



School of Economics and Management
Lund University
Department of Economics

Kan Twitter prediktera oljemarknadens framtida avkastningar?

Kandidatuppsats: NEKH01

Författare: Kristian Axelsson

Handledare: Daniel Ekeblom & Erik Wengström

Abstract

Is it possible to predict the future returns for the crude oil index by looking at the activity on Twitter? It is not possible to predict the market according to the efficient market hypothesis but earlier research on the topic states that it is possible for Twitter to predict the markets. This paper examines this phenomenon further by investigating the relationship between the market and Twitter activity during the “Arab spring”. The relationship is scrutinized by examining the number of times a particular keyword is mentioned on Twitter. The collected time series represents five keywords consisting of Egypt, Yemen, Syria, Kurdistan, and finally Pakistan. The relationship between the five time series and the future returns of crude oil index is examined using a Vector Autoregressive-model. The model is further tested by applying a Granger-non-causality test to examine the keywords predictive power. It turns out that the keywords have a predictive capability about four days prior to the change, which also can predict the direction of the WTI-index future returns.

Key words: WTI returns, Prediction, Arab spring, Twitter, Granger-causalirty

Innehållsförteckning

1 Inledning	1
1.1 Introduktion	1
1.2 Syfte	2
1.3 Frågeställning	2
2 Teori	3
2.1 Tidigare forskning	3
2.1.1 Effektiva marknadshypotesen (EHM)	3
2.1.2 Finansiell Sociologi	3
2.1.3 Twitter som prediktor	4
2.2 Teoretisk referensram	4
2.3 Sammanfattning	5
EMH	5
Finansiell sociologi	5
Twitter som prediktor	5
3 Metod	6
3.1 Material	6
3.2 Metodval	7
4 Empiri	9
4.1 Korsvalidering	9
4.2 Behandling av data materialet	11
4.3 Granger-non-causality test	11
5 Diskussion	16
6 Litteraturförteckning	18
Datakällor	19
Appendix	20
Appendix A	20
Appendix B	21
Appendix C	23
Nollhypoteser bivariat VAR	22
Wald-test	22

1 Inledning

1.1 Introduktion

Är det möjligt att prediktera framtida avkastningar för WTI-råoljeindex?

Denna uppsats behandlar sambandet mellan inläggsaktivitet på Twitter under den ”Arabiska våren” och avkastningarna från WTI-råoljeindex. Sambandet undersöks via nyckelord baserade på de länder som omfattades av ”Arabiska våren”. Det är sambandet mellan avkastningarna beräknade på WTI-indexets spotkurser och det dagliga antalet gånger ett visst nyckelord omnämns via inlägg på Twitter. Nyckelorden består av namnet för de länder som berördes av ”Arabiska våren” samt Pakistan som togs med på grund av sin geografiska placering. Detta arbete fokuserar på att undersöka ifall det råder ett prediktivt samband mellan förekomsten av dessa nyckelord och avkastningarna för WTI-indexet.

Den ”Arabiska våren” är det händelseförlopp som startade i slutet av 2010 där en rad protester och upplopp ägde rum i Mellanöstern och Nordafrika. Detta händelseförlopp är intressant att undersöka på grund av dess omfattning men också på grund av den roll som de sociala medierna spelade för dess utveckling. De sociala medierna utgjorde det huvudsakliga kommunikations- samt organisationsmedlet för dessa protester och uppror. Det medförde också att personer världen över kunde ta del av informationen via de sociala medierna samtidigt som de konventionella medierna, vilket gjorde den ”Arabiska våren” historiskt unik (Comunello & Anzera, 2012, s. 453). Ytterligare en ny företeelse var att information läckte ut från länderna även om de konventionella medierna var reglerade eller förbjudna (Comunello & Anzera, 2012, s. 460-462).

Uppsatsen undersöker ifall ny information på Twitter kan påverka WTI-indexet samt WTI-indexets förmåga att reflektera denna information via sina avkastningar. Valet att undersöka Twitter beror på att inläggen är så pass korta att endast den viktigaste informationen förmedlas. Ytterligare en anledning att undersöka information från Twitter är dess förmåga att representera det aggregerade humöret (Bollen, Mao & Zeng, 2011, s. 7).

Arbetet belyser sambandet mellan oljeindexet och Twitters roll som indikator för protesterna och konflikterna som ägde rum under ”Arabiska våren”. Genom att ta hänsyn till informationen från Twitter kan aktörer förbereda sig på uppkommande prisförändringar på oljemarknaden som är till följd av konflikter i Mellanöstern området.

1.2 Syfte

Detta arbete undersöker inläggsfrekvensen för på förhand valda nyckelord relaterade till den ”Arabiska våren”. De variabler som undersöks har en kronologisk samt geografisk relation till händelseförloppen under den ”Arabiska våren”. Målet är att undersöka ifall inläggsaktiviteten för dessa nyckelord har någon prediktiv förmåga då det gäller de log-normala avkastningarna för WTI-index under tidsperioden. Ytterligare ett mål är att undersöka om prediktionen bidrar med information kring avkastningarnas riktning.

1.3 Frågeställning

Arbetet eftersträvar att möta och besvara följande frågeställning:

Hur ser det prediktiva sambandet ut mellan aktiviteten på Twitter och avkastningen för WTI-råoljeindex?

