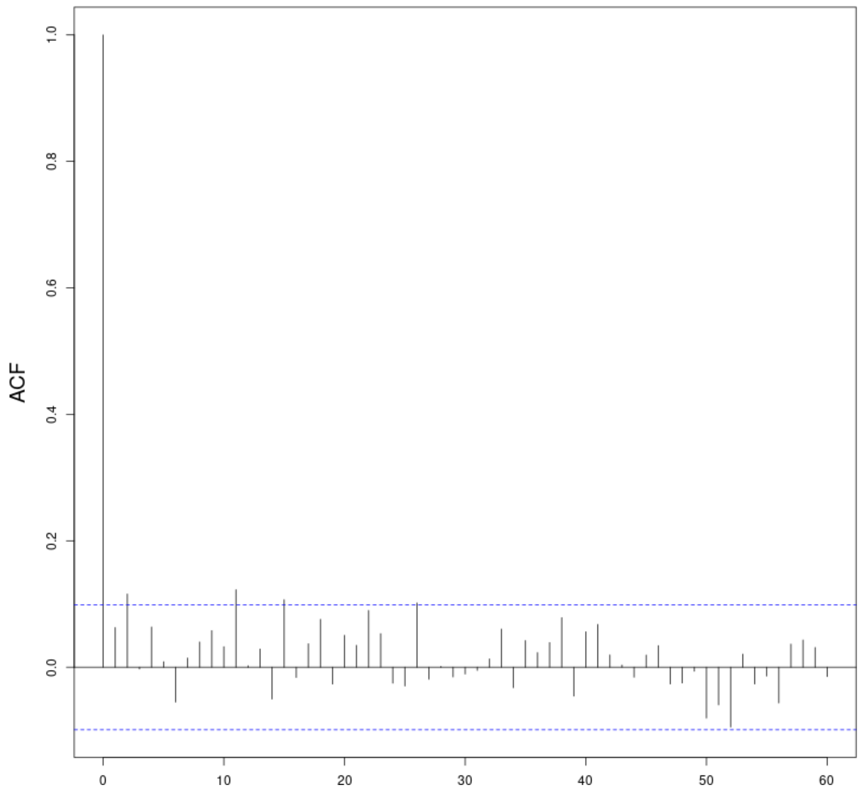
Table 1: Basic statistics of the duration series

N. Obs.	394
Min	0.203
Max	86.048
Mean	9.086
Q1	2.864
Q3	11.328
Median	6.020
SE Mean	0.481
Variance	91.281
Std. Dev.	9.554
Skewness	2.752
Kurtosis	12.660

The autocorrelation function of the duration series indicates a memory-free process, as shown in Figure 10. The simple autocorrelations rapidly decay, ruling out the presence of stochastic trend. The correlogram also suggests the absence of seasonality in significant lags (those lying outside the bands). The simple autocorrelation function demonstrates preliminary stationarity of the duration series, which is further confirmed by conducting the augmented Dickey-Fuller test. The resulting p-value is significantly below 0.05, leading to the rejection of the null hypothesis, which implies the absence of a unit root.

Subsequently, the Box-Jenkins methodology is employed to develop a $GACD(p, q)$ model for the stationary process. Initially, within the framework of model identification,

Figure 10: Autocorrelation function (ACF) plot of the durations



diverse models with different orders are fitted by assigning varying values to p and q. The selection of the most suitable model is based on specific information criteria, as exhibited in Table 2. In this instance, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are utilized, with a preference for Akaike's criterion due to the substantial number of observations in the series. Akaike's criterion evaluates the significance of the chosen parameters as long as there is an adequate amount of data and irrespectively of the model's complexity. The model with the lower information criteria is chosen. Subsequently, parameter estimation is conducted on the selected model, resulting in a $GACD(3,3)$ model. The estimated parameters are presented in Table 3, which includes the parameters of the associated generalized gamma distribution, namely κ, γ, and λ.

