

DETEKTERING OCH VISUALISERING AV LÅNGSIKTIGA AVVIKANDE HANDELSBETEENDEN

MATTIAS PETTERSSON

Examensarbete
2022:E73



LUNDS UNIVERSITET

Matematikcentrum
Matematisk statistik

Master's Theses in Mathematical Sciences 2022:E73
ISSN 1404-6342
LUTFMS-3461-2022
Matematisk statistik
Matematikcentrum
Lunds universitet
Box 118, 221 00 Lund
<http://www.maths.lu.se/>

Abstract

Illegal transactions in the form of money laundering, insider trading and market manipulation is a major problem that is often forgotten. In the year of 2020, 1.6 trillion dollars were successfully laundered, corresponding to 2.7% of the world's total GDP. Insider trading and market manipulation contribute to mistrust of the market, resulting in worse capital distribution from investors to companies as fewer choose to invest. Detecting these problems is very important, and is something that banks and institutions, especially in recent years, have put a lot of resources into. Detecting these problems using statistics is naturally of great importance and is what this report focuses on. Using user and market data to find anomalous trade and transaction patterns is a very useful way to stop these illegalities that often finance very large parts of the world's corruption, terrorism and other illegal activity. In the project, various "measures" are developed and then used to find various types of deviations in individuals' (simulated) trading data. These are visualized in a GUI that makes it easy for the user to apply the measurements to individual participants. Finally, a brief summary is given of how this project and its results can be further developed to find deviations in an even better way and be applied to real scenarios.

Keywords: Money Laundering, Insider Trading, Market Manipulation, Statistics, Outliers, Visualization.

Sammanfattning

Olagliga transaktioner i form av bland annat penningtvätt, insiderhandel och marknadsmanipulation är stora problem som ofta glöms bort. Bara under 2020 tvättades 1.6 biljoner dollar världen över, motsvarande 2.7% av världens totala BNP. Insiderhandel och marknadsmanipulation bidrar till misstro för marknaden vilket gör att den till lägre grad kan allokera pengar till företag från investerare då färre väljer att investera. Att detektera dessa problem är mycket viktigt, och är något som banker och institutioner, speciellt under senare år, har lagt stora resurser på. Att detektera dessa problem på ett statistiskt sätt är naturligt av stor vikt och är vad denna rapport fokuserar på. Att använda sig av aktör- och marknadsdata för att kunna hitta avvikande handels- och transaktionsmönster är ett mycket användbart sätt för att kunna stoppa dessa olagligheter som ofta finansierar mycket stora delar av världens korrruption, terrorism och annan illegala aktivitet. I projektet tas olika mått fram som sedan används för att hitta olika sorters avvikelser i marknadssaktörers (simulerade) handelsdata. Resultatet visualiseras i ett GUI som gör det smidigt för användaren att applicera måtten på individuella aktörer. Slutligen ges en kort sammanfattning av hur detta projekt och dess resultat kan utvecklas för att på ett ännu bättre sätt hitta avvikelser samt appliceras på verkliga scenarion.

Nyckelord: Penningtvätt, insiderhandel, marknadsmanipulation, statistik, extremvärden, visualisering.

Förord

Jag vill tacka mina handledare Björn Thornquist och Christopher Herron på Scila AB för deras idéer och expertis under arbetets gång, samt min handledare Magnus Wiktorsson vid Lunds Tekniska Högskola för hans guidning, kunskap och inspiration under arbetets gång men också under studietiden.

Lund, Sverige

Innehåll

| | | |
|----------|---|-----------|
| 1 | Inledning | 1 |
| 2 | Bakgrund | 1 |
| 2.1 | Insiderhandel | 1 |
| 2.2 | Penningtvätt | 2 |
| 2.3 | Marknadsmanipulation | 3 |
| 2.4 | Hur märks marknadsmissbruk i aktörers datsaserier? | 4 |
| 3 | Instrumentdata | 4 |
| 4 | Aktördata | 8 |
| 4.1 | Olika aktörer | 8 |
| 4.1.1 | Market Maker (MM) | 9 |
| 4.1.2 | HFT-Firma (HFTF) | 9 |
| 4.1.3 | Vanlig Aktör 1 (VA1) | 9 |
| 4.1.4 | Vanlig Aktör 2 (VA2) | 10 |
| 4.2 | Linjär Trend | 10 |
| 4.3 | Extremvärden | 10 |
| 4.4 | Extremvärden i flera dimensioner | 10 |
| 5 | Mått | 11 |
| 5.1 | Extremvärden | 11 |
| 5.2 | Trender | 11 |
| 5.3 | Mahalanobisdistans | 12 |
| 5.4 | Hur många instrument per månad | 13 |
| 5.5 | Köp-/säljvolymkvot | 13 |
| 5.6 | Annulerade/exekverade orderkvot | 13 |
| 5.7 | Medelvärdeskorsning | 13 |
| 5.8 | Genomsnittligt köp- och säljpris | 14 |
| 5.9 | Korrelation till andra aktörers volym | 14 |
| 5.10 | Handel före extrema marknadsrörelser | 14 |
| 5.11 | Sammanfattning av mått | 16 |
| 6 | Implementering i Tkinter | 16 |
| 6.1 | Plot-GUI | 17 |
| 6.2 | Test-GUI 1 | 20 |
| 6.3 | Test-GUI 2 | 23 |
| 6.4 | Test-GUI 3 | 27 |
| 7 | Resultat | 30 |
| 7.1 | Plot-GUI | 31 |
| 7.2 | Vanlig Aktör 1 | 34 |
| 7.3 | HFT-Firma | 43 |
| 7.4 | Market Maker | 44 |

| | | |
|----------|---|-----------|
| 7.5 | Vanlig Aktör 2 | 48 |
| 7.6 | Numeriska Resultat | 50 |
| 7.6.1 | Extremvärden | 50 |
| 7.6.2 | Korrelation till andra aktörer | 51 |
| 8 | Diskussion | 52 |
| 8.1 | Numeriska resultat | 54 |
| 8.2 | Begränsningar för denna typen av undersökning | 55 |

Ordlista

- Instrument: Ett värdepapper, t.ex en aktie eller ett räntepapper.
- Aktör: En individ eller grupp som köper och säljer på marknaden.
- Dataserie: I detta projekt menas t.ex ”köpvoly””, ”säljvoly” eller antal exekverade ordrar. Det är helt enkelt de dataserierna som finns för varje aktör för varje instrument.
- Mått: I detta projekt framtagna metoder för att upptäcka avvikande mönster.
- Likviditet: En marknads likviditet är hur mycket som handlas. Hög likviditet leder till en effektiv marknad med låga spreads samtidigt som marknaden blir svårare att manipulera.
- Market Order: En order som exekveras vid bästa pris, givet nuvarande motparter.
- Limit Order: En order som exekveras vid ett förutbestämt pris.
- Index: En grupp instrument; ofta inom samma sektor, land eller region.
- Beta: Hur mycket ett instruments rörelser beror på övriga marknadens/indexets rörelse.
- Finansinspektionen (FI): Svensk statlig myndighet som ansvarar för tillsyn och övervakning av svenska finansiella marknader.
- United States Securities and Exchange Commission (SEC): USAs motsvarighet till Finansinspektionen. Större befogenheter och ansvarsområde.

1 Inledning

Målet med detta projekt är att tillsammans med värdföretaget Scila AB undersöka och implementera olika metoder för att upptäcka avvikande handelsbeteenden på längre tidsramar, ofta 3-12 månader. En viktig del av arbetet är också visualisering. Att kunna på ett tydligt sätt visa för Scila ABs kunder varför en av deras aktörer har flaggats för avvikande beteende är avgörande för att kunden sedan ska kunna avgöra om de ska skicka detta vidare till Finansinspektionen som i sin tur beslutar om åtal ska väckas. Då just visualiseringen är så pass viktig används inte maskininlärning eftersom detta ofta resulterar i en black-box modell som inte på ett tydligt sätt återger varför en individ flaggades. Rapporten börjar med en kort överblick över vilka olika problem som finns på olika marknader, samt vilka av dessa som detta projekt hoppas kunna upptäcka. Under "Aktördata" inleds diskussionen om hur man genererar dataserier som simulerar olika individers handelsbeteende. Under rubriken "Mått" presenteras de olika tester som har framtagits under processens gång, som sedan är implementerade samt presenterade under rubriken "Implementering i Tkin-ter". Huruvida de olika testerna detekterar olika handelsmönster presenteras under "Resultat", varpå rapporten avslutas med ett diskussionsavsnitt.

2 Bakgrund

2.1 Insiderhandel

Insiderhandel är när en eller flera personer eller enheter köper och säljer värdepapper som de har en personlig koppling till. Detta i sig är inte olagligt, men måste rapporteras till SEC eller Finansinspektionen vilket gör transaktionerna publika. Det man dock ofta syftar på när man pratar om insiderhandel är illegal sådan. Illegal insiderhandel är transaktioner av värdepapper som utförs på grund av information som person(erna) i förtroende har fått ta del av, utan att allmänheten vet om denna information. Sådan information kan till exempel vara en stor nyhet, nyförvärv eller annan information som har stort inflytande över värdepapprets värde. Eftersom olagligheten uppstår när en individ köper eller säljer värdepapper pga information som inte är allmänt känd, är det inte bara chefer och högt uppsatta personer som kan bedriva illegal insiderhandel. Det kan även vara vänner eller familj till dessa, anställda på andra företag som har fått reda på informationen genom att bedriva andra verksamheter med företaget i fråga, regeringsanställda som har tagit del av informationen eller andra personer som på något sätt har kommit över informationen. [1]. Illegal insiderhandel kan till viss del ses som positivt, exempelvis kan det ses som att marknaden blir mer effektiv och exakt, då mer information når marknaden. Om en aktie handlas för 100:-/st och en marknadsaktör som vet om insiderinformation vet att aktien är värd 120:- pga information som släpps om en vecka, kan hen köpa aktien och därmed öka priset till 105:-/st. Detta insiderköp har därmed resulterat i att priset bättre återspeglar all information (även sådan som inte är tillgänglig för allmänheten). Detta överskuggas dock av synvinkeln att marknaden blir orättvis. Att allmänheten fattar misstroende för marknaden pga att de känner sig lurade leder till att de investerar mindre och därmed kan marknaden inte utföra

sin uppgift lika väl. Denna uppgift består av att anskaffa kapital till företag (IPO) och att låta investerare tjäna pengar i förhållande till risken de tar. Marknaden hjälper även regeringar att ta ut skatt [2][3]. Ett av många exempel på insiderhandel är SAC Capital år 2012. I SAC Capitals fall hade (enligt SEC) över 20 olika företag handlats med hjälp av insiderinformation. Ett exempel på information som användes var information gällande en ny Alzheimersmedicin. Genom icke-publik information om specifika företag kunde SAC tjäna hundratals miljoner dollar. De individuella handlarna tjänade också väldigt mycket pengar på bonusar, ofta tiotals miljoner dollar. När fallet uppdagades fick de individuella handlarna böter och/eller fängelsestraff uppemot 10 år. Även SAC Capital fick böta 1.8 miljarder dollar [4]. Finansinspektionen hittar handlare som använder otillåtna medel på främst 2 sätt. Antingen genom att titta på faktiska köp- och säljorder eller liknande för att hitta misstänkta transaktioner, eller om handel görs nära inpå nyhetsläpp. Ett logiskt antagande är att insiderhandlarna oftast köper eller säljer nära inpå en stor nyhet som flyttar värdepapperspriset mycket [5].

Det andra vanliga fallet är att någon tipsar FI. Till exempel om en annan aktör på marknaden har ingått ett kontrakt med insiderhandlaren genom en option, och det underliggande instrumentet ett par dagar senare går upp eller ner så mycket i pris att optionen som aktören sålde nu är värd tio gånger mer, kommer hen antagligen bli misstänksam [6].

2.2 Penningtvätt

Penningtvätt är ett annat stort problem inom den finansiella sektorn. Penningtvätt går ut på att ”tvätta” smutsiga pengar genom att få det att se ut som att de kommer från en legitim källa. Därmed blir ”smutsiga” (olagliga) pengar ”rena” (lagliga) [7]. Penningtvätt är ett enormt problem vilket återfinns i många olika sektorer. Det uppskattas att det endast under 2020 tvättades 1.6 biljoner dollar, motsvarande 2.7% av världens GDP [8]. Det finns flera olika sätt att tvätta pengar. Ett exempel som man kan komma i kontakt med ute i samhället är när unga luras till att bli penningmålvakter. Penningmålvakter tar emot pengar, ofta kontanter, och skickar dessa sedan t.ex med Swish till någon annan, ofta för en liten belöning. Detta gör att mottagaren ser ut att ha tagit emot en vanlig swish, och maskerar alltså det riktiga ursprunget för pengarna. Det kan också ske genom att samma pengar swishas eller överförs många gånger vilket gör det svårare att avgöra var pengarna kom ifrån från början, samt vem de slutligen hamnade hos [9].

Ett annat sätt som pengar tvättas på som är mer relevant i denna rapport är när företag och individer använder sig av värdepapper för att tvätta pengar. Ofta används värdepapper eftersom det försvårar att se hela transaktionskedjan. Till exempel kan en gärningsperson överlåta aktier eller andra värdepapper i utbyte mot smutsiga pengar. När dessa värdepapper sedan säljs ser pengarna ut att komma från en legitim källa. Att använda olika banker och olika värdepapper är ett annat sätt att just försvåra för polis och myndigheter att få en överblick över hur pengar har förflyttat sig. Just denna försvåring är vad som gör att penningtvätt är

vanligt förekommande bland kryptovalutor. Deras nästintill anonyma natur gör att det är enkelt att skicka pengar mellan konton och innehavare utan att bli upptäckt. Ovanpå detta finns det också specialiserade tjänster, så kallade "mixers" som blandar kryptovalutor från olika aktörer och ger dessa nya adresser, vilket gör det i princip omöjligt för myndigheter att följa hur pengarna rör sig [10].

Ett sätt som har blivit implementerat hyfsat nyligen för att stoppa penningtvätt, är så kallade "Know Your Customer (KYC)"-protokoll, där aktören måste ladda upp bevis på identitet i form av ID eller liknande, för att försöka motverka att individer köper och säljer till sig själv med olika konton i olika namn, och på så sätt tvätta pengar [11].

2.3 Marknadsmanipulation

Marknadsmanipulation är ett annat olagligt sätt att använda värdepapper för att tjäna pengar. Detta kan dels ske på mycket korta tidsramar med t.ex layering eller spoofing som är en teknik där man använder falska ordrar för att ändra priset på ett värdepapper. Genom att lägga en stor order på ena sidan trycker man priset åt andra hållet, där man sedan har en order som man vill ska gå igenom. När denna går igenom tar man bort den ursprungliga ordern och har därmed fått ett bättre pris än man egentligen skulle ha fått [12, s.115]. Layering och spoofing används ofta i olika strategier som faller under begreppet momentum ignition. Denna metod går ut på att man i snabb följd köper med stora ordrar för att trigga en hastig uppgång. Förutom att göra andra intresserade av att köpa pga uppgången så triggas andras stoppordrar vilket innebär att dessa också köper. Detta leder i sin tur till att ytterligare stoppordrar triggas ännu högre upp, vilket resulterar i en kaskadeffekt. Därefter säljer man de aktierna man köpte när man startade momentum ignitionen, i hopp om att priset nu är högre än när man initierade strategin [12, s.114].

En annan teknik är pump-and-dump system. Dessa går ut på att artificiellt höja priset på ett värdepapper (pump) i hopp om att sälja det dyrare till någon annan (dump). Detta resulterar ofta i att de som köpt värdepappret på toppen, innan dumpen, sitter med "värdelösa" värdepapper som de har betalt mycket mer för än vad de är värda. Ofta är det värdepapper med låg volym då dessa är lättare att manipulera. Individerna som genomför sådana system trycker ofta upp priset på ett av två sätt. Antingen köper och säljer de till sig själva, alternativt samarbetar med en broker, eller så får de andra att köpa efter att de själva har köpt. Att köpa/sälja till sig själv i annat konto kallas för "Wash trading" och är ett sätt att artificiellt höja värdet eller volymen på ett illikvidt värdepapper. Individerna i fråga köper en aktie för t.ex 5:-/styck, varefter hen säljer för 20:-/styck till sig själv. Det ser då ut som att värdet på aktien är 20:-, medan värdet egentligen inte har ändrats. Det görs dessutom gratis då hen betalar det "dyra" priset till sig själv [13]. Det andra sättet är att först köpa värdepappret själv, för att sedan få *andra* att köpa det, och därmed trissa upp priset. Detta har hänt flertal gånger då sociala medie-stjärnor berättar för sina följare om en ny "fantastisk investering", för att sedan sälja när priset är högt. Detta har blivit speciellt stort med kryptovalutor då amerikanska SEC och andra länders motsvarigheter har haft svårt att komma överens om hur lagstiftning

gen ska se ut. Därmed har det inte varit olagligt att göra pump-and-dumps med kryptovalutor [14]. Pump-and-dumps är på sätt och vis likt momentum ignition, men momentum ignition ses mer som ett resultat av orderboksmanipulation.

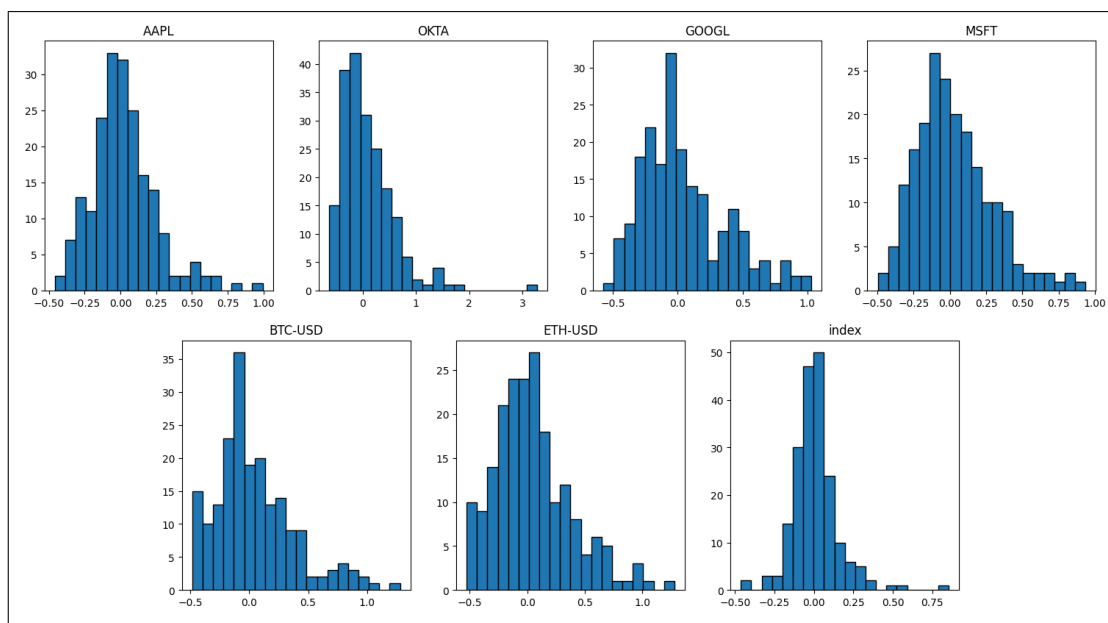
2.4 Hur märks marknadsmissbruk i aktörers dataserier?

Den röda tråden genom alla dessa exempel på olagligheter är avvikande handelsbeteende. Nya instrument, förändrade vanor och förändrad volym är oftast nödvändigt för att kunna tjäna pengar med dessa tekniker. Vid insiderhandel köps en (kanske) tidigare inte handlad aktie då man nu har fått insiderinformation om den. Vid manipulation som pump-and-dump är det också en förhöjd volym av tidigare kanske okända instrument som kan ge ledtrådar. Inom layering och spoofing lägger man väldigt många ordrar som man sedan tar bort innan de exekveras. Även hur ofta man handlar kan ge ledtrådar. Om man tidigare har handlat ett par gånger i månaden men nu helt plötsligt köper oftare, kanske t.om dagligen, kan detta ge ledtrådar om att man skickar aktier fram-och-tillbaka i ett försök att dölja varifrån de kom ifrån från början vilket är en av teknikerna som används inom penningtvätt. Det är därför viktigt att hitta dessa avvikande handelsbeteenden, vilket är fokuset för denna rapport. Vissa av avvikelserna kan dock vara användbara för kunder på andra sätt också, t.ex för att övervaka deras handelssystem. T.ex kan man upptäcka att ett system som ska ha stängt alla positioner över natten ändå håller en nettoposition vid dagens slut.

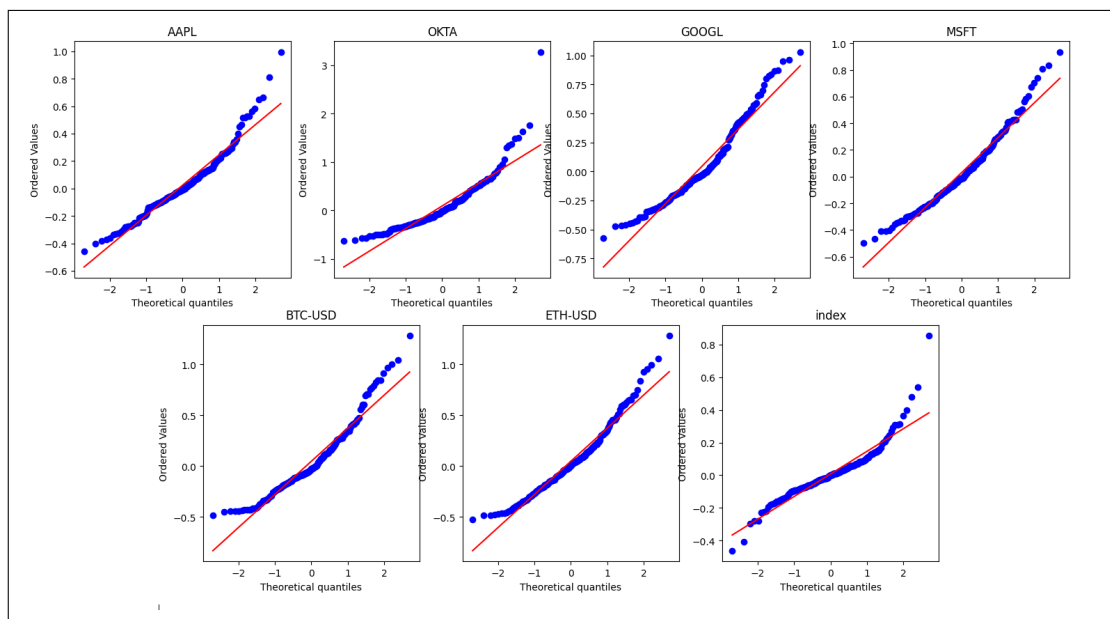
3 Instrumentdata

De olika instrumenten som har valts är aktierna Okta (aktie-ticker "OKTA"), Apple ("AAPL"), Google ("GOOGL"), Microsoft ("MSFT"), samt de två kryptovalutorna (prissatta i Amerikanska dollar) Bitcoin ("BTC-USD") och Ethereum ("ETH-USD"). För samtliga av dessa symboler sätts instrumentpriset till dagens stängningspris, vilket även används till att räkna ut daglig procentuell avkastning genom att räkna ut hur mycket priset har förändrats jämfört med föregående dag. Även instrumentets dagliga volym används, samt dess dagliga omsättning. Omsättningen räknas ut genom att multiplicera instrumentpriset med dagens volym. Från detta räknas sedan den procentuella skillnaden i daglig omsättning ut, på samma sätt som prisets avkastning tidigare räknades ut (i fallet med omsättning används dock log-omsättning för att få normalfördelad data, mer om detta lite senare). All denna data räknas även ut för teknik-indexet Nasdaq100 ("NDX"). Datan hämtas med hjälp av API från Yahoo Finance. Varje instrument kopplas även till ett annat av instrumenten, där idén är att de ligger i samma sektor: Apple-Google, Okta-Microsoft, Bitcoin-Ethereum. Detta används i vissa av måtten. I projektet antas, som ofta inom finans, att den dagliga avkastningen är normalfördelad (figur 5). Däremot är det inte lika vanligt att använda daglig förändring i omsättning. Därför behövs det undersökas vilken fördelning som är lämpligast. Efter att ha plottat datan i ett stapeldiagram (se figur 1 nedan), sågs att svansarna är något större på högerkanten, vilket betyder att de största värdena

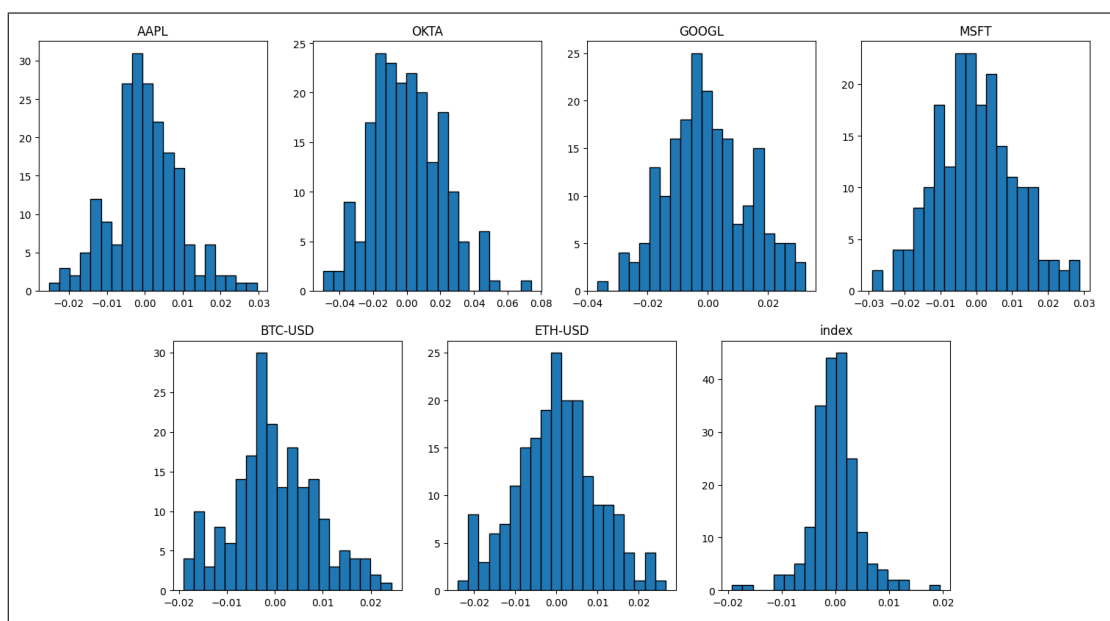
är fler och/eller förhållandevis större än de minsta (när man tittar på avstånd från medelvärdet). Att fördelningen inte är normalfördelad syns också i figur 2 där datan inte följer den röda linjen. Därför transformeras datan med en log-transform för att se om datan då passar bättre in på en normalfördelning. Resultatet syns i figur 3 och 4 nedan. Här syns att datan nu är mycket mer symmetrisk och följer den ideala linjen *mycket* bättre. Den är inte perfekt normalfördelad men för användningsområdet är den bra nog.



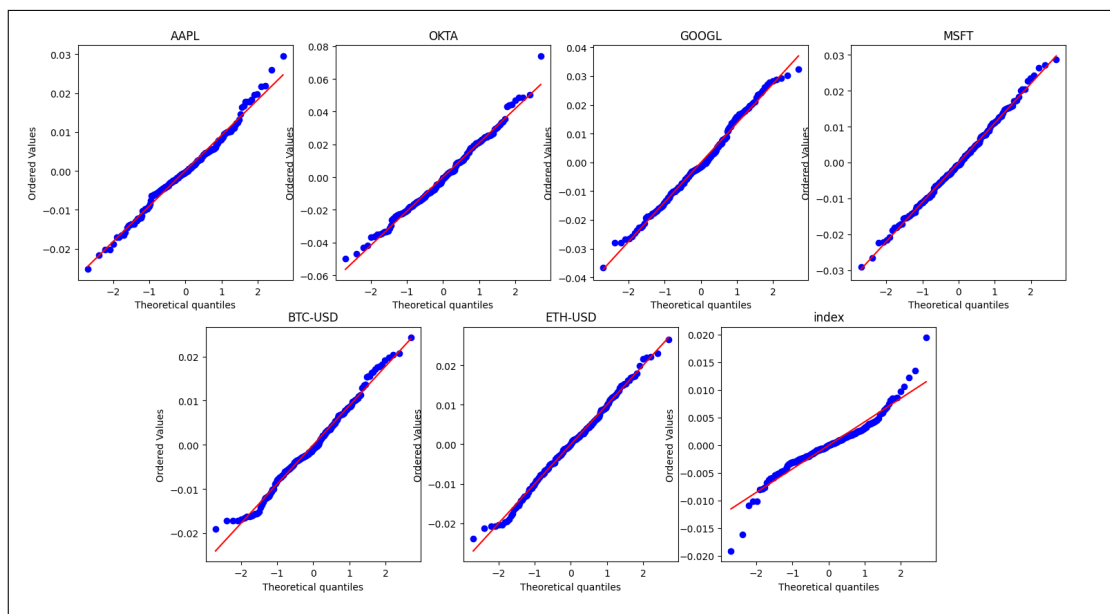
Figur 1: Förändring i omsättning innan log-transformering. X-axeln visar kvoten av dagens omsättning med föregående dags. Y-axeln visar antalet dagar som faller i varje stapel.



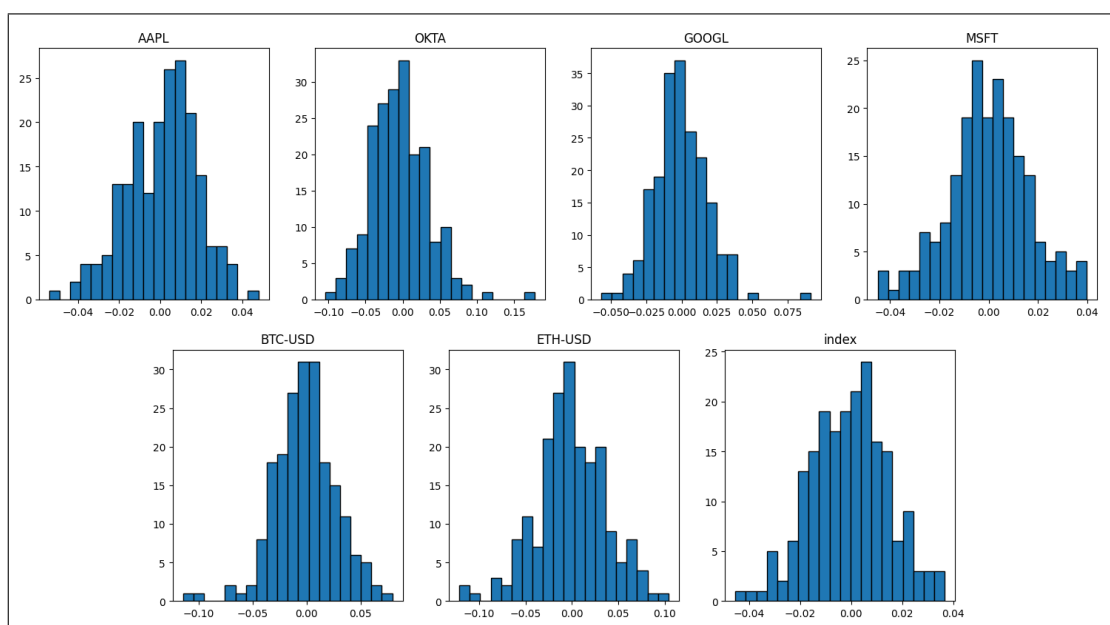
Figur 2: Förändring i omsättning innan log-transformering, plottat i en prob-plot. Det syns tydligt att datan inte är normalfördelad då den inte följer den röda linjen vilket den bör göra.



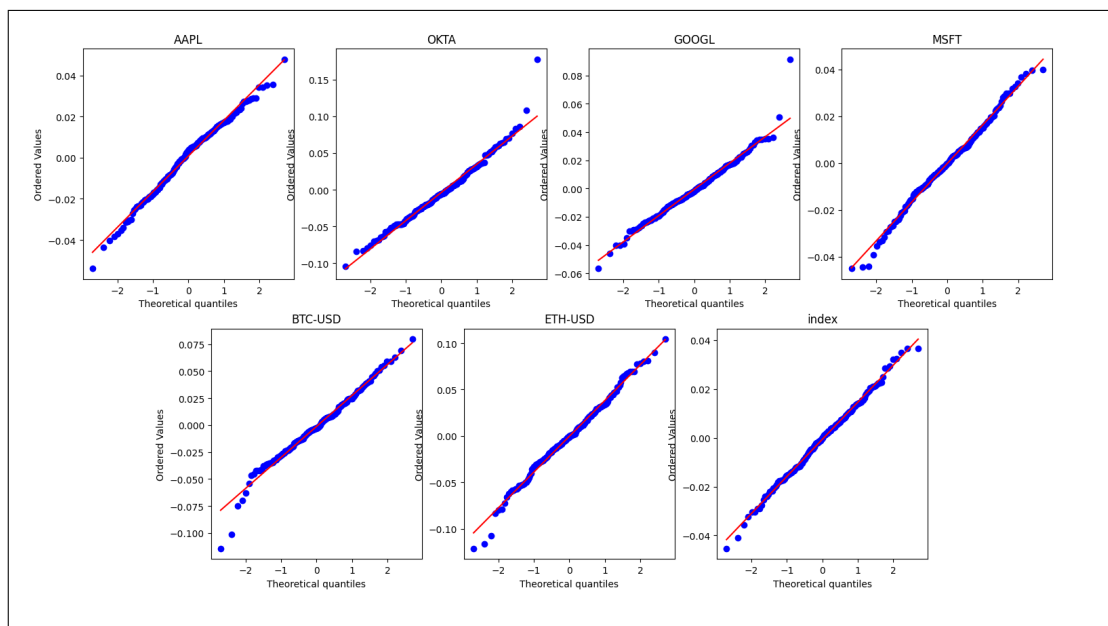
Figur 3: Förändring i omsättning efter log-transformering. X-axeln visar kvoten av dagens omsättning med föregående dags. Y-axeln visar antalet dagar som faller i varje stapel.



Figur 4: Förändring i omsättning efter log-transformering, plottat i en prob-plot. Nu ser datan mycket mer normalfördelad ut. Datat för indexet är fortfarande inte perfekt normalfördelad, men är mycket bättre än före transformeringen.



Figur 5: Avkastning för de olika instrumenten. X-axeln visar kvoten av dagens avkastning med föregående dags. Y-axeln visar antalet dagar som faller i varje bin.



Figur 6: Avkastning för de olika instrumenten, plottat i en prob-plot. Datan är tydligt väldigt nära normalfördelad med endast få avvikande värden.

4 Aktördata

Datan som kommer användas för de olika aktörerna är dagligt aggregerad, alltså finns det en datapunkt per dag per aktör per instrument. Scila AB har även data för mycket kortare tidsramar, men då detta projekt fokuserar främst på längre tidsramar så är detta inte relevant. Dessutom har Scila AB redan färdiga produkter för de korta tidsramarna. Vid projektets start var förhoppningen att kunna använda riktig kunddata men det upptäcktes sedan att detta inte är möjligt pga sekretess. Därför utförs projektet på egen-genererade dataserier. Vad detta betyder är att data genereras i Python för att sedan testa de olika måtten på. Detta är speciellt bra då man själv kan generera data som har de egenskaperna man vill testa och leta efter. Annars hade man i kunders data fått försöka hitta specifika händelser eller egenskaper för att kunna testa måtten. Datan som genereras (volym samt orderhistorik) dras från en normalfördelning (med olika medelvärde och varians beroende på aktör) då aktördata, enligt Scila AB, kan antas ha denna fördelning. Sedan läggs olika speciella egenskaper in. I detta projekt sätts dataseriernas längd till 200 dagar.

4.1 Olika aktörer

Då riktig aktördata inte finns tillgänglig läggs olika fiktiva aktörer in för att göra det tydligare och intressantare för läsaren. På detta sätt förklaras även under vilka förhållanden och situationer som de senare förklarade måtten kan vara användbara. Aktörerna tilldelas dataserier för köp- och säljvolym, samt annullerade och exekverade ordrar som motsvarar hur en sådan aktör handlar på riktiga finansiella marknader.

4.1.1 Market Maker (MM)

En Market Maker är en aktör som har stora mängder värdepapper och handlar dessa med sina kunder. Huvudfunktionen är att erbjuda likviditet till kunderna samtidigt som Market Makern själv tjänar pengar genom att ta ut en spread (skillnad) mellan köp- och säljpriset[15]. I detta projekt karaktäriseras Market Makern av:

- Hög daglig volym (alla instrument).
- Hyfsat jämn daglig volym (alla instrument).
- Högre volym de dagar då turnover ändras mycket (AAPL).
- Både hög sälj- och köpvolym, jämn kvot runt 1 mellan köp-och sälj (alla instrument).
- Handlar alla instrument varje dag.
- Linjär trend (ETH-USD).
- Handlar mer innan stora förändringar i pris, dock inte om man tar bort indexberoendet (OKTA).

4.1.2 HFT-Firma (HFTF)

HFT-Firma, eller High Frequency Trading Firma, är en aktör som handlar otroligt många gånger per dygn. Ofta har positionerna en livslängd på enstaka sekunder. Detta är en form av algoritmisk handel där datorer försöker göra vinster med hjälp av olika strategier. Ett exempel på strategi kan vara att försöka hitta arbitrage mellan olika valutaväxlingskurser. I detta projekt karaktäriseras HFT-Firman av:

- Hyfsat hög volym (inte som market maker).
- Många annullerade ordrar (två instrument).
- Många ordrar (två instrument).
- Lika hög sälj- och köpvolym (två instrument).
- Handlar bara i två instrument.
- Helt plötsligt ändras det så HFT-firman köper mindre än den säljer vissa dagar (GOOGL).

4.1.3 Vanlig Aktör 1 (VA1)

Vanlig Aktör 1, även kallad Normal Trader 1 är en aktör som inte är någon speciell form av handlare. Denna aktör har ett handelsmönster som innehåller:

- Har några extremvärden (BTC-USD).
- Stor Mahalanobisdistans (GOOGL) (Mahalanobisdistans förklaras senare i kapitel 5.3).

- Handlar mycket innan de dagar som priset ändras mycket (OKTA).
- Handlar i GOOGL (samma sektor som OKTA) innan de dagar som priset ökar i OKTA.
- Har ett instrument där Moving average funkar speciellt bra (AAPL).
- Nytt instrument helt plötsligt (ETH-USD).
- Köp- och säljvolymen har stor korrelation till VA2 (nedan) i AAPL.

4.1.4 Vanlig Aktör 2 (VA2)

Vanlig Aktör 2, även kallad Normal Trader 2 är ett konto som tillhör Aktör 1, men som är skrivet i annat namn. Detta för att kunna utföra Wash Trading (tidigare förklarar) alternativt penningtvätt. Den har också enstaka andra egenskaper så som:

- Handlar väldigt likt Vanlig Aktör 1 i AAPL fast omvänt. Detta betyder att VA2 säljer när VA1 köper och vice versa.
- Nettoköpvolym som ökar innan stora prisförändringar, och minskar efter (BTC-USD).
- Handlar bara AAPL och BTC-USD.

4.2 Linjär Trend

En av egenskaperna som läggs in i aktördatan är linjära trender. Dessa genereras genom att för varje dag dra ett värde från en normalfördelning och sedan addera en konstant (σ) som multipliceras med det nuvarande tidssteget:

$$\begin{aligned} y_t &= \sigma t + \epsilon \\ \epsilon &\sim \mathcal{N}(\mu, \sigma^2) \end{aligned} \tag{1}$$

På detta vis fås data som trendar samt har viss variation.

4.3 Extremvärden

Extremvärden introduceras genom att sätta vissa värden (för hand) till ett annat än vad det slumpmässigt hade blivit.

4.4 Extremvärden i flera dimensioner

Extremvärden i flera dimensioner läggs också in. Detta för att visa Mahalanobis-distanst. Målet med dessa punkter är att de inte ska synas i varje dimension för sig, men bli tydliga om man kollar på alla dimensioner samtidigt. Därför är värdena stora/små, men inte extremt stora/små. Alla dessa ligger på samma dag (i olika dataserier).

5 Mått

I denna rapport återfinns ett antal olika mått. Dessa mått är olika tekniker för att hitta avvikande värden i dataserierna. De tittar på många olika saker, allt från enkla 1-dimensionella extremvärden till extremvärden i flera dimensioner, korrelationsförändringar och trender. Vissa av måtten är klassiska statistiska metoder medan andra är utvecklade just för detta projekt tillsammans med Scila AB.

5.1 Extremvärden

För att hitta extremvärden i dataserierna används den klassiska metoden att hitta medelvärdet för datan, och sedan lägga till +/- ett antal standardavvikelser. De datapunkter som sedan hamnar utanför detta konfidensintervall antas vara extremvärden. Eftersom datan förmodas vara normalfördelad bör detta vara en rimlig modell. Innan detta mått används dras den linjära trenden av datan bort. Detta för att få en mer lämplig fördelning för att titta på extremvärden. Tar man inte bort trenden kommer en trendande dataserie inte följa en normalfördelning. [16].

Förutom att applicera detta mått på den råa dataserien (utan trend), appliceras den även på en serie skapad genom att ta skillnaden mellan dagens data och gårdagens. Detta hittar därmed extremvärden i förändringar från dag till dag.

Att bara ta medelvärdet och standardavvikelsen över *hela* dataserien blir inte korrekt då konfidensintervallet isåfall beror på värden som inträffar i "framtiden". T.ex kan ett värde markeras som ett extremvärde om dagarna efter extremvärdet har väldigt små värden, medan det inte var ett extremvärde under dagen som extremvärdet inträffade, eftersom tidigare dagar hade väldigt höga värden. Ett sådant statistiskt konfidensintervall kan därmed resultera i att samma punkt markeras som ett extremvärde ena dagen, men inte nästa. Därför räknas istället ett *glidande* konfidensintervall fram, där användaren specificerar antalet standardavvikelser och längden på intervallet. Sedan markeras extremvärden som hamnar utanför detta *glidande* konfidensintervall. Ett exempel syns i figur 13.

5.2 Trender

För att hitta trender i datan appliceras ett mått för detta. Det mått som har använts är ADF-testet, Augmented Dickey-Fuller testet. Detta test testar om det finns någon enhetsrot för datan. En enhetsrot betyder att det finns en AR-parameter som är lika med 1. Nedan följer en enkel härledning av Dickey-Fuller testet, ett snarlikt test men som endast testar AR(1)-processer medan Augmented Dickey-Fuller testet testar AR-processer med godtycklig längd.

Matematiskt kan en AR(1)-process skrivas som:

$$\begin{aligned}y_t &= \rho y_{t-1} + \epsilon_t \\y_t - y_{t-1} &= (\rho - 1)y_{t-1} + \epsilon_t\end{aligned}\tag{2}$$

där ϵ är den normalfördelade variationen, och ρ är en konstant. Om $\rho = 1$ blir förändringen för varje tidssteg endast lika med den normalfördelade variabeln ϵ . Detta är nollhypotesen för testet:

$$\begin{aligned} H_0 : \rho - 1 &= 0 \\ H_1 : \rho - 1 &< 0 \end{aligned} \tag{3}$$

Ett test kan sedan utföras där $\rho - 1$ testas mot *Dickey-fuller fördelningen* vilket lämnas till den intresserade läsaren. Det lite mer avancerade Augmented Dickey-Fuller testet bygger i huvudsak på samma idé, men där processen inte begränsas till att ha längd 1, och därmed blir mer generell. [17, s.813-836].

I GUIt kan man även rita trendlinjen eller ta bort en trend från dataserien. Detta använder ordinary least squares för att hitta den bäst passande trenden. Detta fungerar genom att minimera avståndet från en trendlinje till samtliga punkter, genom att ändra lutningen på linjen tills det att den passar bäst. Det kan även göras direkt med matriser vilket är mycket snabbare. Härledningen av detta uttryck utelämnas men kan läsas i [18, s.160].

$$\hat{\theta}_{LS} = \arg \min_{\theta} \|y - \mathbf{X}\theta\|_2^2 = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{y}$$

För att detta ska fungera med de datapunkter som finns görs x-värdena om från datum till en sträng av på varandra följande heltal. Därmed är det mycket enkelt att hitta trenden. För att göra en trendande serie stationär (till exempel när man vill hitta extremvärden), drar man helt enkelt bort trend-linjens värde i varje punkt i serien. Måttet returnerar både om ADF-testet hittar en trend eller inte, samt lutningen.

5.3 Mahalanobisdistans

Mahalanobisdistans är ett mått som kontrollerar om det finns extremvärden i flera dimensioner. Detta mått fungerar bäst om datan är normalfördelad samt att det har en linjär beroendestruktur. Vad måttet i princip gör är att mappa varje punkt från ett n-dimensionellt rum till en 1-dimensionell skalär enligt följande. Givet att dataserien följer en N-dimensionell fördelning \mathbb{R}^N där varje x har ett värde $x = (x_1, x_2, x_3, x_4, \dots, x_N)^T$ och medelvärdet för serien är $\mu = (\mu_1, \mu_2, \mu_3, \mu_4, \dots, \mu_N)^T$, så skrivs Mahalanobisdistansen

$$MD = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

där S är kovariansmatrisen för x [16, s.10]. På detta vis kan man hitta punkter som inte ser konstiga ut för varje individuell dimension, men som i flera dimensioner kan ha extremvärden. Ett exempel är att en man som är 190 cm lång inte är något konstigt, inte heller om en man väger 50 kilo. Däremot om detta är samma person så är det absolut ett extremvärde i två dimensioner. Sådana situationer hittar Mahalanobisdistansen. I denna rapport används köpvolym, säljvolym, exekverade ordrar, samt borttagna ordrar, alltså blir det ett 4-dimensionellt rum. För att identifiera om en punkts Mahalanobisdistans faktiskt är ett extremvärde jämförs

värdet med en chi-kvadratfördelning. Genom detta fås ett p-värde för varje punkt vilket sedan kan användas för att identifiera hur extremt ett värde är [19, 10]. Även i detta fall är medelvärdena för varje dimension glidande.

5.4 Hur många instrument per månad

Detta är ett egengjort test som plottar antalet gånger ett instrument har handlats för varje månad. Här kan man enkelt se om ett nytt instrument har introducerats, tagits bort, eller om ett instrument helt plötsligt handlas mycket oftare än andra. För varje månad får varje instrument en stapel. Denna stapels höjd är antalet dagar som instrumentet handlats.