In the third step, following the Box-Jenkins methodology, the residuals are scrutinized to ensure the absence of pertinent information and validate the selected model. The autocorrelation function (ACF) plot of the residuals (Figure 11) illustrates that the lags, as they do not surpass the significance bands, lack any discernible structure. This finding is corroborated by the Ljung-Box test, where the p-values for various lag lengths are well above 0.05, thereby leading to the acceptance of the null hypothesis. This acceptance affirms that the residuals are independently and identically distributed, as indicated in Table 4. Furthermore, both the density plot and Q-Q plot of the residuals exhibit a reasonably good fit to the original distribution, i.e., the generalized gamma distribution,

Table 2: Order of candidate models and their respective information criteria

p	q	Log-likelihood	AIC	BIC	MSE
3	3	−1239.433	2496.866	2532.653	86.760
1	1	−1244.326	2498.652	2518.534	89.897
1	3	−1242.961	2499.921	2527.756	89.810
2	1	−1244.209	2500.419	2524.277	89.986
1	2	−1244.994	2501.988	2525.846	89.261
2	2	−1244.455	2502.910	2530.744	89.946
3	1	−1244.650	2503.299	2531.134	89.903
2	3	−1243.704	2503.407	2535.218	89.847
3	2	−1244.558	2505.117	2536.927	89.683

Table 3: Estimation of model parameters and the generalized gamma distribution

	Coefficient	Standard Error
ω	0.791	0.775
α_1	0.075	0.027
α_2	0.068	0.029
α_3	−0.007	0.029
β_1	−0.345	0.042
β_2	0.187	0.052
β_3	0.938	0.039

Parameters: $\kappa = 3.98206$, $\gamma = 0.52799$, $\lambda = 0.060318$.

as depicted in Figure 12. Lastly, an examination of the basic statistics of the residual series in Table 5 reveals that its mean (0.99) closely approximates the theoretical value of 1, indicating the suitability of the $GACD(3,3)$ model for accurately modeling this series of price durations.

Table 4: Ljung-Box test results at various lag lengths

Lag Length	Control Statistic	Degrees of Freedom	p-value
5	3.029	5	0.696
10	6.873	10	0.737
15	13.078	15	0.596
20	15.251	20	0.762
25	20.597	25	0.715
30	23.835	30	0.779

3.2 Marked point processes

In this study, an $ACD(p,q)$ model is developed to analyze the durations d_i of prices in an irregularly distributed time series. The durations d_i can be represented as the product of two components: ψ_i and ε_i. Here, ε_i is independent and identically distributed (i.i.d.) and follows a distribution with a mean of 1 and a variance of σ_ε^2:

Figure 11: Autocorrelation function (ACF) plot of the residuals

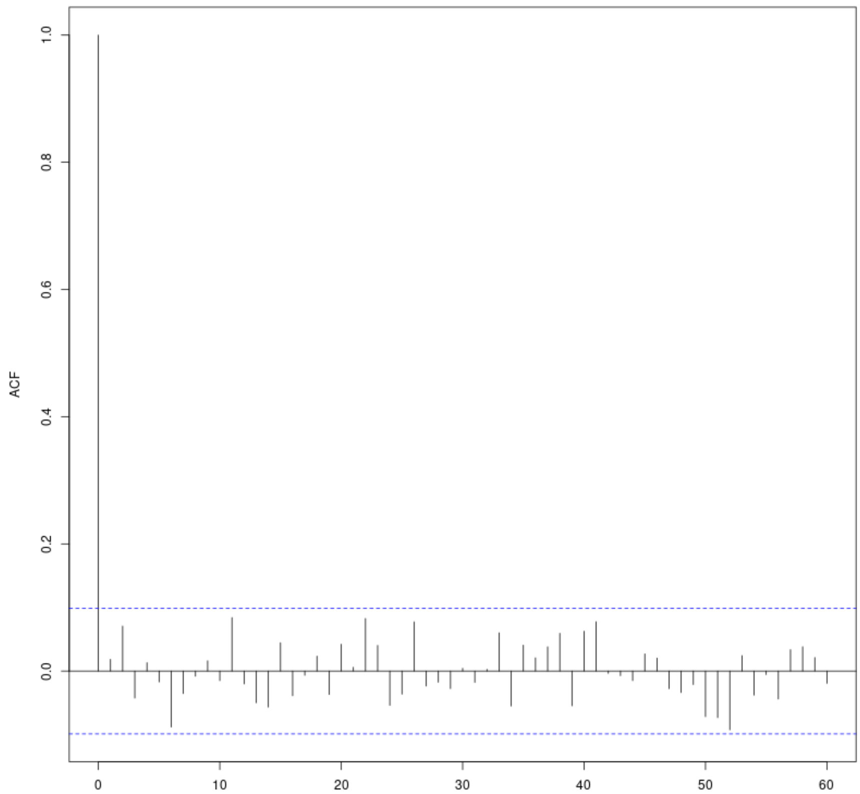
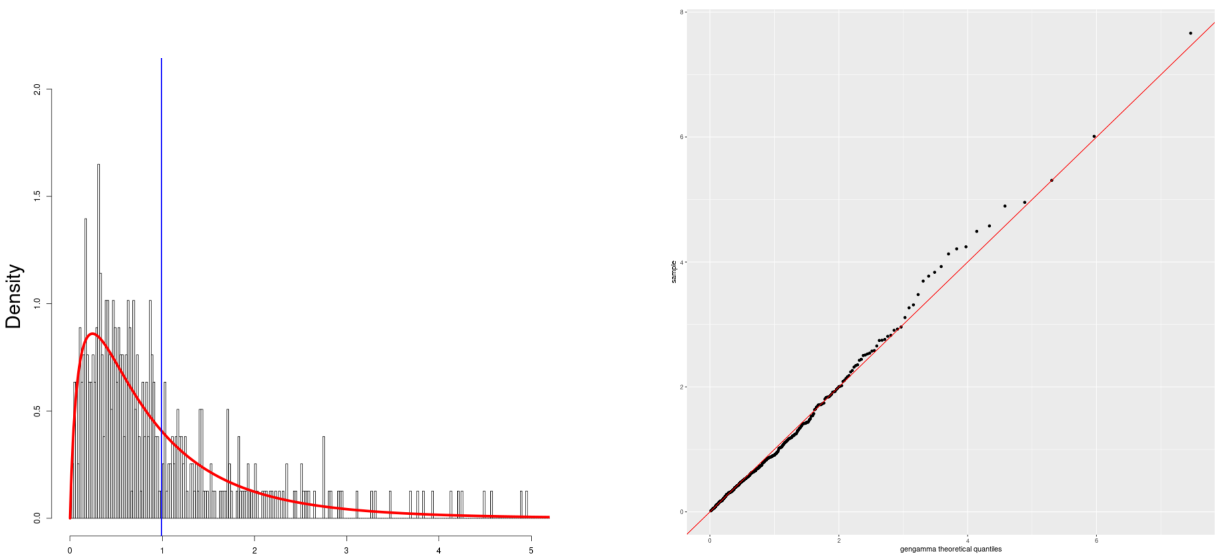


Figure 12: Distribution of the residuals compared to the generalized gamma distribution



$$d_i = \psi_i \varepsilon_i, \qquad \varepsilon_i \sim \text{i.i.d.}\left(1, \sigma_\varepsilon^2\right) \tag{10}$$

The component ψ_i represents the conditional expectation of d_i given the past information $F_{t_{i-1}}$, denoted as $\psi_i = E\left(d_i | F_{t_{i-1}}\right)$ and is defined by the autoregressive conditional duration (ACD) model of order $(1, 1)$:

Table 5: Basic statistics of the residuals

N. Obs.	394
Min	0.016
Max	7.662
Mean	0.992
Q1	0.339
Q3	1.236
Median	0.657
SE Mean	0.051
Variance	1.031
Std. Dev.	1.016
Skewness	2.439
Kurtosis	8.137

$$\psi_i = \omega_d + \alpha d_{i-1} + \beta \psi_{i-1} \tag{11}$$

ACD models are univariate models that capture stylized properties of durations and their correlations, including seasonality, clustering, density, long-term memory processes, and spreads. However, these models do not incorporate the time-dependent influence of price durations on other variables in the model, such as prices, volumes, or returns.

Alternatively, the ACD model could be applied independently to study prices, with each ACD model providing information solely about the modeled variable. Nevertheless, a more comprehensive understanding of market microstructure and improved predictions can be achieved by considering the simultaneous dynamics of multiple variables, given their interdependence. To explore this joint dynamics, multivariate models are necessary.

A simple multivariate model to start with is a bivariate model that includes the duration d_i of prices and the prices p_i as variables. This model allows for the examination of the relationship between these variables and the behavior of prices conditioned on their durations. Following Engle (2000), the joint density of duration and prices can be expressed as the product of the duration density and the conditional price density:

$$(d_i, p_i) \sim f\left(d_i, p_i | F_{t_{i-1}}, \theta\right) = g\left(d_i | F_{t_{i-1}}, \theta_d\right) \cdot h\left(p_i | F_{t_{i-1}}, d_i, \theta_p\right) \tag{12}$$

The $(1,1)$ order model is completed by incorporating the evolution of the conditional expectation of durations, $\psi_i = \omega_d + \alpha_{11} d_{i-1} + \beta_{11} \psi_{i-1}$. Similarly, the dynamics for the conditional expectation of prices given durations, $\varphi_i = E\left(p_i | F_{t_{i-1}}, d_i\right)$, are established as follows:

$$\varphi_i = \omega_p + \alpha_{22} p_{i-1} + \beta_{22} \varphi_{i-1} + \alpha_{21} d_{i-1} + \beta_{21} \psi_{i-1} + \alpha_1 d_i \tag{13}$$

It is important to note that the present value of d_i is involved in these equations. By introducing the martingales $\eta_i = d_i - \psi_i$ and $\zeta_i = p_i - \varphi_i$, a multivariate $ARMA(1,1)$ model is obtained for the variables (d_i, p_i). To incorporate the dynamics of returns r_i into the duration-price model described above, a GARCH model can be utilized:

$$r_i = \mu_i + \sigma_i \zeta_i, \zeta_i \sim \text{i.i.d.}\,(0,1) \tag{14}$$

where the variance of the returns follows the dynamics:

$$\begin{aligned}\sigma_i^2 = \omega_\sigma + \alpha_{33} r_{i-1}^2 + \beta_{33}\sigma_{i-1}^2 + \alpha_{31} d_{i-1} + \beta_{31}\psi_{i-1} \\ + \alpha_{32} p_{i-1} + \beta_{32}\varphi_{i-1} + \alpha_2 d_i + \alpha_3 p_i\end{aligned} \tag{15}$$

Once again, the present values of d_i and p_i are involved in these equations. Ultimately, this model establishes a causal relationship in which durations condition prices, and both durations and prices condition volatility σ_i^2. This relationship is inspired by the work of Manganelli (2005). In terms of martingales, this model can be expressed as a multivariate $VARMA(1,1)$ model for the variables d_i, p_i, and σ_i^2. If one desires to incorporate dependencies among all variables and consider the joint distribution of the three series:

$$(d_i, p_i, r_i) \sim f\left(d_i, p_i, r_i | F_{t_{i-1}}, \theta\right) \tag{16}$$

each of the three dynamics includes present values of all three variables as well as lagged values, capturing the causal effects and feedback among all variables in the model. The model can be expressed analytically using matrices and would remain a VARMA model in terms of martingales.

Furthermore, it is possible to increase the order (p,q) of dependence in the dynamics, which has been restricted to $(1,1)$ in this section for the sake of presentation simplicity. All these models fall under the class of Multiplicative Error Models (MEM), extensively studied in Chapter 7 of Hautsch (2012).

3.2.1 Empirical properties of price durations

Bivariate duration-price models provide valuable insights into the relationships among variables such as prices, volatility, and volumes. Previous research by Pacurar (2008) reveals several empirical properties associated with price durations, including positive autocorrelations, a high level of statistical dispersion characterized by fat tails, strong right-skewness, and asymmetry.

Given that price durations exhibit an inverse relationship with market volatility, they can serve as indicators or signals for analyzing intraday volatility patterns. This, in turn, enables the prediction of price behavior, specifically the amplitude of spreads, which tend to be wider when durations are shorter and narrower when durations are longer.

Furthermore, these models can be utilized to gain insights into traded volumes, as short durations are often indicative of information-based trading, which is typically accompanied by higher trading volumes. Informed traders, possessing insider trading information, have a competitive advantage over uninformed agents or market makers, as explained by O'Hara (1995) in the context of information-based models.

4 Concluding remarks and further extensions

In conclusion, this investigation into high-frequency FX market dynamics through Autoregressive Conditional Duration (ACD) models demonstrates the significant improvements these models offer in understanding the market's microstructural aspects.

Contrary to earlier assumptions, higher frequency data does not obscure insights into market behavior; rather, it enhances them by providing a more detailed understanding of price movements and liquidity dynamics. The utilization of ACD models in this new context not only broadens the theoretical applications of econometric tools but also delivers practical implications for market analysis.

However, it is crucial to acknowledge the evolving nature of the FX market as a potential limitation to this study's findings. As the market continues to undergo significant transformations, the entry of diverse new participants is reshaping trading dynamics and market structure.