5.5 Köp-/säljvolymkvot

Detta test plottar förhållandet mellan köp- och sälj volym för varje dag. Detta kan visa om en handlare helt plötsligt köper eller säljer mer än hen brukar i förhållande till hur mycket hen säljer eller köper. Att bara titta på volym ger inte en tydlig bild av vad en individ gör då detta inte visar om man köper mer och säljer mindre, eller om man köper och säljer lika mycket etc. Om en handlare slutar köpa och istället börjar sälja är detta ett avvikande handelsbeteende, även om den totala volymen är samma som tidigare. Ett annat sätt att visualisera detta är att titta på aktörens nettoköpvolym vilket är köpvolymen minus säljvolymen för varje dag. Ett annat exempel på när detta mått kan användas är för High Frequency Traders (HFT). Dessa handlar på väldigt kort tidsram (sekunder till minuter), och stänger oftast alla sina positioner innan dagens slut. Detta betyder att köp- och säljvolymen för varje dag borde vara lika stor (alla positioner som öppnas, stängs också). Därmed borde kvoten mellan dessa vara 1. Om helt plötsligt kvoten förändras från 1, visar detta ett avvikande beteende. Detta behöver dock inte tyda på olagligheter, det kan även betyda att deras system inte fungerar önskvärt (att hålla positioner över natten/olika sessioner betyder ofta förändrad riskprofil).

5.6 Annullerade/exekverade orderkvot

Detta mått är ett mycket intressant mått. Det visar på förhållandet mellan annullerade och exekverade ordrar. Detta förhållande bör vara ganska lågt då varje order rimligtvis har målet att exekveras. Att då ha ett högt antal annullerade ordrar per exekverad order är anledning till att bli misstänksam. Om kvoten är väldigt hög kan det också vara ett tecken på layering eller spoofing. HFT-Firmor kan dock ofta utslutas från detta, då de använder höga antal annullerade ordrar på ett legitimt sätt.

5.7 Medelvärdeskorsning

Ett annat mått är att använda olika glidande medelvärden (Moving Averages, MA) på en dataserie. Att använda t.ex ett medelvärde med längd 10 och ett annat med längd 20, och titta på när dessa korsar varandra, kan vara ett enkelt sätt att se att handelsbeteendet har ändrats. Om det MA med längd 10 korsar det MA med

längd 20, betyder detta att den kortsiktiga trenden har brutit sig loss från den långvariga. Detta är ett bra komplement till att kolla på total trend över tidsperioden, då medelvärdeskorsningen kan hitta mindre avvikelser.

Om dataserien är väldigt konstant kan det korta MA t korsa det långa väldigt ofta. Därför har ett filter lagts in i form av att ett minsta avstånd (som användaren specificerar) mellan medelvärdena måste ha uppnåtts sedan förra korsningen, för att nästa korsning skall räknas. Detta betyder att endast större svängningar fås med, inte små korsningar som uppstår när två MA ligger i princip på varandra och korsar varandra nästintill dagligen.

5.8 Genomsnittligt köp- och säljpris

Detta mått plottar snittpriset för köp- och säljvolym för varje dag. Måttet är tänkt som en övergripande bild av hur handlandet går för aktören. Det är ett mycket användbart mått om en individ handlar i detta instrument väldigt sällan (3-4 transaktioner på 6 månader), då en sådan aktör är svår att analysera på övrigt sätt. Om t.ex det genomsnittliga köppriset är 100:-/aktie och säljpriset en vecka senare är 160:-/aktie, kan detta vara anledning till fortsatt undersökning.

5.9 Korrelation till andra aktörers volym

Ett annat mått som analyserar en specifik aktör. Genom att titta på korrelationen mellan en aktörs köpvolum och alla andra aktörers säljvolym och vice versa kan man se om det finns aktörer som handlar samma volym, fast när den ena säljer köper den andra. Detta kan vara ett sätt att hitta penningtvätt eller wash trading. Om korrelationen är hög mellan två konton (korrelation mellan aktör 1's köpvolum och aktör 2's säljvolym), är chansen stor att de två kontoinnehavarna tvättar pengar genom att sälja och köpa till varandra, eller helt enkelt att båda kontona innehavs av samma individ. I kapitel 7.6.2 beskrivs nackdelen med att detta mått kan ge falska utslag om en handelsplattform har (väldigt) många aktörer då detta kan resultera i höga korrelationer endast på grund av slump. I samma kapitel visas dock också att risken för detta fenomen antagligen är liten för de flesta handelsplattformarna.

5.10 Handel före extrema marknadsrörelser

- Ett vanligt sätt att upptäcka insiderhandel är att en aktör köper eller säljer en aktie före en kraftig upp- eller nedgång. Detta mått räknar ut snitt-volymer under de n (specificerat av användaren) dagar som leder upp till en dag med en avkastning som ligger över ett visst antal standardavvikelser från medelnivån. Måttet ger sedan kvoten mellan detta snitt och snittet för alla andra dagar aktören har handlat. Precis som i tidigare fall så är standardavvikelsena glidande med en längd som sätts av användaren.
- Det kan dock vara så att den stora förändringen i pris en viss dag beror på en stor marknadsförändring, t.ex en sektor-specifik nyhet eller något som påverkar marknaden som t.ex Corona-viruset gjorde, eller en höjning i räntor från en

centralbank. Isåfall kommer detta mått i princip triggas för alla aktörer, trots att de inte har haft insiderkunskap om en viss aktie. Därför finns det även ett mått som ger snittet för de dagar som aktien har rört sig utanför x antal standardavvikelser, men där de dagar som även indexet rörde sig utanför x antal standardavvikelser (för indexets prisserie) har tagits bort. Detta i förhoppning om att ta bort dagar som hela marknaden rör sig ovanligt mycket.

- Det finns dock ett problem med denna lösning. Om en akties pris går upp otroligt högt, kanske utanför 4 standardavvikelser, men indexet bara rör sig utanför med 2 standardavvikelser, så kommer man sortera bort denna dag, trots att aktiens rörelse bara till viss del förklaras av index-rörelsen. Därför finns det en tredje version, nämligen beta-justerad avkastning. Denna funktion räknar ut betat för aktien under hela tidsperioden och använder sedan detta för att filtrera bort indexets rörelse från aktierörelsen. Detta lämnar sedan hur aktiens pris har rört sig, oberoende av hur indexet rörde sig. På denna avkastning används sedan samma metod med x antal standardavvikelser och kvoten mellan de dagar innan stora rörelser, och övriga dagar. Betat är inte glidande utan antas samma under hela tidsperioden.
- Förutom stora förändringar i aktiepriset så kan det också ske stora förändringar i omsättning för en aktie. T.ex vid penningtvätt kan ett litet instrument helt plötsligt få väldigt hög omsättning, medan priset hålls artificiellt lågt för att undvika misstanke. Detta kan t.ex göras genom att två aktörer säljer till varandra genom limit orders, istället för market orders. Därför finns även exakt samma tre mått som ovan, fast istället applicerat på omsättning. Alltså hur mycket en aktör har handlat de dagar som leder upp till ovanligt hög omsättning. Hög omsättning kan också visa på dagar som priset faktiskt går upp/ner, trots att det kanske inte riktigt går upp/ner så många standardavvikelser som behövs för att avkastnings-måttet ovan ska triggas.
- För dessa olika mått har även kvoterna för en *annan* aktie lagts in. Alltså hur en aktör har handlat i en annan aktie (i detta fall i samma sektor), under de dagar som leder upp till en stor förändring i den första aktien. Detta för att försöka detektera individer som vet om att en stor prisökning kommer ske i t.ex tekniksektorn eftersom hen jobbar på Google, men som istället handlar i Apple för att undvika att bli upptäckta. Detta då de vet att folk ofta har noggrannare koll på hur en högt uppsatt person i ett företag handlar i sitt eget företag, än vad hen handlar i en konkurrent eller annat företag i samma sektor.

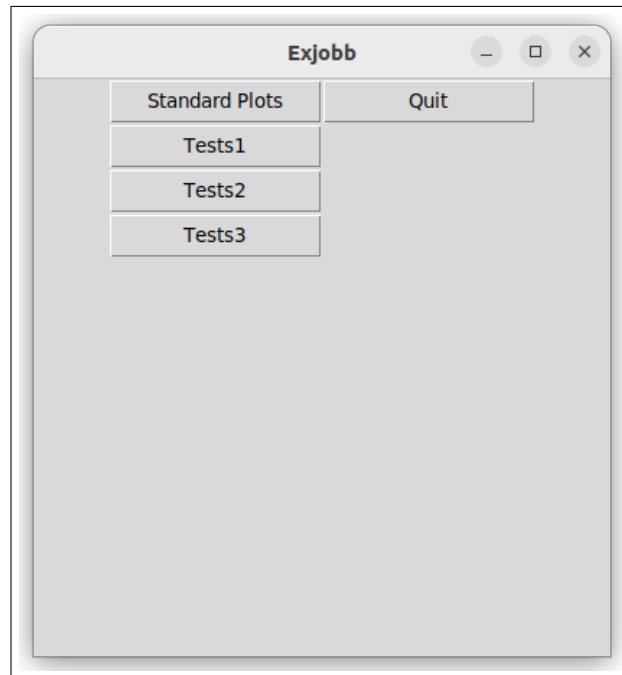
Användaren kan också välja att visa hur aktören handlade de dagar som aktien har gått *ner* mer än x standardavvikelser. Detta kan vara intressant om man tittar på om aktören har hög säljvolym inför dessa dagar. Slutligen finns det ett mått som visar om aktören hade speciellt hög volym de dagar som ett av avkastning-måtten och även ett av omsättnings-måtten gav utslag. Vilka två av dessa måtten väljer användaren själv i menyn. Läs mer i Tkinter-avsnittet nedan.

5.11 Sammanfattning av mått

- Extremvärden - Detekterar värden utanför x standardavvikelse från det glidande medelvärdet.
- Trender - Detekterar om det finns någon total trend över hela fönstret (200 dagar).
- Mahalanobisdistans - Detekterar extremvärden i flera dimensioner.
- Antal instrument - Visar hur många instrument som handlas varje månad.
- Köp-/säljvolymkvot - Detekterar om förhållandet mellan köp- och säljvolym förändras.
- Annullerade/exekverade orderkvot - Detekterar om förhållandet mellan annullerade och exekverade ordrar förändras.
- Medelvärdeskorsning - Detekterar förändringar i trend på kortare tidsramar än över hela fönstret.
- Genomsnittligt köp- och säljpris - Ger en överblick över vilka priser en aktör köper och säljer för.
- Korrelation till andra aktörer - Detekterar om en aktör handlar mycket med andra specifika aktörer.
- Handel före extrema marknadsrörelser - Detekterar om en aktör köper eller säljer ovanligt många kontrakt inför stora instrument- eller marknadsrörelser.

6 Implementering i Tkinter

I Tkinter (Python) skapas de visuella hjälpmedlena. Utgångspunkten är ett fönster (figur 7) med knappar som antingen tar användaren till plot-GUI:t eller till ett av de tre test-GUI:n.



Figur 7: Startfönster där man väljer vilket plot- eller testfönster som skall öppnas.

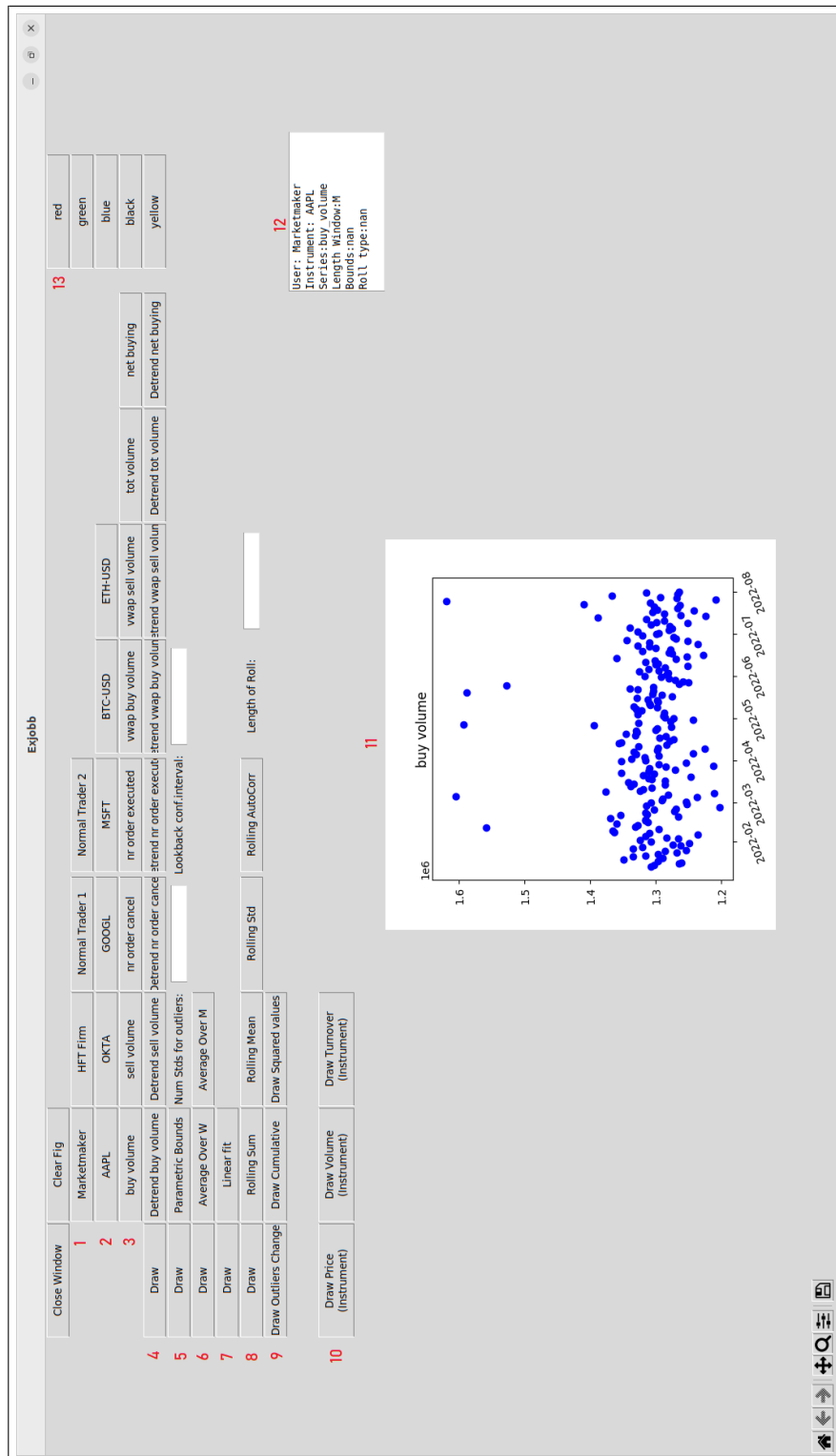
6.1 Plot-GUI

Plot-GUIt är ett GUI som låter användaren plotta en specifik serie genom att välja user (aktör), instrument och dataserie. Användaren kan sedan välja olika färger för att plotta olika saker på denna serie, till exempel övre och undre gräns för extremvärden, trender, ta bort trender eller ett genomsnittligt värde för varje vecka eller månad. Det finns också en plott som visar ett glidande fönster med en användar-bestämd längd (x) som plottar summa, standardavvikelse eller genomsnitt för (x) punkter bakåt i tiden.

Här nedan finns fönstrets alla funktioner förklarade. Alla referenser till figurer i denna tabell syftar på siffror/figurer i tillhörande bild 8.

1. Aktör.
2. Instrument.
3. Dataserie.
4. Dataserie där den linjära trenden har subtraherats, samt draw-knapp som ritat valet som gjorts i 1-4.
5. Välj vilken sorts glidande konfidensintervall (just nu finns bara parametriskt, men möjligheten finns att lägga till andra), samt draw-knappen för konfidensintervallet. Det finns även fält där användaren skriver in hur många standardavvikelser samt hur många dagar konfidensintervallet räknas ut med.

6. Välj om man ska rita upp snittet för varje vecka eller månad, samt draw-knappen som ritar det gjorda valet i figuren.
7. Vilken sorts trend-fit man ska göra, samt tillhörande draw-knapp. Även i detta fall finns bara en sorts trend, men det finns möjlighet att lägga till andra.
8. Olika glidande plottar. Summa, medelvärde, standardavvikelse, samt autokorrelation kan ritas på plotten. Ett fält för hur långt det glidande fönstret skall vara finns med, samt en draw-knapp.
9. De tre knapparna på denna rad ritar:
 - (a) Förändringen från föregående dag i dataserien.
 - (b) Kumulativ summa av dataserien.
 - (c) Kvadraten på värdena i dataserien.
10. De tre knapparna på denna rad ritar:
 - (a) Priset för det valda instrumentet.
 - (b) Volumen för det valda instrumentet.
 - (c) Omsättning för det valda instrumentet.
11. Plot-fönstret där allt ritas. Y-axeln är olika beroende på vilken dataserie som valts. T.ex. för volym är det antalet kontrakt, medan för antalet avbrutna ordrar är det antalet ordrar som visas på y-axeln. X-axeln visar datum.
12. I detta fönster visas några av de olika valen som har gjorts:
 - Aktör (user).
 - Instrument.
 - Dataserie.
 - Om vecka (V) eller månad (M) har valts på rad 6 (length window).
 - Vald färg.
 - Vald typ av konfidensintervall (bounds).
 - Vald typ av glidande fönster (roll type).
13. Här väljs vilken färg man vill plotta med.



Figur 8: Figur som visar hur plot-fönstret ser ut. De röda siffrorna används för att hänvisa till olika delar av figuren.

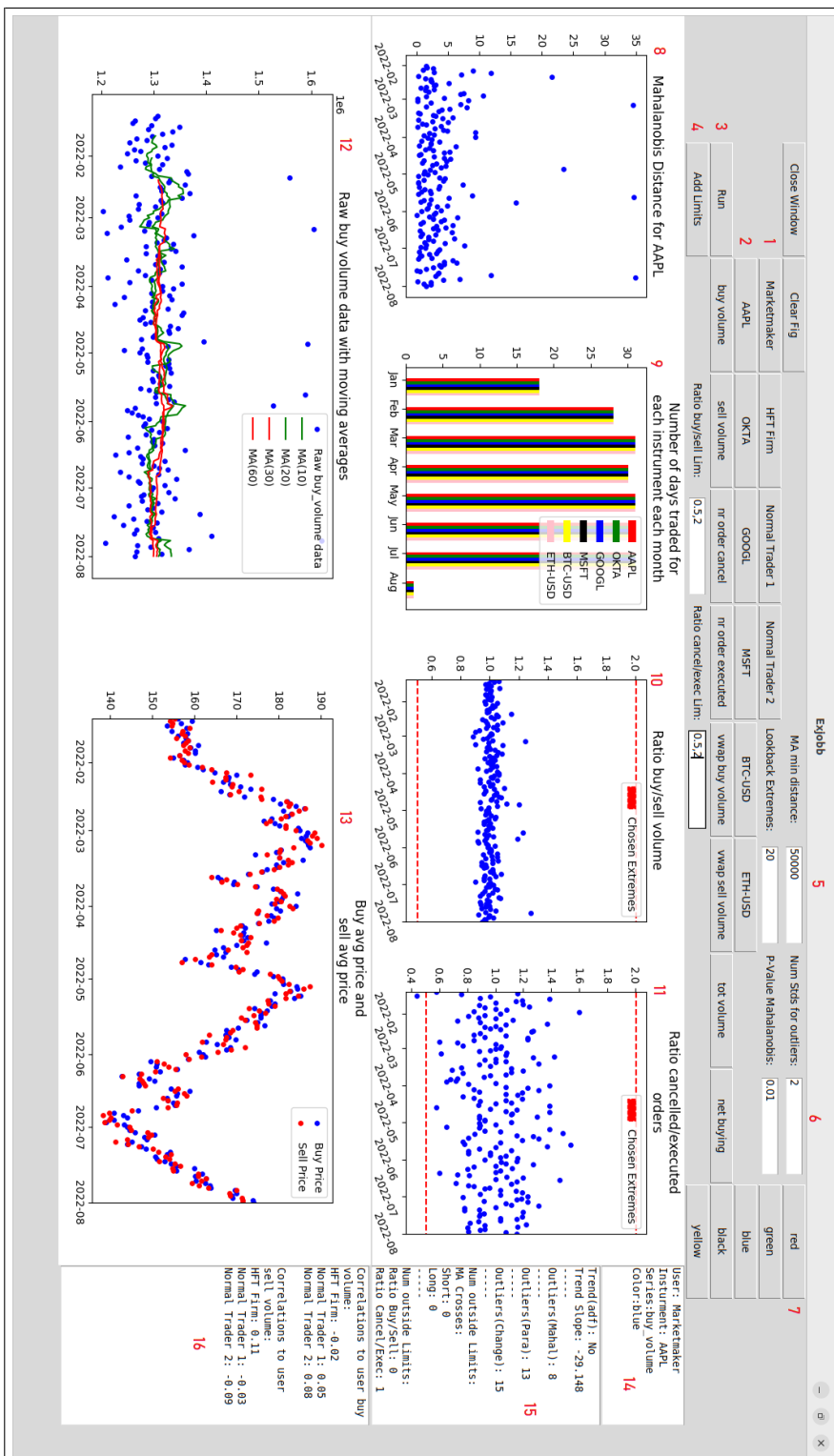
6.2 Test-GUI 1

Detta test-GUI fokuserar på en specifik aktör. Här väljs aktör, instrument samt dataserie som man vill analysera, t.ex köp-, sälj-, totalvolym eller antal exekverade ordrar. Sedan plottas Mahalanobisdistans, hur många instrument per månad, köp-/säljvolymkvot, kvot mellan cancelleerade och exekverade ordrar, glidande medelvärden, samt genomsnittligt köp- och säljpris i varsitt plot-fönster. I marginalen presenteras vald aktör och instrument samt dataserie. Det skrivs också ut om någon trend har hittats, gradienten på denna trend, antalet extremvärden (Mahalanobis), antalet extremvärden (parametriska), samt extremvärden på förändringen från dag till dag. Det skrivs också ut hur många medelvärdeskorsningar som uppmäts i fönstret som visar glidande medelvärden. I sista rutan i marginalen skrivs också ut det tidigare nämnda måttet för korrelation mellan köp- och säljvolym för vald aktör och övriga aktörer.

Här nedan finns fönstrets alla funktioner förklarade. Alla referenser till figurer i denna tabell syftar på siffror/figurer i tillhörande bild 9.

1. Aktör.
2. Instrument.
3. Dataserie.
4. På denna rad finns två fält där man kan lägga in max- och minvärde som sedan ritas in i figur 10 respektive 11. Antalet punkter utanför dessa linjer räknas och summan visas längst ner i fönster 15 (ratio buy/sell samt ratio cancel/exec).
5. Två fält:
 - "MA min distance" är minsta avståndet mellan två MA sedan förra korsningen för att det ska räknas som en ny korsning.
 - "lookback extremes" bestämmer hur många dagar som ska användas när man räknar ut konfidensintervallet till de tre olika extremvärdena i fönster 15.
6. Två fält:
 - "Num Stds for outliers" är antalet standardavvikelse som används vid uträkning av konfidensintervallen som används vid de tre olika extremvärdena i fönster 15.
 - "P-value Mahalanobis" bestämmer vilket p-värde som ska räknas som extremt för Mahalanobisdistansen. En punkt med ett värde under P-värdet innebär att punkten räknas som ett extremvärde.
7. Vald färg.
8. Figur för Mahalanobisdistans. På Y-axeln är punktens distans (enhetslös).

9. Figur som visar hur många dagar per månad varje instrument handlas.
10. Figur som visar kvot mellan köp- och säljvolym.
11. Figur som visar kvot mellan avbrutna och exekverade ordrar.
12. Figur som visar den valda dataserien (blå punkter) samt två par glidande medelvärden. Ett "kort" par med längderna 10 respektive 20, same ett "långt" par med längderna 30 respektive 60.
13. Figur som visar genomsnittligt köppris (blå) samt genomsnittligt säljpris (röd) varje dag.
14. Här väljs vilken färg man vill plotta figur 8, 10, 11 samt 12 med.
15. I detta fönster visas några av de olika valen som har gjorts:
 - Aktör (User).
 - Instrument.
 - Dataserie.
 - Vald plotfärg.
16. Datafönster. Fönstret visar, i ordning:
 - Om en trend hittats i figur 12.
 - Trendens lutning.
 - Mahalanobis extremvärden (för vald dataserie, visualiserad i figur 12).
 - Parametriska extremvärden (för vald dataserie, visualiserad i figur 12).
 - Extremvärden i förändringen per dag (för vald dataserie, visualiserad i figur 12).
 - Antalet korsningar de två paren glidande medelvärdena gör (minus de korsningar där avståndet mellan medelvärdena var mindre än valet gjort i 5, sedan föregående korsning).
 - Antalet punkter över+under de två gränserna valda på rad 4, visualiserade i figur 10 respektive 11.
 - Korrelation mellan vald aktörs köp- samt säljvolym till de andra aktörernas sälj- samt köpvolym.



Figur 9: Figur som visar hur det första test-fönstret ser ut. De röda siffrorna används för att hänvisa till olika delar av figuren.

6.3 Test-GUI 2

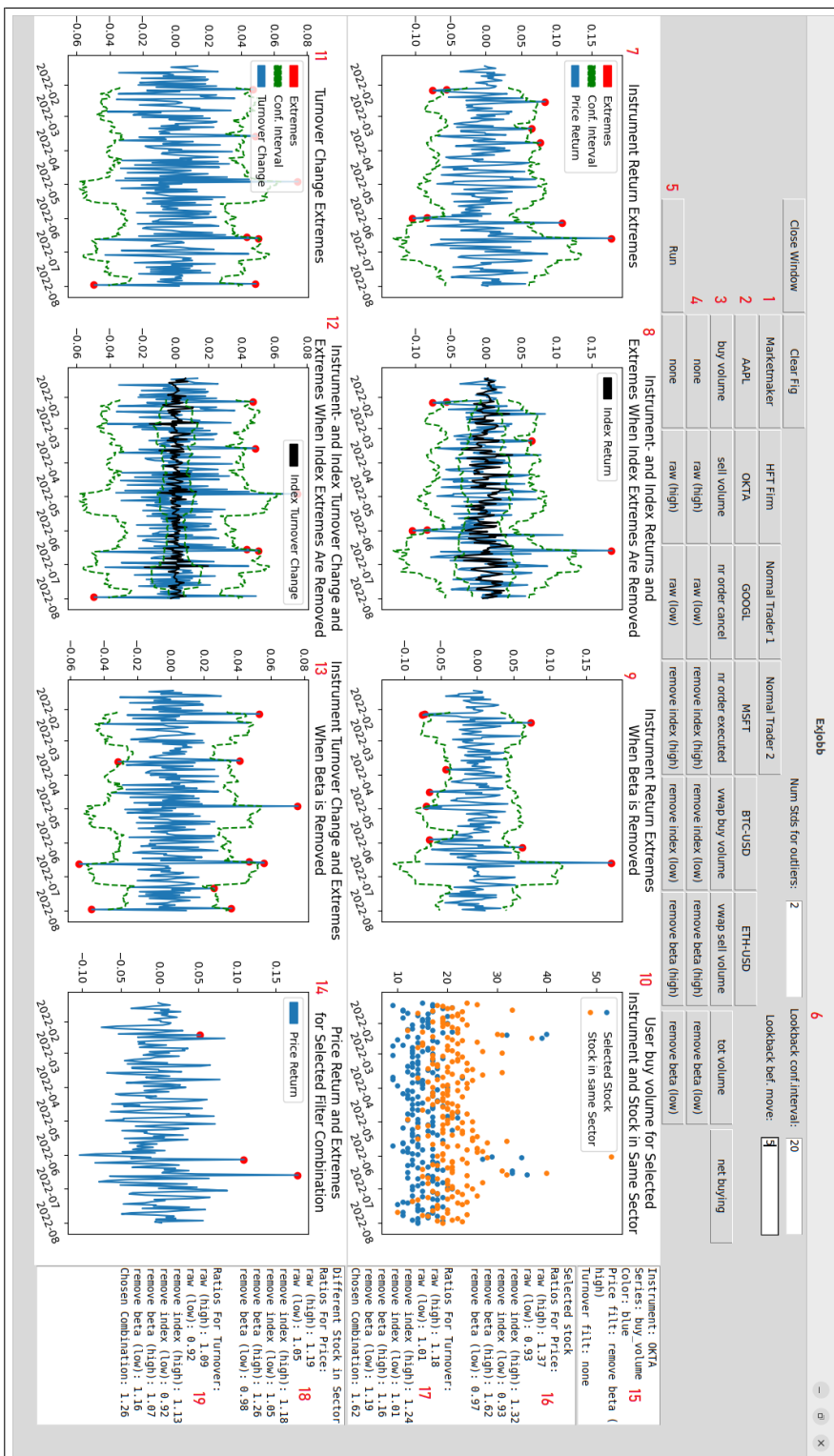
Precis som i Test-GUI 1 väljer man aktör, instrument och serie. Det som är speciellt här är att man väljer ett av sex val för avkastning, samt ett av sex val för omsättning. Dessa är olika filter för att visa extremvärden för det valda instrumentet. Det finns raw (high/low), remove index (high/low) samt remove beta (high/low). Dessa olika finns beskrivna ovan i ”5.2.10 Handel före extrema marknadsrörelser”-avsnittet. Användaren väljer ett mått från varje rad (alltså ett för avkastning och ett för omsättning). ”High” betyder att det är extremvärden som går ovanför medelvärdet + x standardavvikelse, medan low betyder dagar då värdet går under medelvärdet - x standardavvikelse. De två valda filtrena används för att hitta dagar som stämmer in på båda dessa kriterier. I fönstret finns också de olika plottarna som visar vilka dagar som hade extrem avkastning eller omsättning. I högermarginalen visas återigen vilken aktör, instrument samt serie som är vald. Det står också vilket avkastnings- samt omsättnings filter (på extremvärdena) användaren har valt. Under det finns det utskrivet kvoten mellan den valda seriens medelvärde de n dagar före varje extremvärde för det specifika filtret, och den genomsnittliga siffran för samma serie över hela tidsperioden. Detta finns också mer ingående förklarat längre upp. Nedanför detta finns ”Different Stock in Sector”. Detta är samma metod fast där kvoten är uträknad för en annan akties serie, dock med samma dagar som indikeras på skärmen.

Här nedan finns fönstrets alla funktioner förklarade. Alla referenser till figurer i denna tabell syftar på siffror/figurer i tillhörande bild 10.

1. Aktör.
2. Instrument.
3. Dataserie.
4. Filtertyp för prisavkastning.
5. Filtertyp för log-omsättningsförändring.
6. Tre olika fält:
 - Antalet standardavvikelse för uträkning av extremvärden.
 - Längden på det glidande fönstret för konfidensintervallet.
 - Hur många dagar innan extremvärden som används för att räkna ut om aktören köpt mycket innan extremvärden.
7. Figur som visar instrumentets avkastning (blå), konfidensintervall (grön streckad), samt vilka dagar avkastningen är större/mindre än konfidensintervallet (röd).
8. Figur som visar instrumentets avkastning (blå), konfidensintervall (grön streckad), indexets avkastning (svart), indexets konfidensintervall (grön streckad), samt vilka dagar avkastningen är större/mindre än konfidensintervallet när man tar bort de dagar som även indexet är utanför sitt konfidensintervall (röd).

9. Figur som visar instrumentets avkastning när man har tagit bort indexets påverkan med hjälp av beta-värdet för instrumentet (blå), konfidensintervall (grön streckad), samt vilka dagar denna justerade avkastning är större/mindre än konfidensintervallet (röd).
10. Figur som visar dataserien för aktören för det valda instrumentet samt för det andra instrumentet i samma sektor.
11. Figur som visar instrumentets dagliga förändring i log-omsättning (blå), konfidensintervall (grön streckad), samt vilka dagar denna förändring är större/mindre än konfidensintervallet (röd).
12. Figur som visar instrumentets dagliga förändring i log-omsättning (blå), konfidensintervall (grön streckad), indexets dagliga förändring i log-omsättning (svart), indexets konfidensintervall (grön streckad), samt vilka dagar förändringen är större/mindre än konfidensintervallet när man tar bort de dagar som även indexet är utanför sitt konfidensintervall (röd).
13. Figur som visar instrumentets dagliga förändring i log-omsättning när man har tagit bort indexets påverkan med hjälp av beta-värdet för instrumentet (blå), konfidensintervall (grön streckad), samt vilka dagar denna justerade förändring är större/mindre än konfidensintervallet (röd).
14. Figur som visar instrumentets avkastning (blå) samt en liknande röd prick som i de andra tabellerna, de dagar som den valda filterkombinationen ger utslag på samma dagar. I denna specifika bild har hög beta-justerad avkastning valts (inget omsättningsfilter).
15. I detta fönster visas några av de olika valen som har gjorts:
 - Aktör (User).
 - Instrument.
 - Dataserie (col).
 - Vald plotfärg.
 - Pris-filtertyp.
 - Omsättning-filtertyp.
16. Här visas kvoten av medelvärdet av aktörens tidsserie under de dagar som ledde upp till en stor avkastningsförändring (det exakta filtret står utskrivet), jämfört med snittet över hela tidsserien.
17. Här visas kvoten av medelvärdet av aktörens tidsserie under de dagar som ledde upp till en stor förändring i log-omsättning (det exakta filtret står utskrivet), jämfört med snittet över hela tidsserien.
18. Här visas kvoten av medelvärdet av aktörens tidsserie för en annan aktie i samma sektor som den valda, under de dagar som ledde upp till en stor avkastningsförändring (det exakta filtret står utskriver), jämfört med snittet över hela tidsserien.

19. Här visas kvoten av medelvärdet av aktörens tidsserie för en annan aktie i samma sektor som den valda, under de dagar som ledde upp till en stor förändring i log-omsättning (det exakta filtret står utskriver), jämfört med snittet över hela tidsserien.



Figur 10: Figur som visar hur det andra test-fönstret ser ut. De röda siffrorna används för att hänvisa till olika delar av figuren.

6.4 Test-GUI 3

Test-GUI 3 är ett fönster som visar data för alla olika aktörer (I detta fall fyra olika). Här plottas den valda serien för det valda instrumentet för de olika aktörerna i varsitt fönster. Det plottas även vertikala linjer som överensstämmer med det extremvärdesfilter man har valt upp i menyn. Idén med detta är att man i Test-GUI 2 hittar en kombination av avkastning- och omsättningsfilter man vill titta närmare på, sedan går man in i Test-GUI 3 och tittar på hur de olika individerna handlade runt dessa dagar. Sedan kan man gå ännu djupare i Test-GUI 1.

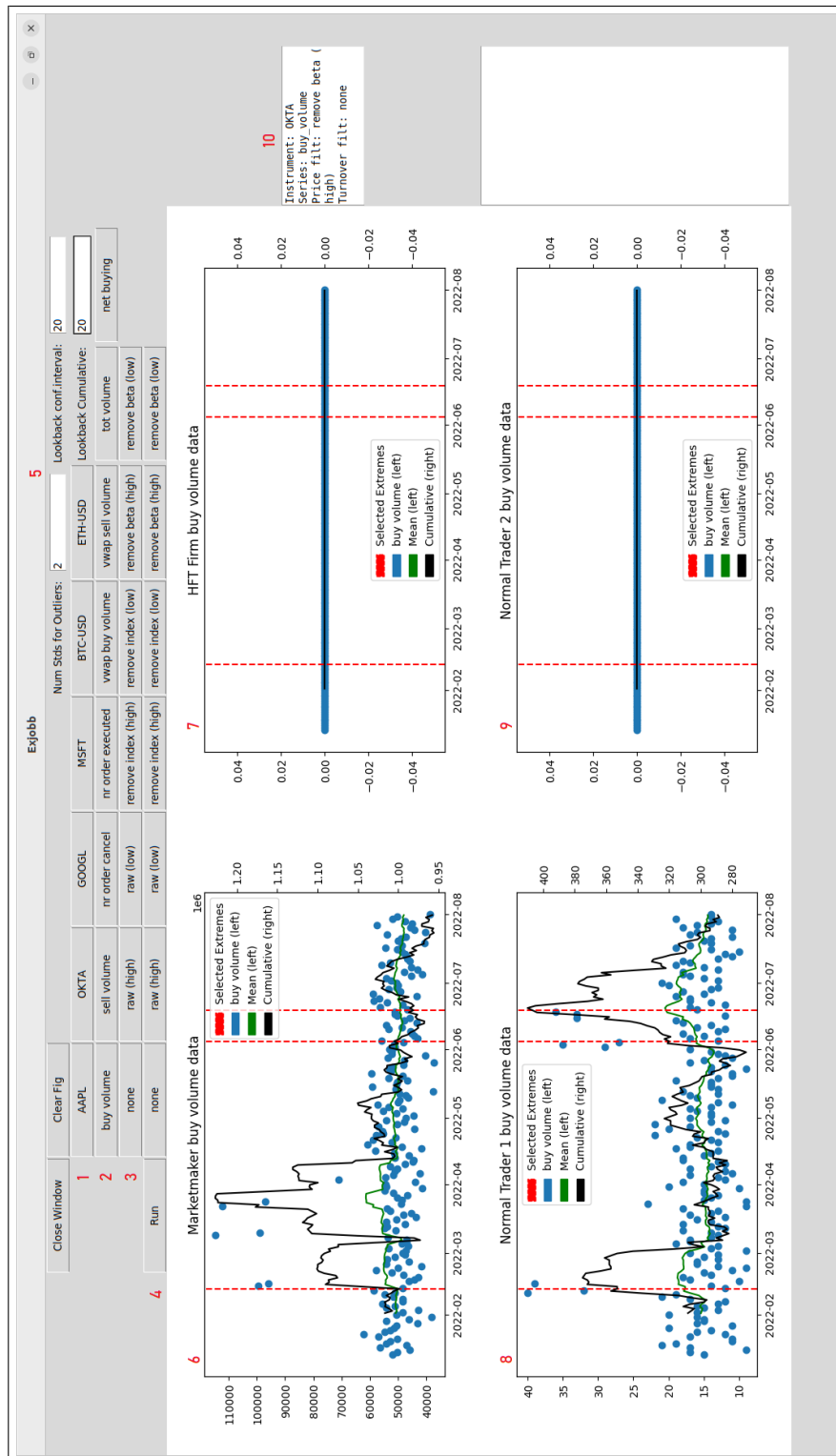
Här nedan finns fönstrets alla funktioner förklarade. Alla referenser till figurer i denna tabell syftar på siffror/figurer i tillhörande bild 11.

1. Instrument.
2. Dataserie.
3. Filtertyp för prisavkastning.
4. Filtertyp för log-omsättningsförändring.
5. Tre olika fält:
 - Antalet standardavvikelser för de glidande konfidensintervallen som används för att hitta extremvärdena på samma sätt som i Test-GUI 2. Extremvärdena ritas som röda vertikala streck i figur 6-9.
 - Längden på de glidande konfidensintervallen.
 - Hur många dagar som används när man räknar ut den glidande kumulativa summan i figurerna.
6. Figur som visar vald dataserie (blå, axel till vänster) för aktör market maker samt extremvärdena som stämmer in på båda de valda filtrena (röd). Dessa är precis samma dagar som de röda punkterna i Test-GUI 2. Glidande medelvärdet för den valda dataserien (grön, axel till vänster), samt kumulativ summa för den valda dataserien (svart, axel till höger).
7. Figur som visar vald dataserie (blå) för aktör HFT-Firma samt extremvärdena som stämmer in på båda de valda filtrena (röd). Dessa är precis samma dagar som de röda punkterna i Test-GUI 2. Glidande medelvärdet för den valda dataserien (grön), samt kumulativ summa för den valda dataserien (svart).
8. Figur som visar vald dataserie (blå) för Vanlig Aktör 1 (Normal Trader 1) samt extremvärdena som stämmer in på båda de valda filtrena (röd). Dessa är precis samma dagar som de röda punkterna i Test-GUI 2. Glidande medelvärdet för den valda dataserien (grön), samt kumulativ summa för den valda dataserien (svart).
9. Figur som visar vald dataserie (blå) för Vanlig Aktör 2 (Normal Trader 2) samt extremvärdena som stämmer in på båda de valda filtrena (röd). Dessa är precis

samma dagar som de röda punkterna i Test-GUI 2. Glidande medelvärdet för den valda dataserien (grön), samt kumulativ summa för den valda dataserien (svart).

10. I detta fönster visas några av de olika valen som har gjorts:

- Instrument.
- Dataserie.
- Pris-filtertyp.
- Omsättning-filtertyp.

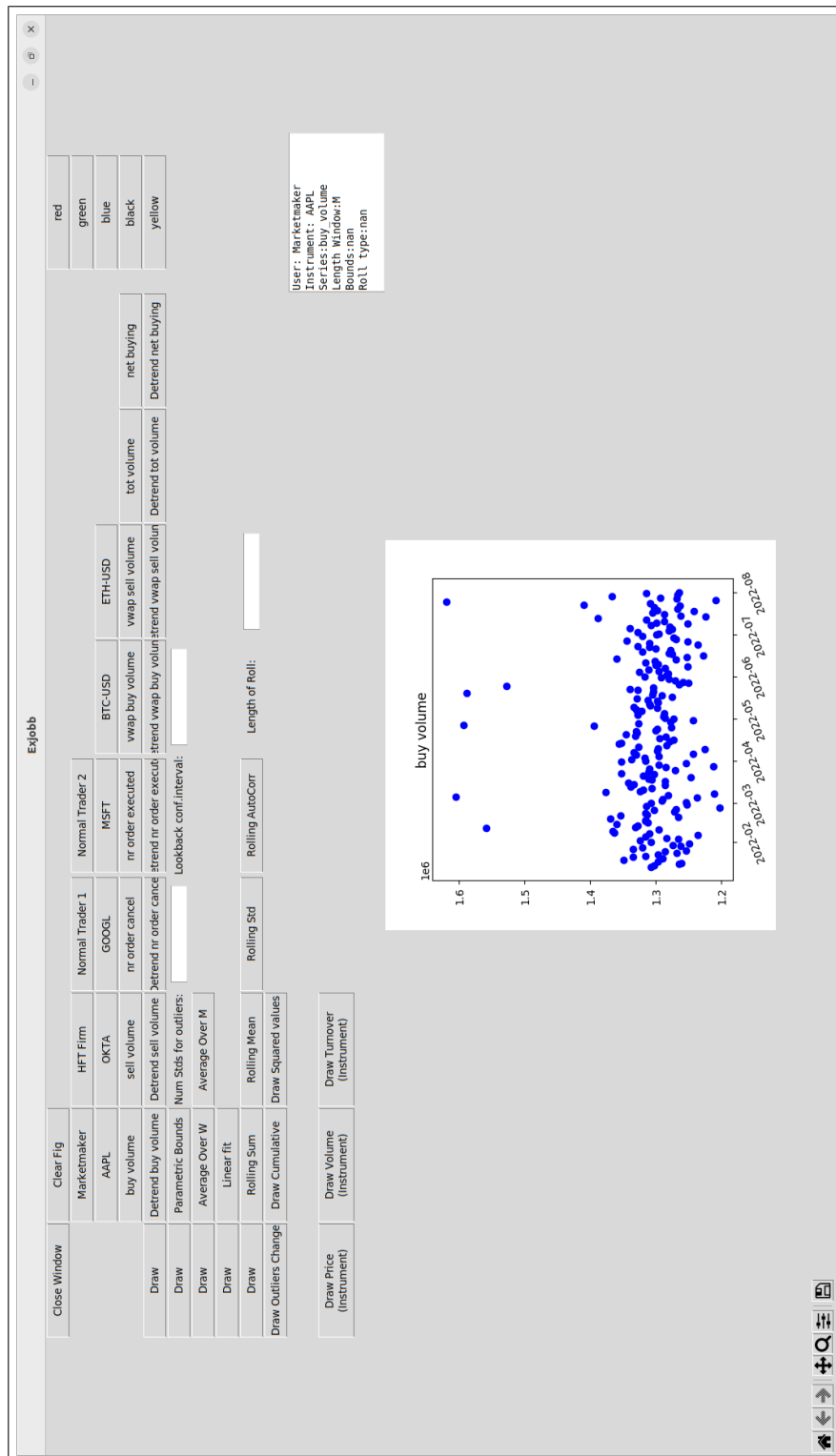


Figur 11: Figur som visar hur det tredje test-fönstret ser ut. De röda siffrorna används för att hänvisa till olika delar av figuren.