These changes include the growing involvement of retail investors and high-frequency trading (HFT) firms. Once considered marginal, retail investors have now become a focal point for banks as their aggregated transactions attract considerable attention. This segment often misinterprets market signals, confusing noise with actionable information, leading to frequent trading losses, as described by King et al. (2013). Despite these challenges, their activities contribute to enhanced market liquidity during periods of lower institutional activity, a dynamic noted by Black (1986).

Furthermore, the widespread adoption of algorithmic trading by HFT firms has introduced new complexities. These firms exploit informational advantages to predict price movements within milliseconds, effectively crossing spreads and securing consistent profits, as discussed by Brogaard et al. (2014).

However, such advancements also bring challenges such as adverse selection, which can disadvantage slower traders and lead to a potentially less stable market environment, a concern highlighted by Biais et al. (2015). Amidst these technological advances, the persistence of voice brokerage underscores the continued importance of human interaction, especially in managing large-volume trades or complex instruments under volatile conditions (King et al., 2013).

Moreover, as banks increase automation, a significant degree of risk internalization occurs, with institutions now offsetting up to 80% of transactions within their own client portfolios to optimize liquidity and reduce hedging costs. This trend has decreased interbank liquidity and contributed to market fragmentation, while the rise of non-bank market makers has forced traditional banks to heavily invest in technology, raising barriers to entry and reshaping the competitive landscape.

As these structural changes continue to unfold, they underscore the need for in-depth research to fully grasp their implications and ensure that econometric models remain relevant and robust in capturing the true dynamics of the FX market.

References

Aldridge, I. (2013). *High-frequency trading: A practical guide to algorithmic strategies and trading systems*. John Wiley & Sons, Inc.

Bank for International Settlements (2019). Triennial central bank survey: Foreign exchange turnover in April 2019.

Biais, B., Foucault, T., and Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial Economics*, 116(2):292–313.

Black, F. (1986). Noise. *The Journal of Finance*, 41(3):528–543.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306.

Carpenter, A. and Wang, J. (2007). Herding and the information content of trades in the Australian dollar market. *Pacific-Basin Finance Journal*, 15(2):173–194.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., and Pictet, O. V. (2001). *An introduction to high-frequency finance*. Academic Press.

Drost, F. C. and Werker, B. J. M. (2004). Semiparametric duration models. *Journal of Business and Economic Statistics*, 22(1):40–50.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162.

Evans, M. D. D. and Rime, D. (2019). Microstructure of foreign exchange markets. Available at SSRN: https://ssrn.com/abstract=3345289.

Hautsch, N. (2012). *Econometrics of financial high-frequency data*. Springer.

Ito, T. and Hashimoto, Y. (2006). Intraday seasonality in activities of the foreign exchange markets: Evidence from the electronic broking system. *Journal of the Japanese and International Economies*, 20(4):637–664.

Kallianiotis, J. N. (2013). *Exchange rates and international financial economics: History, theories, and practices*. Palgrave Macmillan.

King, M. R., Osler, C. L., and Rime, D. (2012). Foreign exchange market structure, players, and evolution. In James, J., Marsh, I. W., and Sarno, L., editors, *Handbook of exchange rates*, pages 1–44. John Wiley & Sons, Inc.

King, M. R., Osler, C. L., and Rime, D. (2013). The market microstructure approach to foreign exchange: Looking back and looking forward. *Journal of International Money and Finance*, 38:95–119.

Maechler, A. M. (2020). FX execution algorithms and market functioning. Markets Committee Papers No. 13. Swiss National Bank. https://www.bis.org/publ/mktc13.pdf.

Manganelli, S. (2005). Duration, volume, and volatility impact of trades. *Journal of Financial Markets*, 8(4):377–399.

O'Hara, M. (1995). *Market microstructure theory.* Blackwell Publishers Inc.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270.

Osler, C. (2009). Market microstructure, foreign exchange. In Meyers, R. A., editor, *Encyclopedia of complexity and systems science*, pages 5404–5438. Springer.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22(4):711–751.

Sun, W. (2007). *Quantitative methods in high-frequency financial econometrics: Modeling univariate and multivariate time series.* PhD thesis, Universität Karlsruhe (TH).

Zhang, M., Russell, J. R., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104(1):179–207.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.