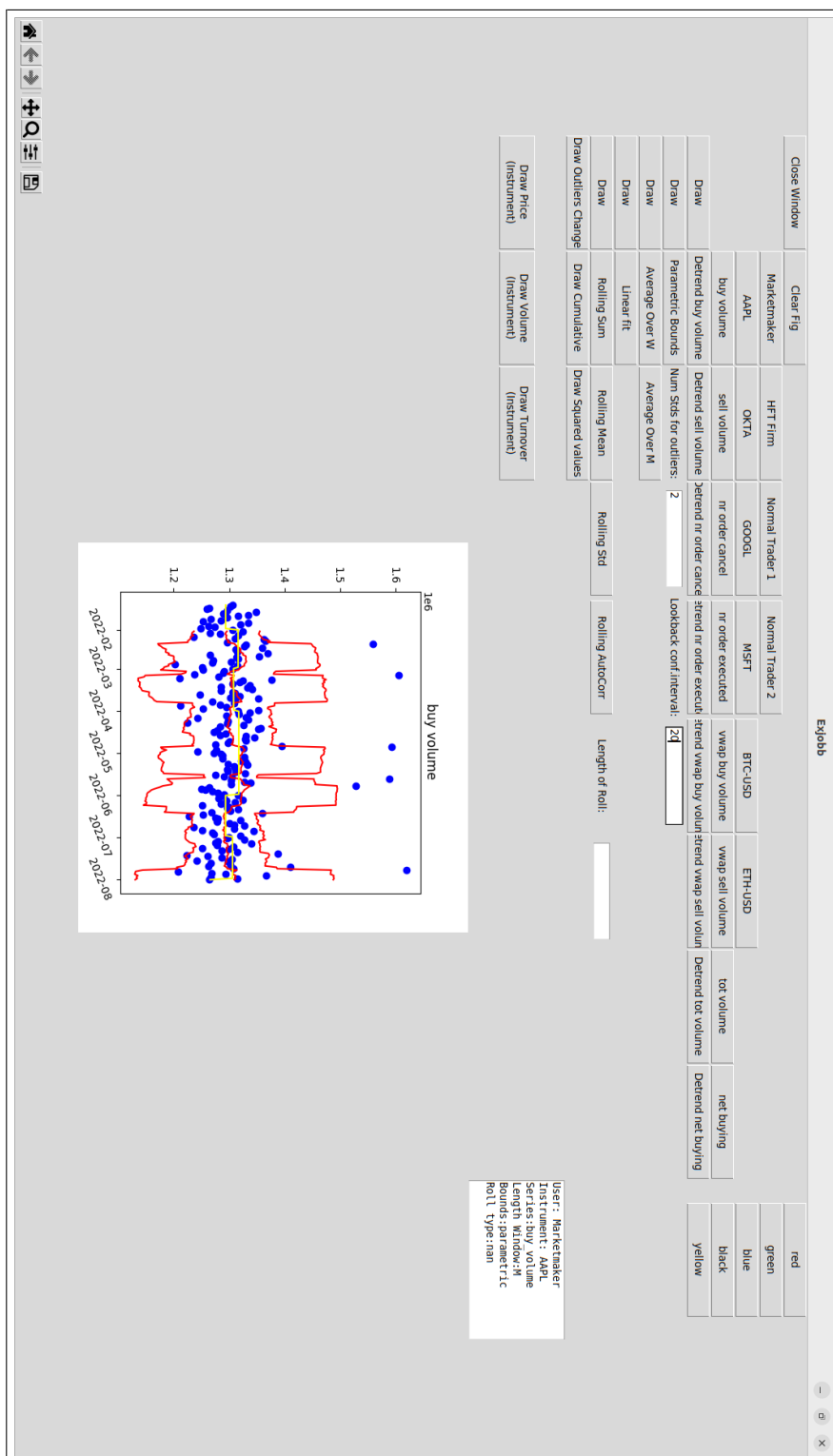
7 Resultat

I resultatet visas först tre bilder från plot-GUIt. I plot-fönstret görs ingen analys utan det är till för att kunna titta närmare på en specifik aktörs dataserie. Nästkommande bilder visar sedan hur olika scenarion för varje aktör detekteras av test-GUI 1, 2 och 3.

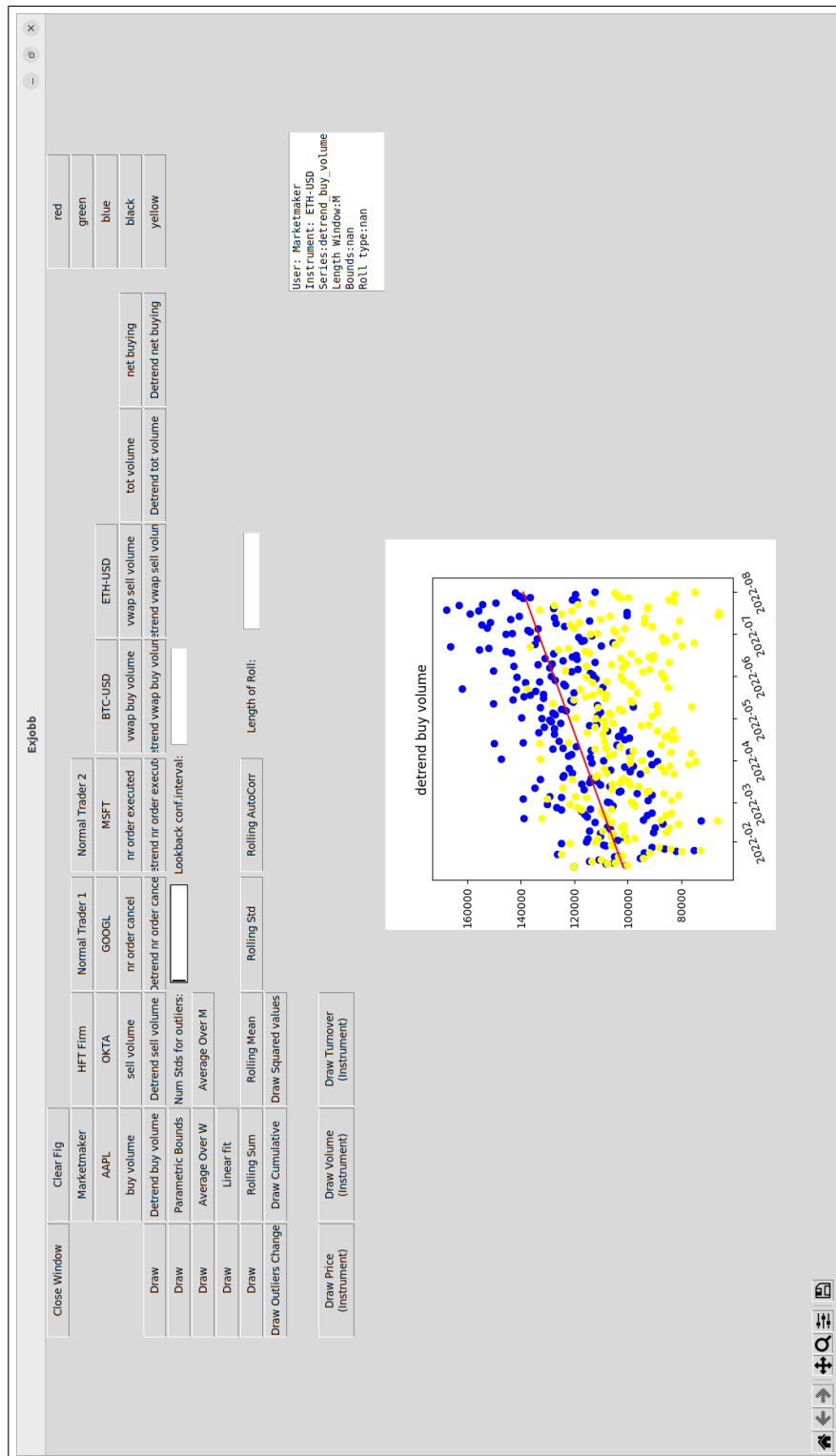
7.1 Plot-GUI



Figur 12: Bild som visar standard-plot-fönstret där man kan välja att plotta olika data. I bilden har endast köpvolum plottats i blått.

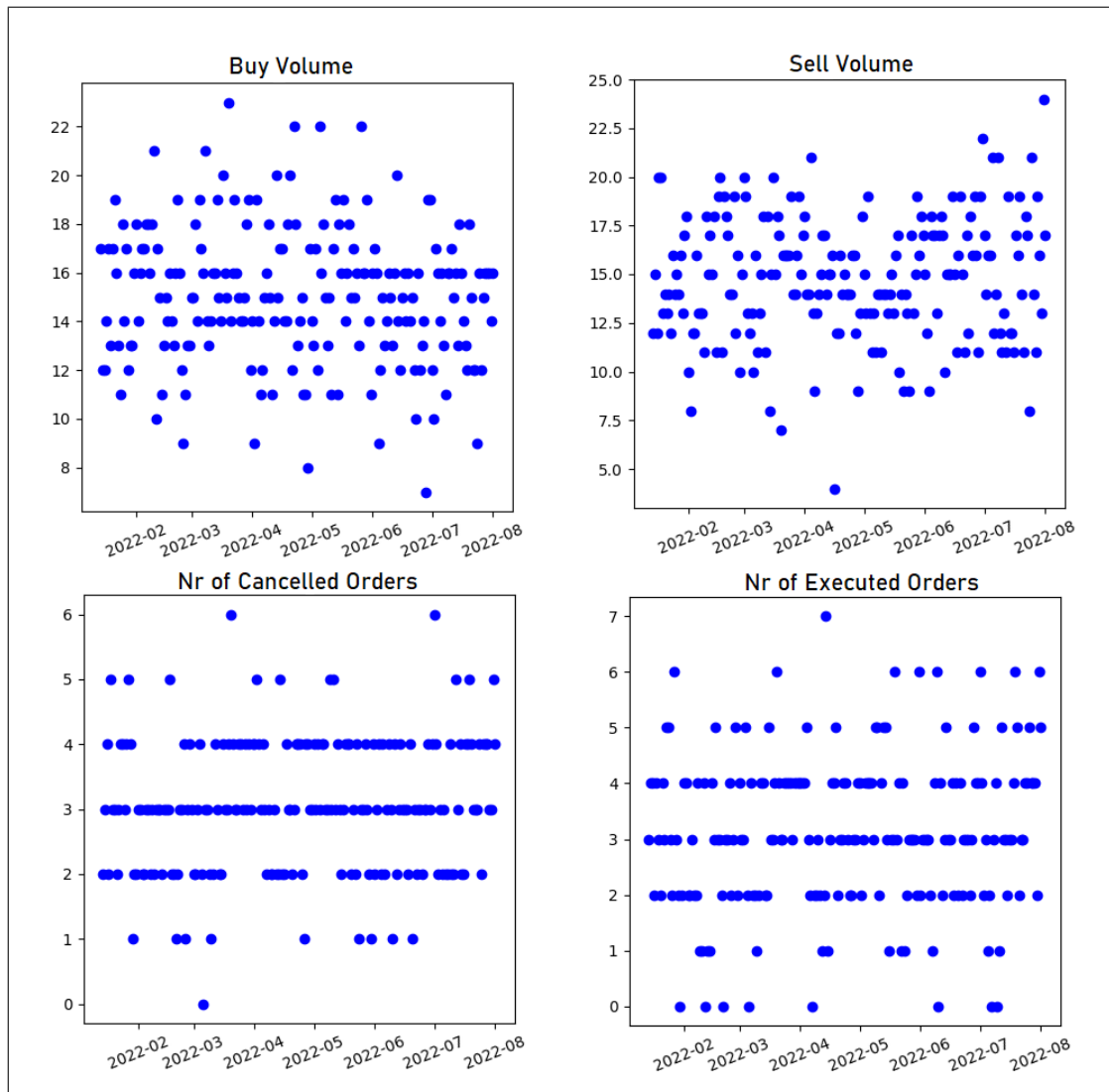


Figur 13: Exempelbild som visar hur plot-fönstret plottar köpvoly (blå), glidande medelvärde och konfidensintervall (rött), samt snittvolymen för varje månad (gult). Längden för det glidande medelvärdet och konfidensintervallet är satt till 20, samt 2 standardavvikelser.

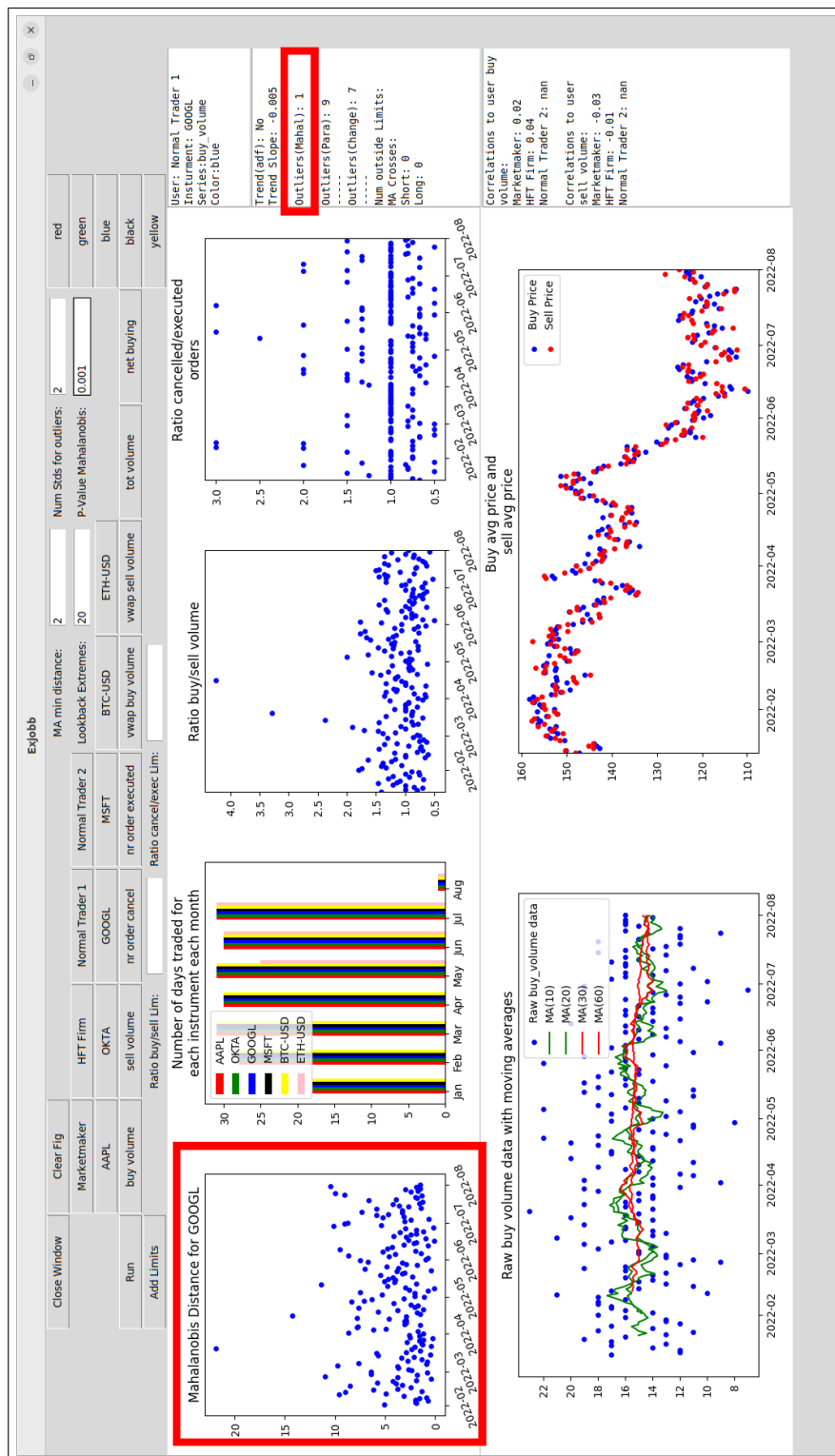


Figur 14: Exempelbild som visar hur plot-fönstret plottar köpvoly (blå), trenden för köpvolyen (röd), samt köpvolyen om man tar bort trenden (gul).

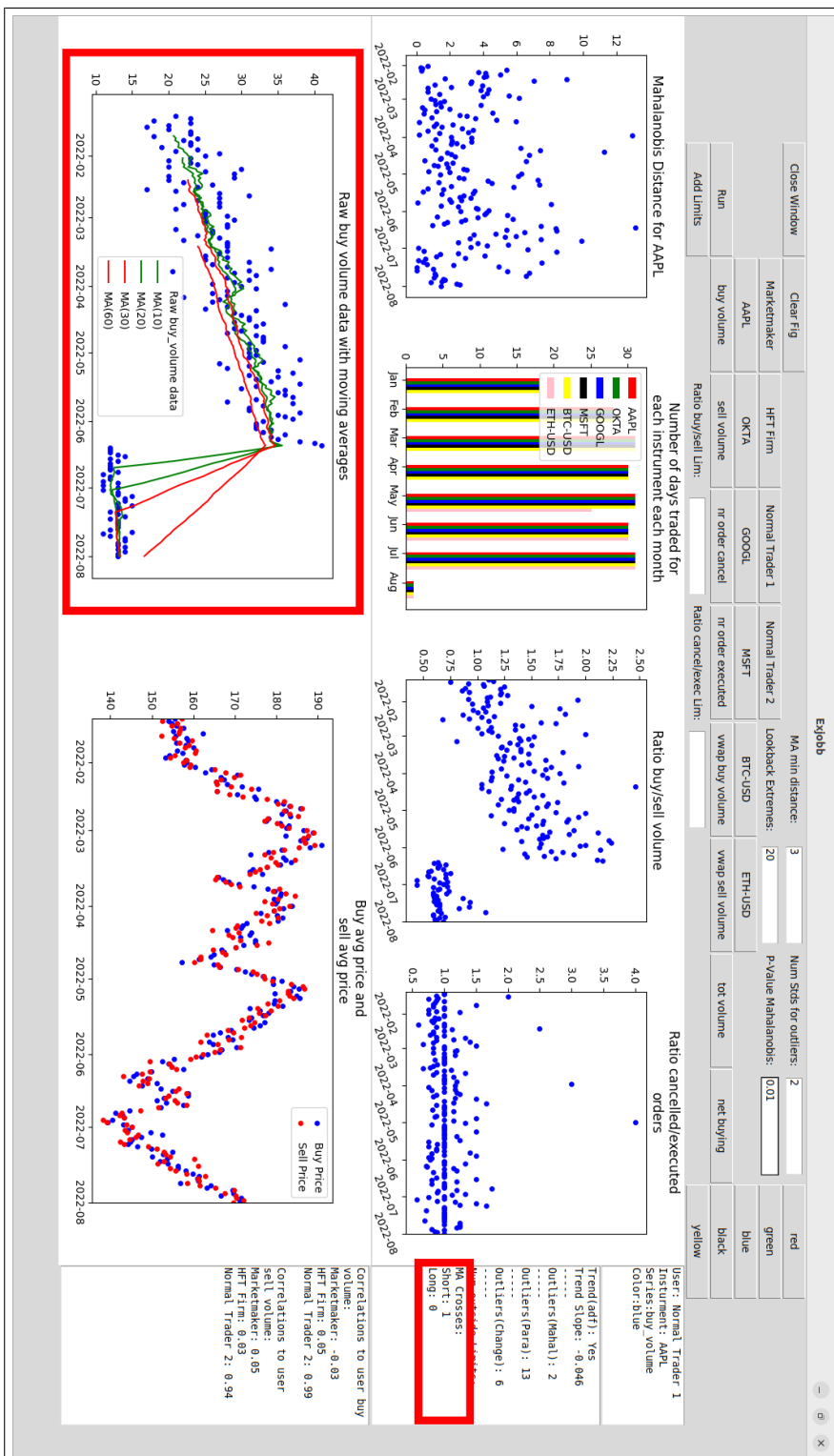
7.2 Vanlig Aktör 1



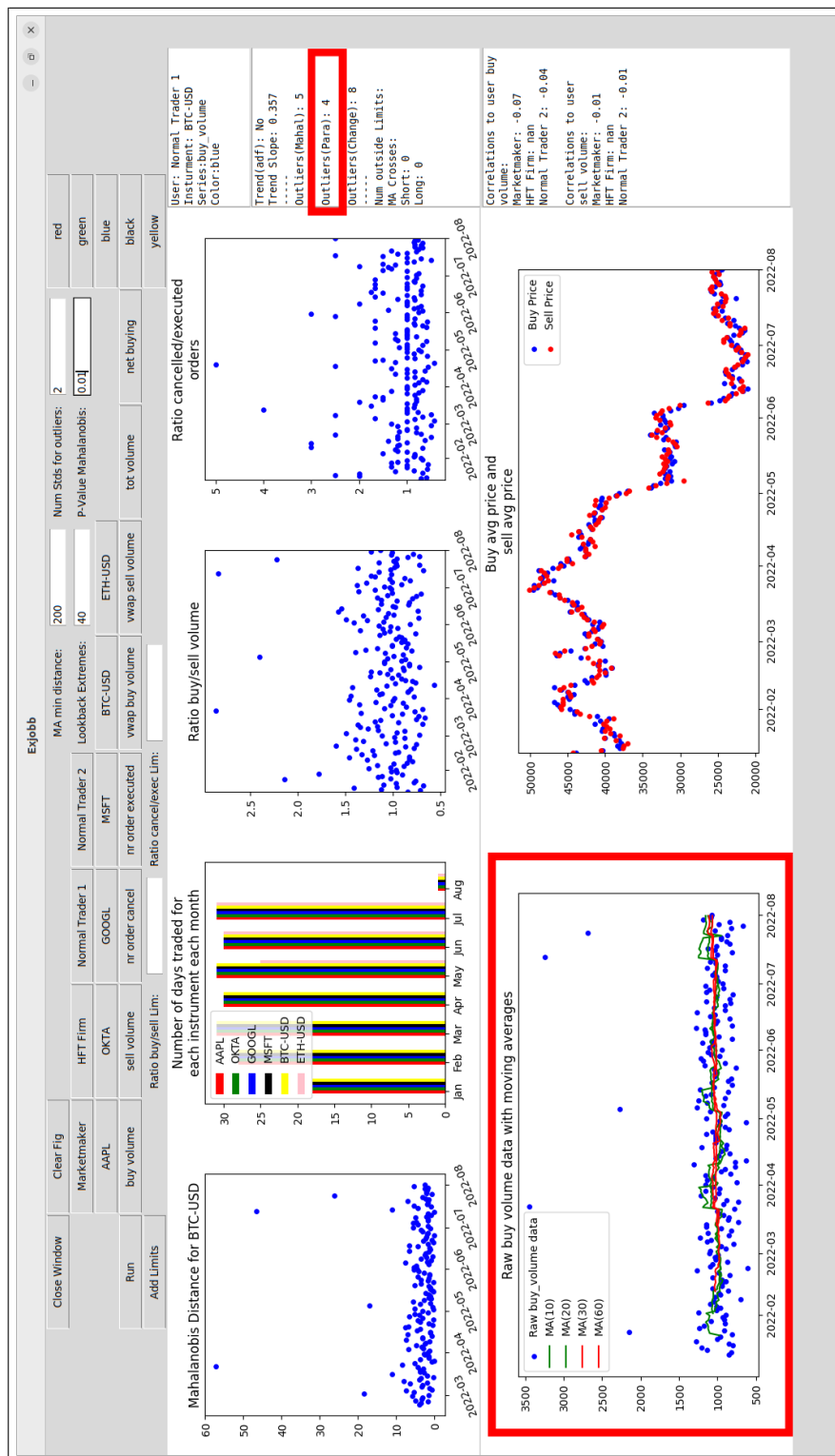
Figur 15: Exempel som visar de fyra dataserierna som används för att räkna ut Mahalanobisdistans. I dessa bilder syns inga extremvärden, men om man räknar ut mahalanobisdistansen ser man i figur 16 nedan att det finns extremvärden som endast dyker upp om man tittar på alla fyra dimensioner samtidigt. X-axeln visar datum för varje punkt i serierna.



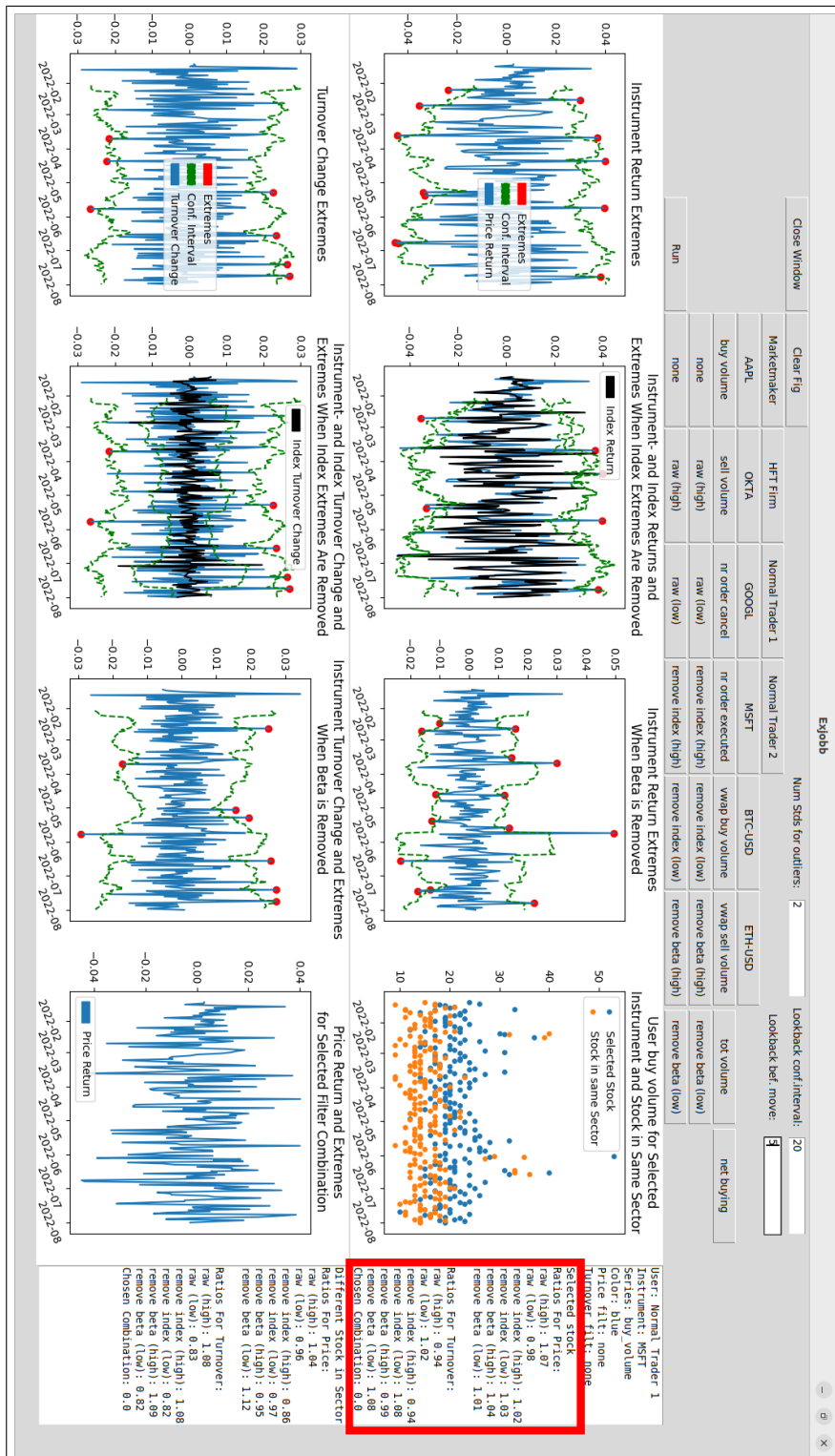
Figur 16: I första plotten syns ett Mahalanobi-extremvärde som inte syns i rådatan i figur 15 ovan. Om P-värdet sätts till 0.001 är detta det enda extremvärdet som hittas.



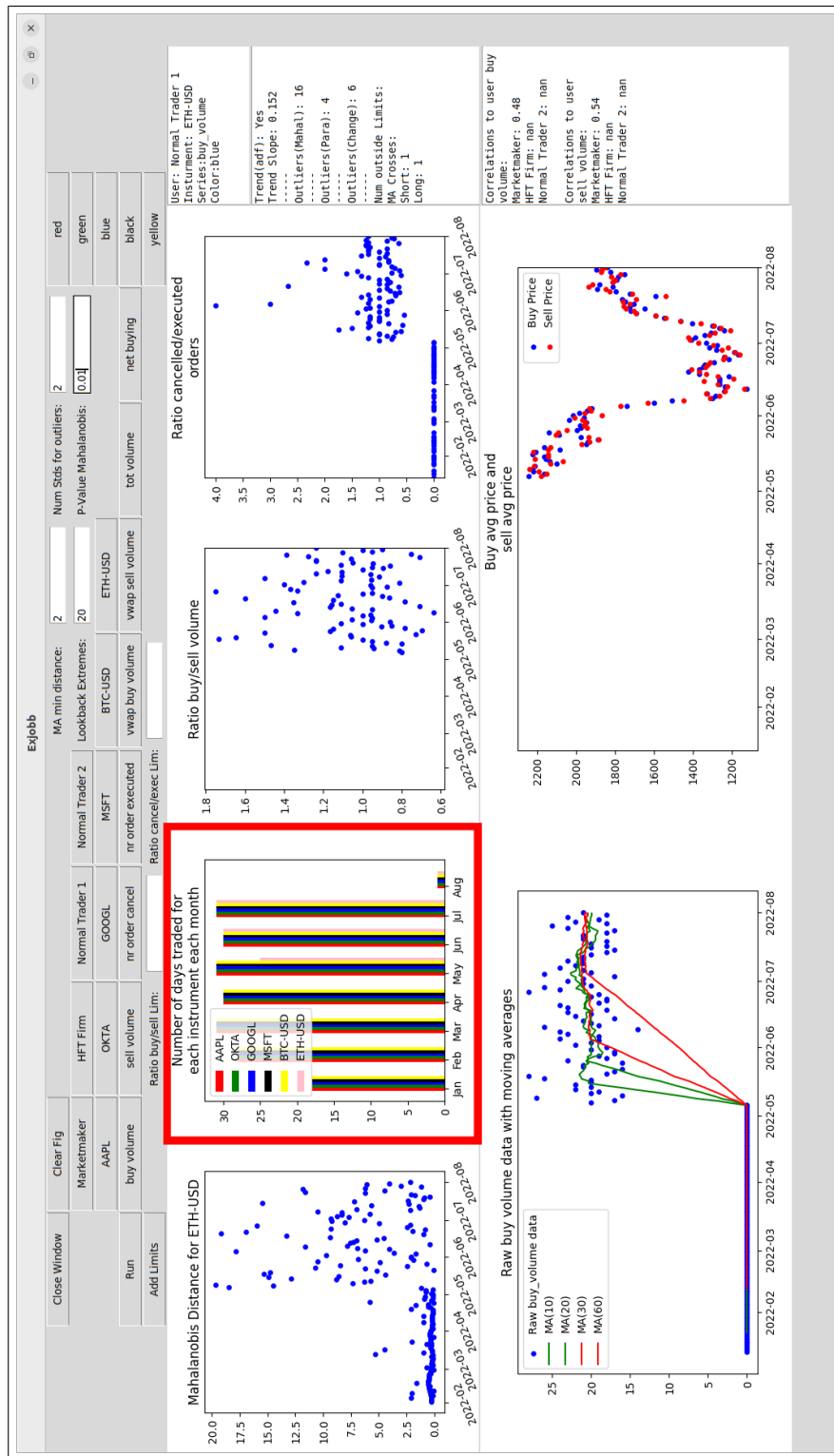
Figur 17: Visar hur medelvärdeskorsningar fungerar. I bilden nere till vänster syns tydliga korsningar. I högermarginalen syns att programmet hittar en sådan korsning (givet att filter-distansen är 3).



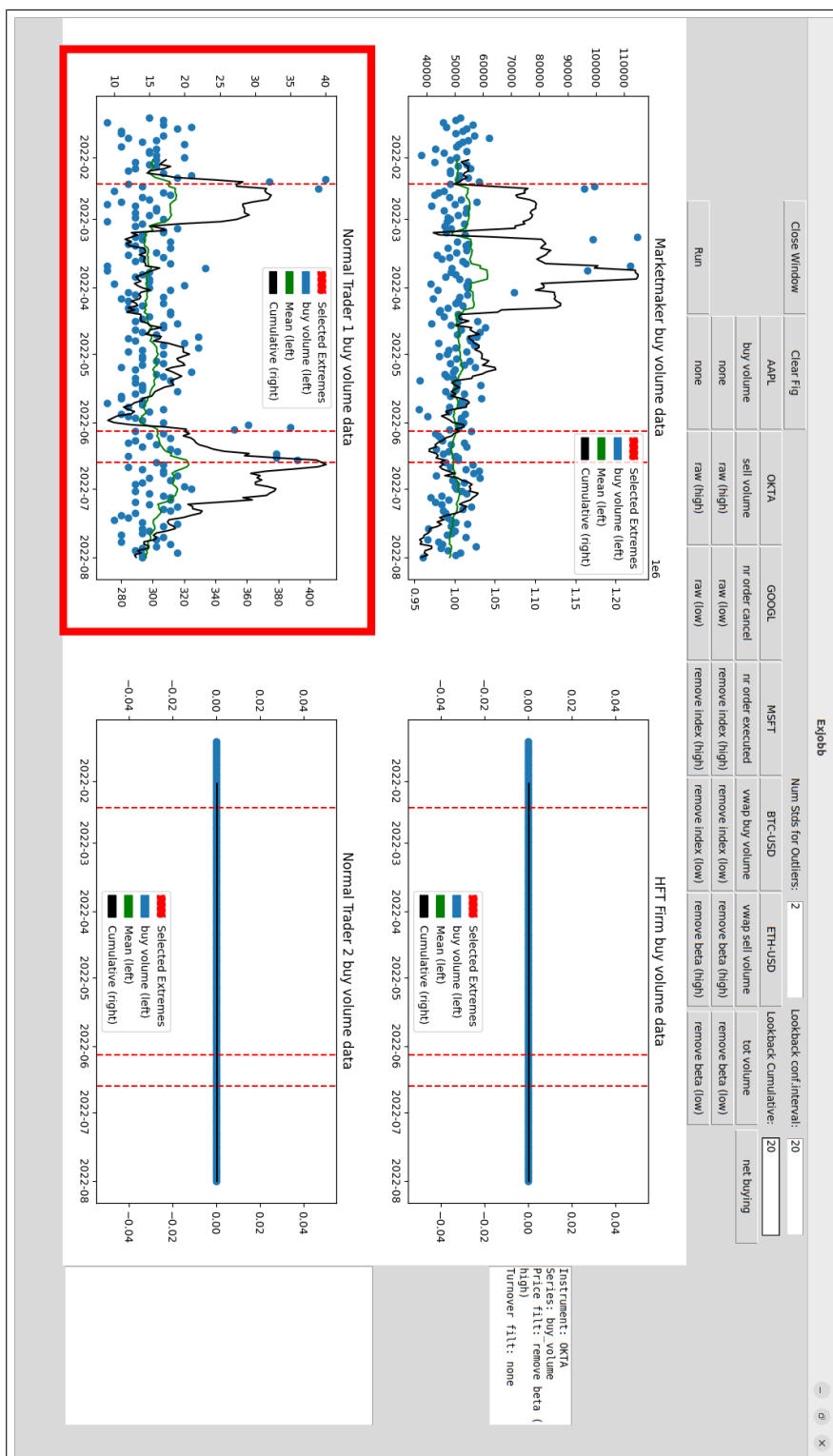
Figur 18: Visar några extremvärden i BTC-USD köpvolymsdata för Vanlig Aktör 1 när man sätter antalet standardavvikelser till 2 och antalet dagar den tittar tillbaka på till 40. Anledningen till att det till synes är 5 extremvärden men bara 4 detekteras är just att det glidande fönstret är 40 dagar långt. Därmed detekteras inga extremvärden innan dag 40.



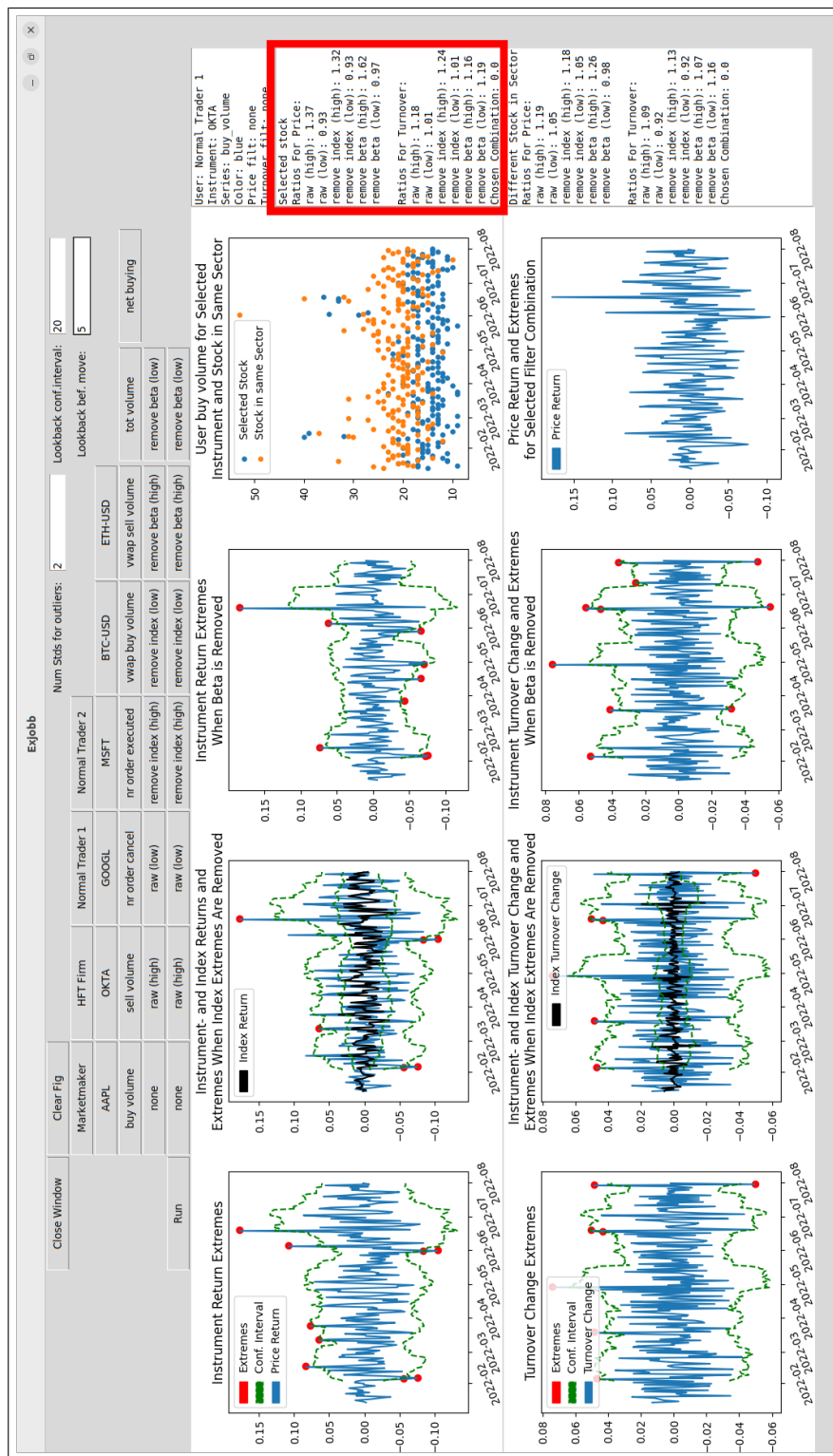
Figur 19: Figur som visar att inget konstigt upptäcks för hur aktören köper instrumentet i förhållande till dagar med stora prisförändringar eller omsättningsförändringar (alla siffror är runt- eller under 1).



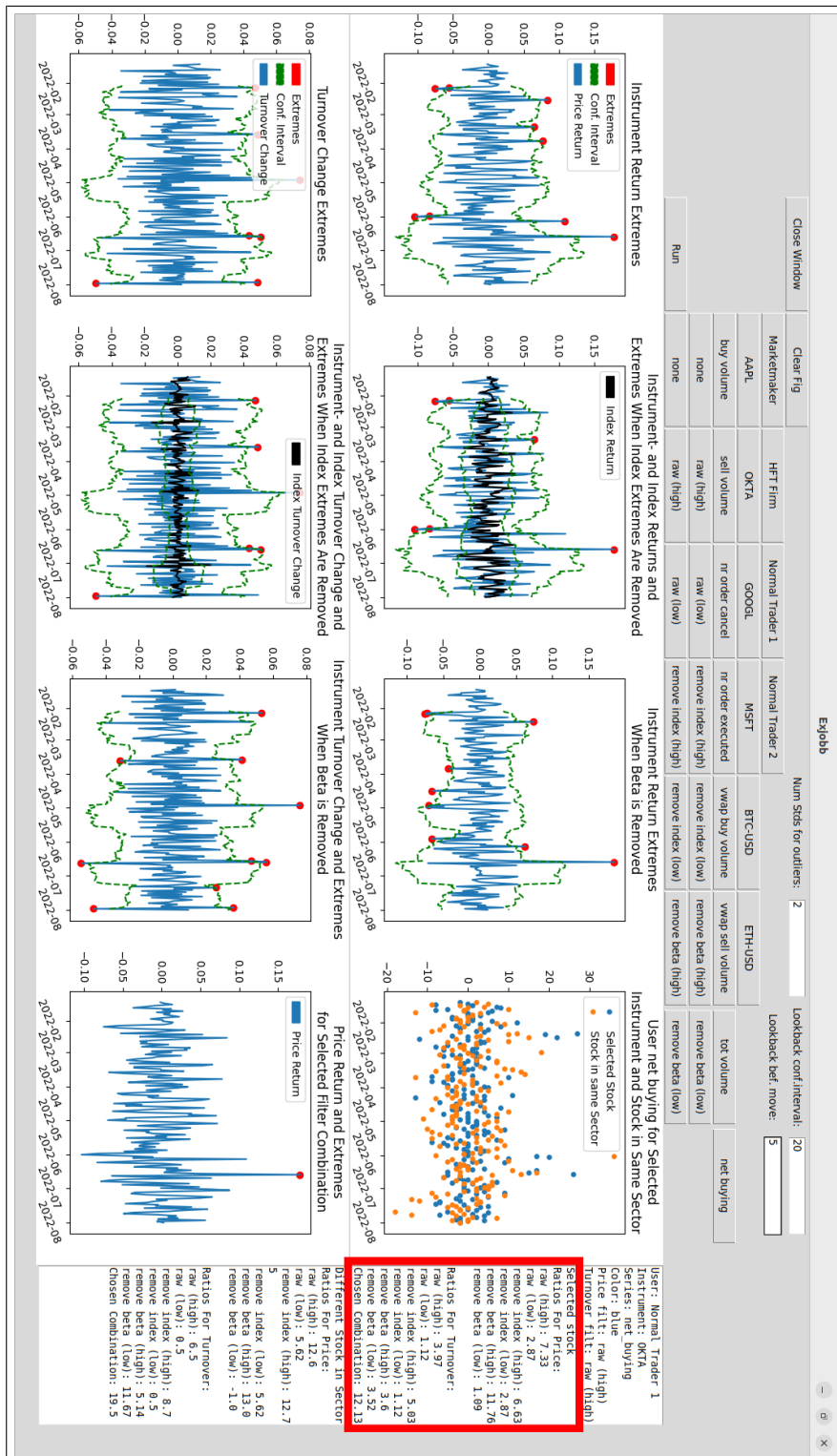
Figur 20: Figur som visar hur det ser ut när ett nytt instrument introduceras. Detta syns tydligast i figur 2, där man ser att den rosa stapeln introduceras efter ungefär halva tiden. Det syns dock även i de andra figurerna.



Figur 21: Figur som visar hur de olika aktörernas köpvolymer ser ut i aktien OKTA. Man ser att varken HFT-firman eller VA2 har köpt något alls, medan marketmakern och VA1 har köpt ofta. De röda linjerna visar de dagarna som filtreras fram. I detta fall är filtret inställt på dagar då avkastningen är hög om man tar bort indexberoendet (med hjälp av beta). Som kan ses så handlar VA1 misstänkt mycket under de dagar som är precis innan dessa stora prisförändringar.

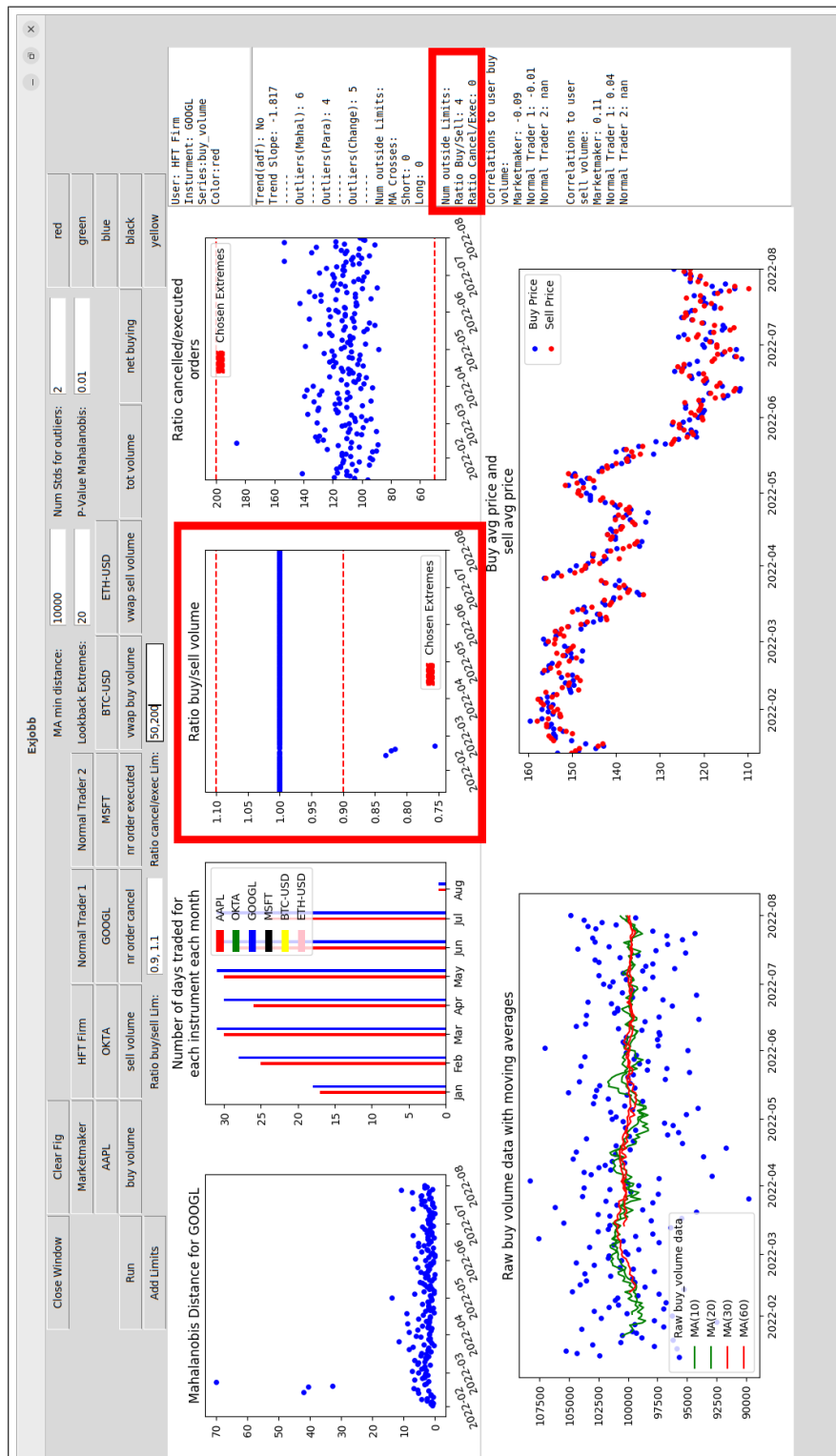


Figur 22: Efter figur 21 blir man nyfiken på om det finns några faktiska konstigheter med VA1 och hur denne handlar i OKTA. Tittar man i denna figur så ser man (i högermarginalen) att VA1 handlar förhållandevis mycket både före "raw (high)" (1.37x vanlig köpvolum), 1.32x före de dagar då det är hög avkastning men man har subtraherat de dagar då indexet också rör sig mycket, samt hela 1.62x mer under de dagar som leder upp till hög avkastning då man har räknat bort beroendet till index (beta).



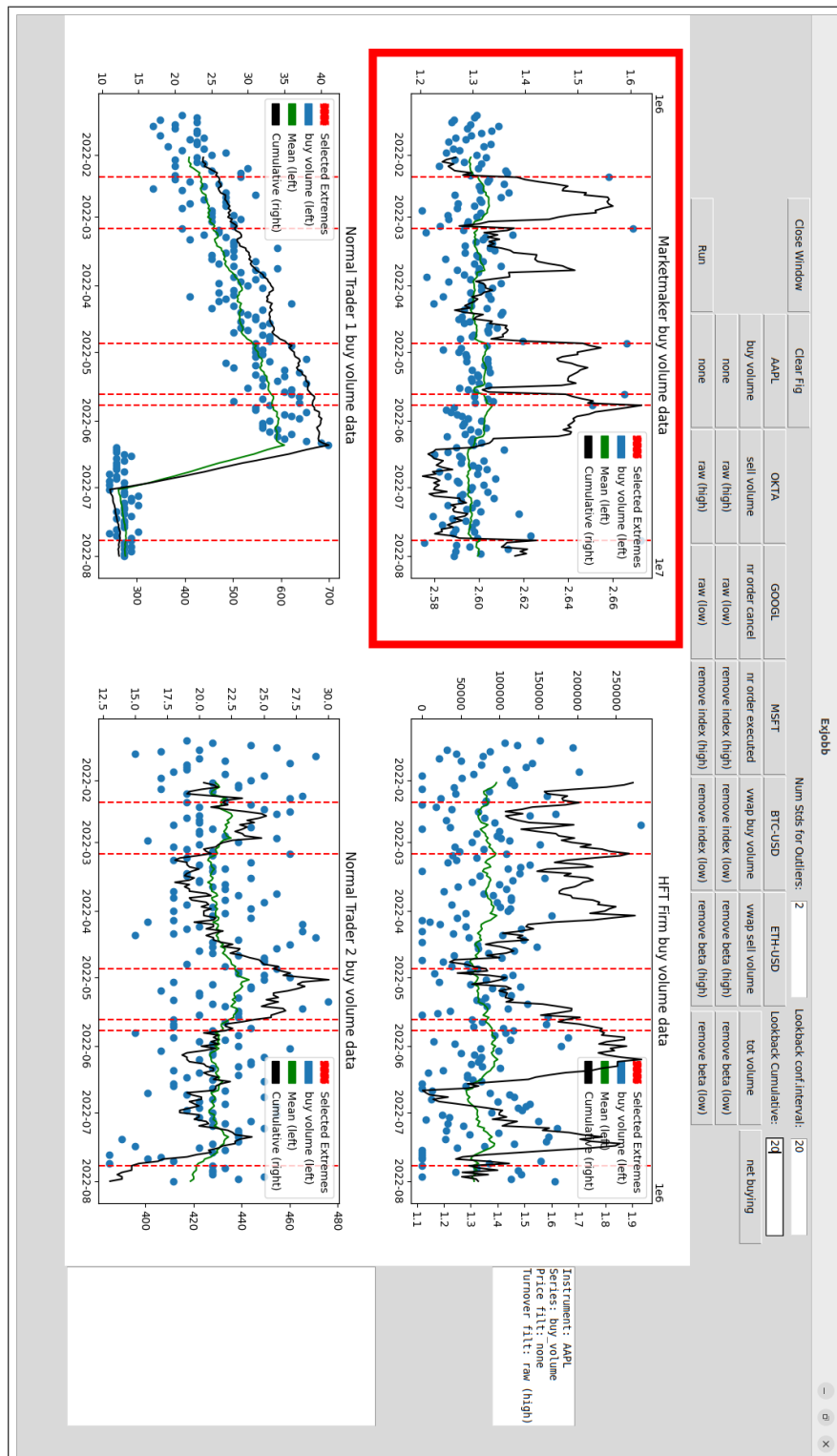
Figur 23: Denna figur visar samma sak som figur 22 ovan, men istället för köpvoly m så visas nettoköpvoly m. Här syns att nettoköpvoly men är mycket högre de dagar innan aktiepriset går upp mycket, för alla tre "high"-prisfiltertyper (precis som i figur 22) ovan.

7.3 HFT-Firma

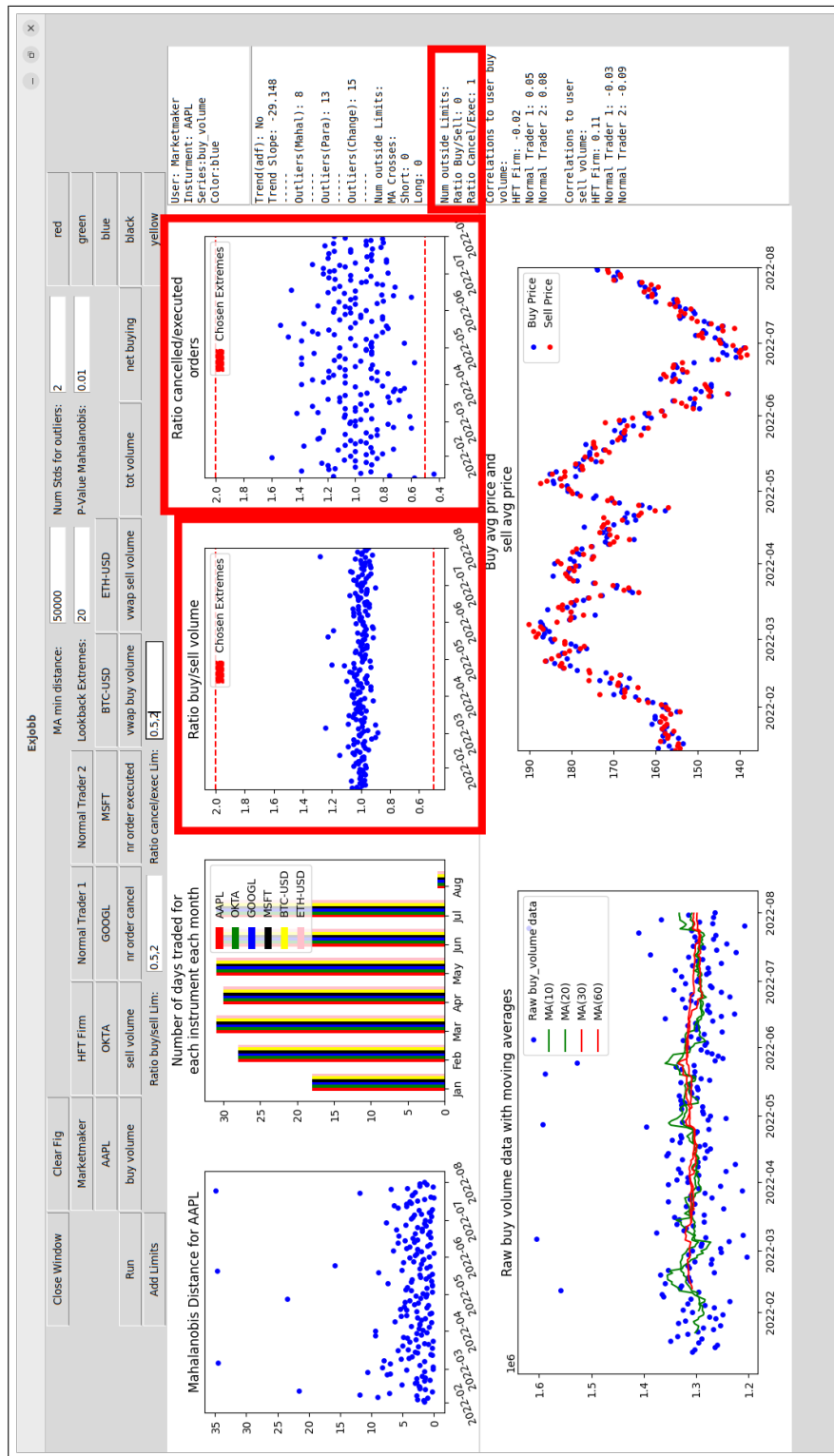


Figur 24: I plotten som visar förhållandet mellan köp- och säljvolym syns några extremvärden. Detta är förvånande då en HFTF-firma ofta bara håller sina positioner i sekunder. Att de skulle ha en skillnad i köp-och säljvolym en dag betyder att de håller positioner över natten. Här har gränser lagts in vid (0.5,1.1) för köp-/säljkvoten. Dessa ger därmed utslag för de fyra punkter som ligger under 0.5 vilket visas i högermarginalen.

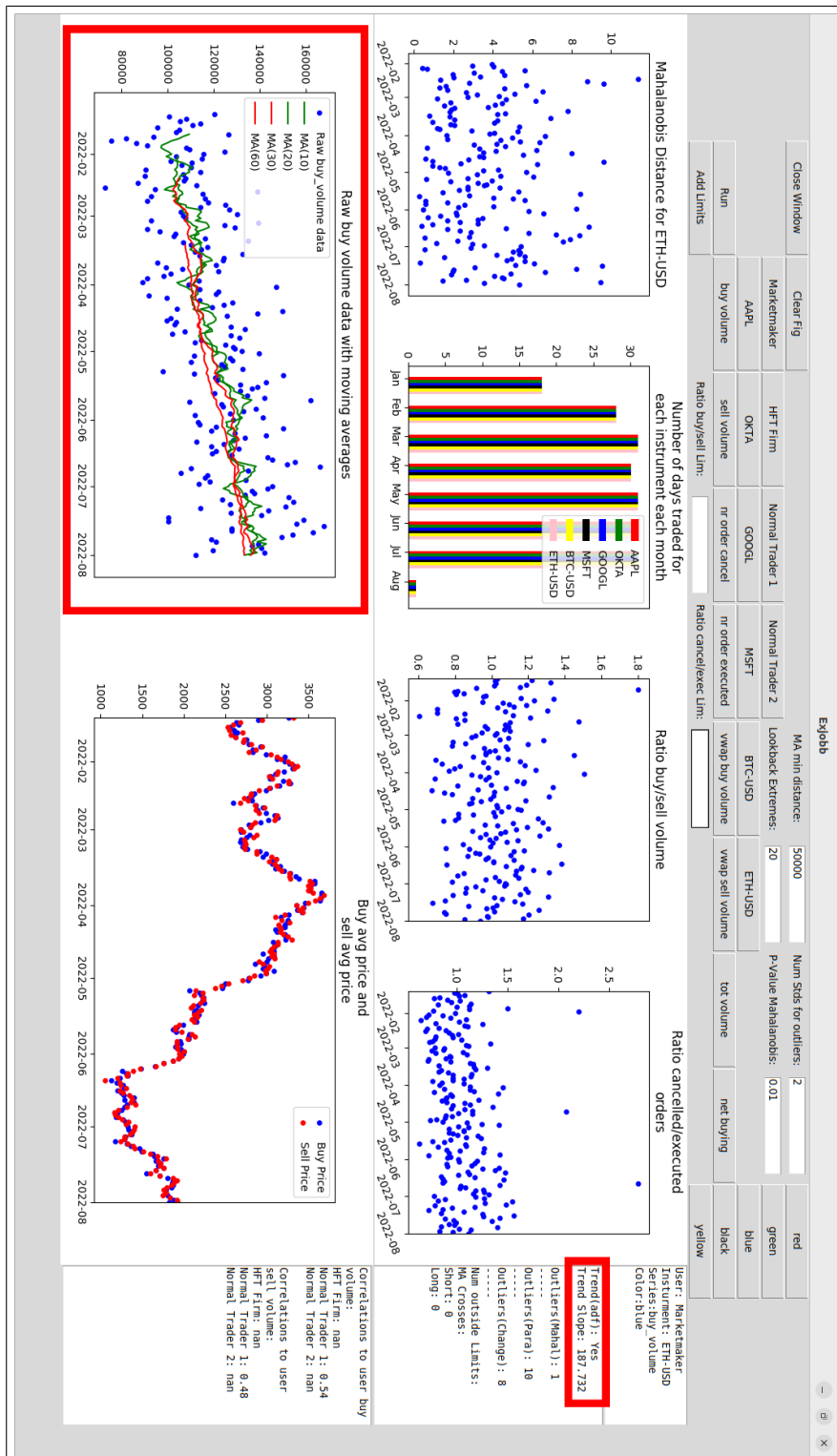
7.4 Market Maker



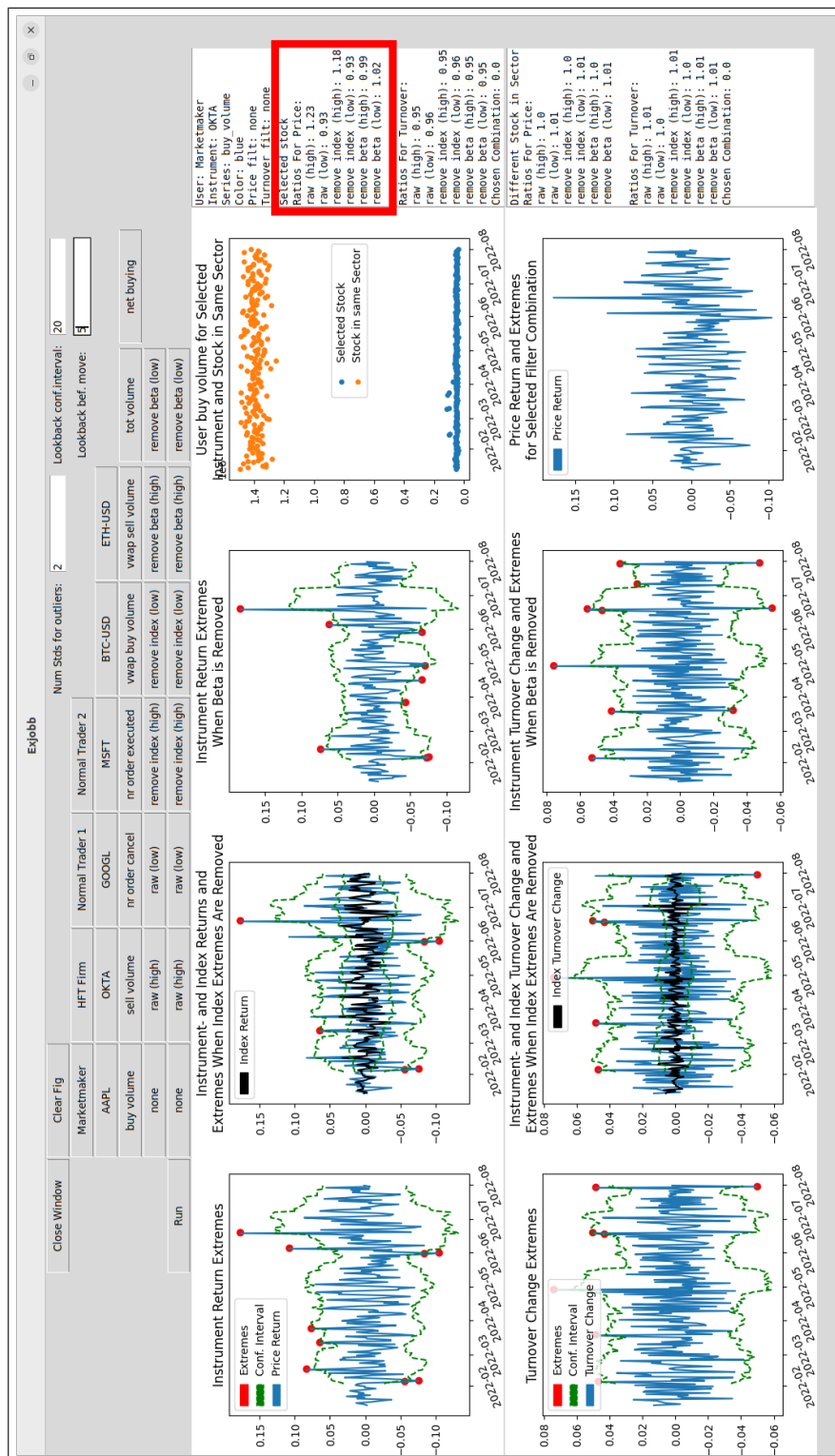
Figur 25: Mycket handel för market makern under de dagar då aktien har hög omsättning. Detta visas genom att inte välja något avkastningsfilter, men att välja omsättningsfiltret till de dagar då förändringen i log-omsättning är högre än 2 standardavvikelser (om man använder ett glidande fönster på 20 dagar). Detta plottar de röda linjerna.



Figur 26: Bild som visar hur limits fungerar i test-GUI 1 för market makern. Här har gränserna valts till 0.5 och 2 för både köp-/säljvolymförhållandet, och cancel/executed order förhållandet. I högra marginalen visas hur många dagar som hamnar utanför dessa gränser (0, respektive 1).

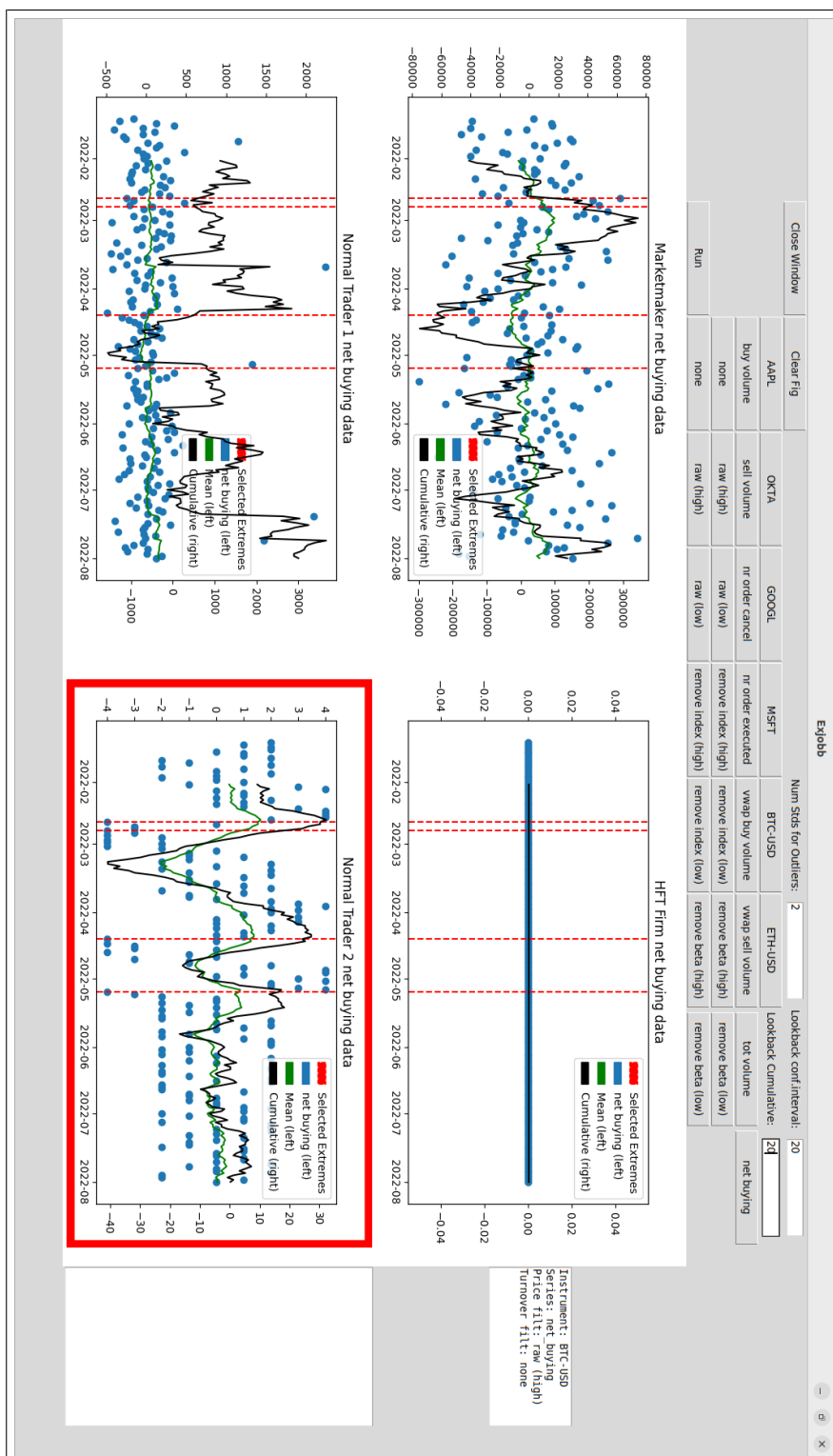


Figur 27: Figuren visar att köpvolymer trendar uppåt. Detta ger ett utslag i högermarginalen där Trend(ADF) visar "yes" samt att lutningen är ~ 187 . Detta betyder att volymen i snitt ökar med 187 kontrakt per dag. Detta kan vara rimligt för en market maker om den precis har börjat erbjuda ETH-USD till sina kunder, och fler och fler kunder börjar handla det.

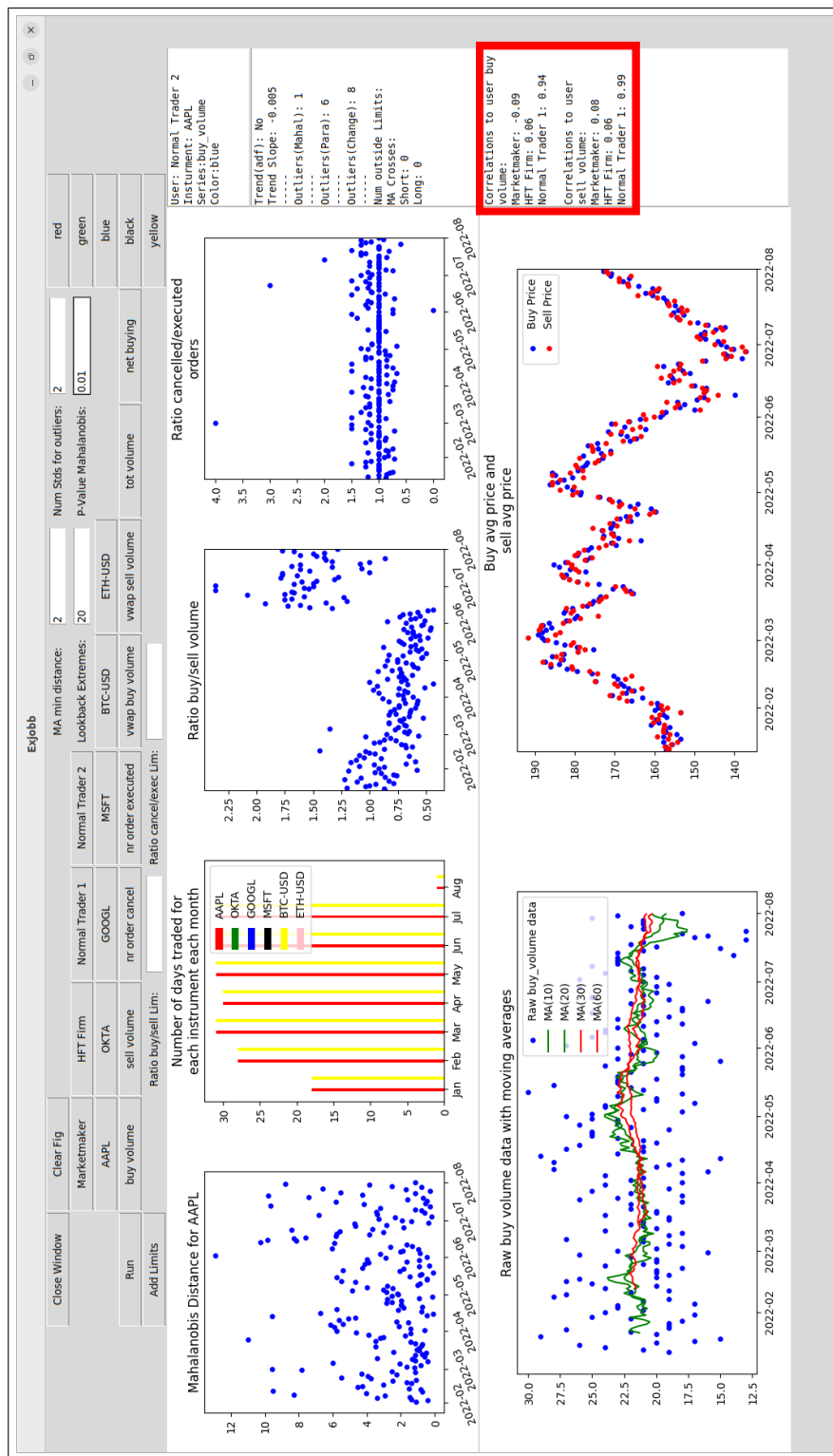


Figur 28: Här ser man återigen en förhöjd köpvolym de dagar som leder upp till dagar med höjd avkastning. Dock så ser man i detta fall att det endast är högre köpvolym för de två första avkastningsfiltrena, inte när man tar bort indexet med hjälp av beta. Därmed kan man antagligen säga att de volymökningar som skedde inte skedde pga insiderinformation, utan för att de som köpte misstänkte att indexet eller sektorn skulle röra sig mycket, t.ex vid räntesänkningar. Detta är inte olagligt.

7.5 Vanlig Aktör 2



Figur 29: Här syns ett exempel på misstänkt nettoköpvolum. I figuren för Vanlig Aktör 2 i nedre högra hörnet, syns att det ofta köps mer än det säljs under dagarna före en stor avkastningshöjning, och sedan säljs det efteråt. Detta syns speciellt i den kumulativa serien (svart).

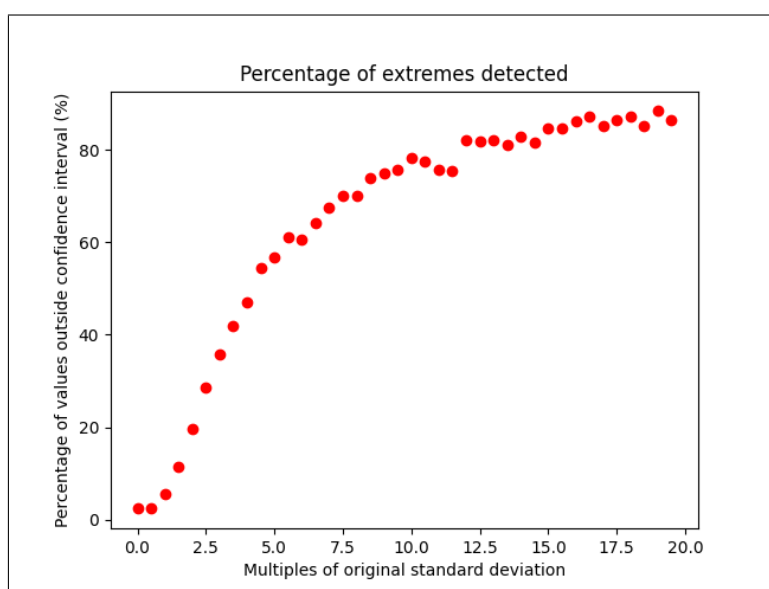


Figur 30: I nedre högra marginalen syns att VA2 har en mycket hög korrelation till VA1 när det gäller köp- och säljvolym. Detta är en indikation på att dessa två konton handlar med varandra, vilket kan vara ett tecken på wash-trading eller penningtvätt.

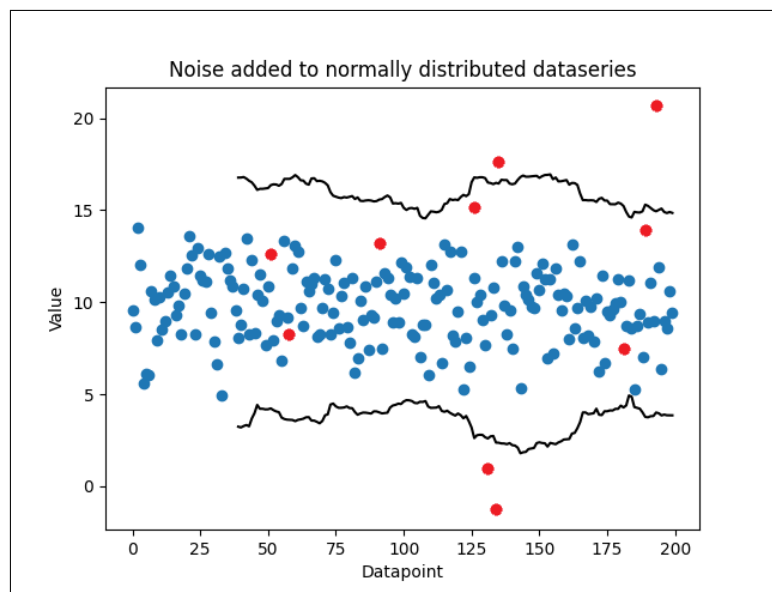
7.6 Numeriska Resultat

7.6.1 Extremvärden

Även om numeriska resultat är svåra att visa för detta projekt så visas ett exempel nedan i figur 31. I exemplet har 10 extrempunkter från en normalfördelning (bruset) $\mathcal{N}(0, n \cdot \sigma^2)$ lagts på en annan normalfördelning med fördelning $\mathcal{N}(\mu, \sigma^2)$ (ses i figur 32) där n är en variabel som går att variera för att testa måttets lämplighet. För att generera bilden varieras n i intervaller om 0.5 från 0 till 20. För varje n körs 100 simuleringar där resultatet (antal extremvärden utanför konfidensintervallet) noteras. Snittet för varje n räknas ut och plottas i figuren. Parametrarna för måttet är 3 standardavvikelse samt att längden på det glidande medelvärdet är 40 dagar. I figuren syns att måttet redan vid $n = 4$ lyckas identifiera 50% (5) av extremvärdena. Vid $n = 10$ är träffsäkerheten 80%. Alla extremvärden läggs in efter att det glidande medelvärdet blir aktivt, alltså mellan dag 40 och dag 200.



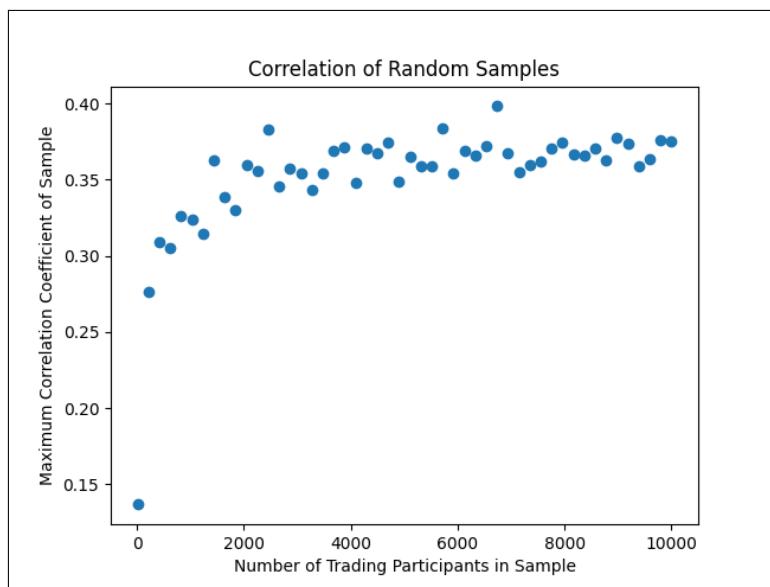
Figur 31: Här syns huruvida extremvärdesmåtten hittar extremvärden. Andelen punkter som detekteras visas på y-axeln och multipeln av varians syns på x-axeln.



Figur 32: Här syns en av dataserierna som testas för att skapa figur 31. I detta exempel är $n = 5$. Originalserien är plottad i blå och extremvärdena genererade med hjälp av bruset är plottade i rött.

7.6.2 Korrelation till andra aktörer

För måttet som kontrollerar korrelation mellan olika aktörer finns det en osäkerhetsgrad som ökar med antalet aktörer. Eftersom korrelationen testas mellan alla olika aktörer finns det en risk att det med många aktörer blir hög korrelation endast på grund av slump. Ett oändligt antal aktörer hade resulterat i att det alltid finns aktörer som har korrelation 1 till varandra. Lyckligtvis är det inte ett oändligt antal aktörer på de finansiella marknaderna. I figur 33 nedan syns hur den maximala korrelationen förändras med olika antal aktörer. Detta test gjordes genom att simulera 50 olika instanser med 10-10,000 aktörer (varje simulering bestod av en normalfördelad dataserie per aktör, med 200 "dagars" data med olika standardavvikelser och medelvärden). För varje simulering noterades den högsta korrelationen mellan två aktörer i simuleringen.



Figur 33: Här syns den maximala simulerade korrelationen för olika val av antal aktörer.

Detta test visar att korrelationen ökar med antal aktörer, men också att om värdet för när en aktör flaggas för ”hög korrelation” sätts till ett värde som är högre än dessa empiriska värden, bör man ändå kunna urskilja de som faktiskt handlar med varandra systematiskt, från de som endast har samma korrelation på grund av slump. För ett säkrare resultat bör detta test replikeras många gånger för varje antal aktörer, men det ger ändå en fingervisning om att måttet är användbart (om än inte för de allra största handelsplattformarna, beroende på hur mycket marginal man vill ha mellan flaggning och empiriskt slumpmaximum).

8 Diskussion

Till att börja med bör sägas att kryptovalutor och kryptotokens inte är helt förenliga med traditionella ekonomiska tankesätt. Den centraliserade aktiemarknaden gör att man ofta tar för givet att priser och volymer alltid är i princip exakta, medan så inte är fallet för kryptovalutor. Dessa värden, inklusive korrelationsbeta mellan t.ex bitcoin och övriga kryptovalutor (kryptoindex), är väldigt varierande beroende på vilken kryptobörs man hämtar data från [20, Appendix 2]. Även om kryptodata i det begränsade numeriska materialet i denna rapport verkar indikera att kryptovalutor fungerar för ändamålet, bör deras lämplighet undersökas ytterligare.

Metoderna som har tagits fram lyckas väl med att hitta de scenarion som har testas. De olika måtten kan tillsammans upptäcka ett stort antal olagliga handelsmönster vilket i slutändan gör det svårt att både tvätta pengar och att tjäna pengar på insiderhandel. T.ex är det svårt att tvätta pengar om detta system enkelt visar om en individ handlar mer än vanligt, har bytt instrument, samt om den handlar väldigt mycket med en speciell aktör. Det är på samma sätt svårt att lyckas med

insiderhandel om man inte kan handla mer än vanligt innan de dagar som en stor förändring i aktiepriset sker, eller om man inte kan göra större ändringar i antal ordrar, köpvolum eller instrument utan att det märks. Att helt plötsligt börja med layering och spoofing blir också svårt om ens aktiemäklare enkelt har koll på hur många ordrar en aktör lägger, annullerar och exekverar (även om just layering och spoofing är väldigt avancerat och antagligen hade behövt ytterligare mått för att upptäckas mer generellt). Tyvärr så finns det ingen riktig kund-data att tillgå då sådan data är mycket känslig, men de flesta av måtten som utvecklats bör fungera väl oavsett hur datan ser ut. Det goda resultatet till trots så finns det dock ändå ett antal problem med denna typ av analys. Ett problem med att hitta individer som gör olagligheter på finansiella marknader är att det är ett väldigt brett spektrum. I analysen ovan tittas det t.ex främst på insiderhandlare som handlar hyfsat nära in på förändringar i instrumentpriset. Dock så finns det med säkerhet individer som vet att det *antagligen* blir t.ex ett uppköp av en konkurrent många månader i förväg, och därmed kan köpa aktien utan att det syns i måttet. I dessa fall får man istället titta på specifika individers handel vilket är väldigt tidskrävande (att titta närmare på alla individer som har någon koppling till något börsnoterat bolag är i princip omöjligt) och svårt. Samtidigt så blir inte köpet självklart vinstbringande om det görs alldeles för långt i förväg, då andra händelser kan inträffa mellan köp-datumet och datumet då nyheten publiceras. I den andra änden av spektrumet återfinns individer som sitter högt uppsatta i företag och som ofta köper en aktie. De vet säkerligen om alla nyheter men om de alltid köper aktien och hade tänkt göra det oavsett, agerar de då olagligt?

Det absolut största problemet med resultatet är dock att olika aktörer är just olika. En privatperson handlar ganska sällan och vanligtvis i olika mängder och instrument, medan en professionell handlare håller sig till några få instrument och ofta har en tydligare metod till antalet aktier och storlek på sina positioner. Att försöka hitta gemensamma mått, och framförallt gemensamma parametrar för dessa mått, är väldigt svårt. Ännu svårare blir det om man även blandar in aktörer som market makers och HFT-firmor. Market makers använder sig av otroligt stora volymer för att kunna bibehålla en mer effektiv marknadsplats för sina kunder (och tjäna pengar på spread). En HFT firma brukar å andra sidan ha mindre volymer, men otroligt många annullerade ordrar. Detta gör att dataserierna som ska försöka buntas ihop med de valda måtten är väldigt olika. Ett sätt som detta kan försöka lösas på är genom att försöka hitta egenskaper som gör olika aktörer "lika". Om man t.ex kunde bunta ihop alla HFT-firmor (t.ex med maskininlärning) och titta på dessa gemensamt skulle det vara något lättare att hitta gemensamma egenskaper och därmed välja parametrar för måtten som ger rimliga, och inte för många, utslag. Även detta kan dock vara svårt då även olika HFT-firmor handlar på olika sätt gentemot varandra, med t.ex olika orderstorlekar, instrument och hur ofta de handlar. Att de flesta HFT-firmor använder sig av olika strategier gör inte problemet lättare.

Ett annat sätt som maskininlärning skulle kunna användas på är genom att träna en komplett modell på de tidigare beskrivna måtten. Genom att träna modellen och låta den avgöra vilka mått som är betydelsefulla, samt hitta de parametervärden som

ger högst sannolikhet att hitta förövare, kan man göra mjukvaran mer lättanvändlig och effektiv. Samtidigt tillfredställer man finansinspektionens krav på transparens genom att måtten och den grafiska förklaringen fortfarande finns tillgänglig. Maskininlärningen blir alltså endast till ett ytterligare filter som filtrerar bort mått som inte är effektiva, samtidigt som det väljer parametrar som har störst chans att hitta förövare. Ett problem med denna approach kan dock vara att få tag på data att träna på. Denna data, alltså transaktionsdata från dömda förövare, kan vara mycket svår att få ta del av, om den ens finns sparad hos myndigheter.

Vid implementering av detta system i ett verkligt scenario hade ett larmsystem implementerats vilket kan användas för att varna då olika siffror och mått ger utslag. Sedan kan användaren aktivera olika mått och tröskelvärden för att få signalerat om någon aktör gör något olagligt, utan att behöva sitta och manuellt titta på varje aktör eller instrument. Som tidigare beskrivit så är det dock inte helt enkelt att bestämma dessa tröskelvärden. I nuläget (utan de idéer som har presenterats i diskussionsavsnittet) kan utgivaren av mjukvaran sätta rimliga default-värden men det är ändå svårt för kunden att välja värden som hittar rätt individer, samtidigt som det inte ska bli för många utslag då detta skapar mer jobb både för kunden och för finansinspektionen. Detta resulterar i ett problem med typ-1 och typ-2 fel. Om man gör systemet känsligare kommer man få fler falska utslag (typ-2 fel) samtidigt som bördan (och orimligheten) ökar exponentiellt för Finansinspektionen. I detta scenario hittar man dock i bästa fall alla som handlar på illegala grunder. Om man istället sätter mer extrema nivåer på parametrarna kommer detta resultera i typ-1 fel eftersom man då kommer missa fler fel aktörer som missbrukar de finansiella marknaderna, men gör det i begränsad skala. Individer som handlar förhållandevis låga volymer och liknande kommer antagligen inte bli upptäckta då dessa inte ger utslag i system som har för extrema parametrar.

Enligt Scila AB så löser de problemet med att välja rätt nivåer på två olika sätt. Ett sätt är genom erfarenhet. Den som använder mjukvaran kan t.ex veta från egen erfarenhet att en market maker brukar ha en kvot mellan köp- och säljvolym som ligger (exempelvis) nära 1, därför kan det vara rimligt att lägga in larm som signalerar om denna kvot överstiger 3.

En annan metod som Scila använder sig av är backtesting. Vad detta innebär är att användaren kan skapa larm, sedan köra dessa larm på gammal data, för att se hur många larm som triggades etc. Om för många larm aktiveras kan användaren titta närmare på varför och sedan trimma larmen för att få ett mer önskvärt antal. Även här hade det kunnat vara intressant att titta på möjligheter att med maskininlärning träna en algoritm som väljer bättre gränsvärden.

8.1 Numeriska resultat

Det numeriska resultatet är väldigt begränsat dels då ingen riktig användardata användes, men också för att dessa fenomen är svåra att simulera och testa. Något som kan sägas om resultatet i figur 31 och 32 är dock att det är lovande. Redan vid

en standardavvikelseökning med 5x hittar testet i snitt hälften av alla extremvärden. Detta betyder att om en aktör manipulerar en marknad genom att ändra t.ex sin köpvoly m vissa dagar, till synes på ett slumpmässigt sätt, behöver inte standardavvikelsen på hans avvikande ordrar vara extremt annorlunda från hans vanliga handelsmönster för att de med hög sannolikhet skall upptäckas, speciellt inte om flera av dessa transaktioner görs. Är transaktionerna ännu mer extrema hittas de med ännu större sannolikhet.

8.2 Begränsningar för denna typen av undersökning

Slutligen finns problemet med att det med denna sorts system ofta är svårt att *bevisa* att en aktör har handlat olagligt. Att säga att en aktör har handlat mycket innan en specifik dag då en stor nyhet kom ut räcker ofta inte för att bevisa något. Efter att en bank eller annan användare av detta program ser något potentiellt olagligt måste sedan åklagare ta fram bevis för att något olagligt har skett. Tyvärr är detta mycket svårt och det uppskattas att endast runt 20% av alla insiderhandlare blir åtalade [21]. Som tidigare berättat är det också otroligt mycket penningtvätt som sker, trots försök att motverka detta. System som det som har utvecklats i detta projekt är dock en viktig del för att avskräcka och motverka dessa problem.

Referenser

- [1] A. Doffou, “Insider trading: A review of theory and empirical work,” *Journal of Accounting and Finance Research*, vol. 11, no. 1, 2003.
- [2] L. Smigel, “Why Does The Stock Market Exist?” 2022, [Accessed: Nov 2022]. [Online]. Available: <https://analyzingalpha.com/why-does-the-stock-market-exist>
- [3] A. Thomson-DeVeaux, “What’s So Wrong With Insider Trading Anyway?” 2016, [Accessed: Oct 2022]. [Online]. Available: <https://fivethirtyeight.com/features/whats-so-wrong-with-insider-trading-anyway/>
- [4] “SEC Charges Hedge Fund Firm CR Intrinsic and Two Others in \$276 Million Insider Trading Scheme Involving Alzheimer’s Drug,” 2012, [Accessed: Nov 2022]. [Online]. Available: <https://www.sec.gov/news/press-release/2012-237htm>
- [5] Securities and E. Commission, “Enforcement Surveillance of Markets,” 1997, [Accessed: Nov 2022]. [Online]. Available: <https://www.sec.gov/oig/reportspubs/aboutoigaudit246finhtm.html>
- [6] M. A. Robertson, “Speech by SEC Staff: Insider Trading - A U.S. Perspective,” 1998, [Accessed: Oct 2022]. [Online]. Available: <https://www.sec.gov/news/speech/speecharchive/1998/spch221.htm>

- [7] J. Zdanowicz, “Trade-based money laundering and terrorist financing,” *Review of Law Economics*, vol. 5, pp. 3–3, 2009.
- [8] United Nations, “Tax abuse, money laundering and corruption plague global finance,” 2020, [Accessed: Nov 2022]. [Online]. Available: <https://www.un.org/development/desa/en/news/financing/facti-interim-report.html>
- [9] “Så undviker du att bli penningmålsvakt,” 2021, [Accessed: Oct 2022]. [Online]. Available: <https://polisen.se/utsatt-for-brott/skydda-dig-mot-brott/penningtvatt/undvik-att-bli-penningmalvakt/>
- [10] N. Oberheiden, “Crypto laundering: Bitcoin + money laundering,” *The National Law Review*, vol. 11, no. 287, 2021.
- [11] A. Batan, *A Quick Overview of Know Your Customer (KYC) in Financial Industry*, 2021.
- [12] S. R. McNamara, “The law and ethics of high-frequency trading,” *Minnesota Journal of Law, Science Technology*, vol. 17, no. 1, pp. 71–152, 2016.
- [13] S. Imisiker and B. K. O. Tas, “Wash trades as a stock market manipulation tool,” *Journal of Behavioral and Experimental Finance*, vol. 20, pp. 92–98, 2018.
- [14] M. Gruet, “Tracking crypto pump-and-dump operations on social media,” 2022, [Accessed: Oct 2022]. [Online]. Available: <https://techxplore.com/news/2022-07-tracking-crypto-pump-and-dump-social-media.html>
- [15] A. Bloomenthal, “Market Maker definition: What it means and how they make money,” 2021, [Accessed: Oct 2022]. [Online]. Available: <https://www.investopedia.com/terms/m/marketmaker.asp>
- [16] V. Hodge and J. Austin, “A survey of outlier detection methodologies,” *Artificial Intelligence Review*, vol. 22, no. 2, p. 85–126, 2004.
- [17] G. Elliott, T. J. Rothenberg, and J. H. Stock, “Efficient tests for an autoregressive unit root,” *Econometrica*, vol. 64, no. 4, pp. 813–836, 1996.
- [18] A. Jakobsson, *An Introduction to Time Series Modeling*, 5th ed. Studentlitteratur, 2019.
- [19] R. Brereton, “The chi squared and multinormal distributions,” *Journal of Chemometrics*, vol. 29, pp. 9–12, 2015.
- [20] C. Alexander and M. Dakos, “A critical investigation of cryptocurrency data and analysis,” 2019.
- [21] V. Patel and T. J. Putniņš, *How Much Insider Trading Happens in Stock Markets?* American Finance Association (AFA) Annual Meeting, 2020.