2 Teori

2.1 Tidigare forskning

2.1.1 Effektiva marknadshypotesen (EHM)

EHM är ett fundament för den finansiella teorin och utgår från principerna om att marknadspriset är en effektiv informationsbärare. EMH baseras på antagande om att en effektiv marknad är arbitragefri, diversifierbar samt att marknadens aktörer direkt och korrekt justerar priset efter ny information (Ashgarian & Nordén, 2010 s. 18). Marknadseffektiviteten kan ha olika styrkor enligt EMH, vilka som uppges nedan (Ashgarian & Nordén, 2010 s. 19).

- Svag marknadseffektivitet – Priset återspeglar historisk information.
- Halvstark marknadseffektivitet – Priset återspeglar all publikt tillgänglig information samt den historiska informationen.
- Stark marknadseffektivitet – Priset återspeglar insider information utöver all den publika samt historiska informationen.

Utöver dessa antagande och styrkor finns det tre olika typer av information, vilka som ligger till grund för att informationen ska återspegla sig i priset (Fama, 1970, s. 290). De tre informationstyperna utgörs av risk reduktion, förbättrad beslutgrund för drift och handelsvinster för investerare till följd av ensamrätt på ny information.

Enligt ovanstående antagande och teorier ska resultaten från denna studie inte påvisa något prediktivt samband. Detta beror på att arbetet undersöker om den historiska informationen från Twitter har ett förklarande samband för oljeindexets rådande avkastningar.

2.1.2 Finansiell Sociologi

Detta forskningsområde utgår från grunden om att människor inte är rationella och därmed inte heller marknaderna. Istället läggs fokus på människans psykologiska samt socialpsykologiska aspekter. Där människans preferenser samt beslut har sin grund i de upplevda sinnestillstånden och känslorna snarare än rationellt beslutsfattande (Lowenstein, 2000, s. 426). Priserna på marknaden återspeglar de beslut som investerare fattar och besluten som fattas återspeglar investerarens humör, vilket påverkas av samhällets aggregerade humör (Nofsinger, 2005, s.144-145). Under osäkerhet tenderar människor att fatta sina beslut efter andras beslut. Detta fenomen kallas för ”vallning” och är ett grupp beteende som tar sig uttryck på marknaden (Pretcher & Parker, 2007, s. 93). Ytterligare en grund för marknadens priser är de geografiska och socio-kulturella avstånd som existerar mellan marknaden och händelsen (Lucy & Zhang, 2009, 75-76).

I enlighet med de ovannämnda teorierna bör sociala uppslutningar återspegla de aggregerade sinnestillstånden, vilket bör leda till att prediktiva samband uppstår. Implikationen av detta är således att den historiska informationen från Twitter avspeglar sinnestillstånden som präglar aktörernas framtida agerande på oljemarknaden.

2.1.3 Twitter som prediktor

Sambandet mellan Twitter och aktiemarknaden har undersökts på olika sätt samt från olika infallsvinklar. Bollen, Mao och Zeng undersökte den prediktiva förmågan av det sociala humöret genom att skapa en mjukvara som tog hänsyn till innehållet för inlägg (tweets) på Twitter. Forskningen visade på att det existerade ett prediktivt samband mellan det aggregerade humöret på Twitter och framtida prisförändringar för Dow Jones Industrial Average (Bollen, Mao & Zeng, 2011, s. 7). En annan undersökning granskade relationen mellan nyckelord baserade på sinnesstämningar och olika aktieindex. Det var aktiviteteten för nyckelorden hopp, rädsla samt oro som undersöktes och det visade sig råda korrelation mellan aktiviteten i dessa nyckelord och indexen DJIA, S&P500, NASDAQ samt VIX (Zhang, Fuehres & Gloor, 2011, s. 61). Ytterligare en studie som berör Twitters prediktiva samband har genomförts. Där undersöks nyckelord baserade på ekonomiska termer, som guld, olja, jobb, ekonomi och dollar mot tillgångens respektive index. I denna studie undersöktes huvudsakligen aktiviteten för svaren (retweets) på inläggen och även i detta fall visade det sig finnas ett prediktivt samband mellan Twitter och avkastningen på DJIA (Zhang, Fuehres & Gloor, 2012, s. 30-33).

Denna studie skiljer sig från de ovannämnda i avseendet att undersöka sambandet mellan oljemarknaden och Twitters roll som informationskälla samt indikator under en tid som präglas av konflikter och protester. Kriserna och konflikterna har makroekonomiska implikationer såväl som konsekvenser för det aggregerade humöret, vilket fångas upp av Twitter. Detta arbete undersöker med hjälp av Twitter hur väl oljeindexet tar hänsyn till denna information.

2.2 Teoretisk referensram

Den grundläggande teorin för detta arbete är att de aggregerade tweetsen innehåller ytterligare information som förbises av marknaden. Arbetet har därav valt att fokusera på ett specifikt händelseförlopp under ett specifikt tidsintervall för att undersöka den påverkan som ny icke-institutionell information har på avkastningarna för oljeindexet. Denna typ av information ska enligt EMH inte ha någon inverkan på indexets pris eller avkastningar (Fama, 1970, s. 92). Det visar sig trots det råda statistiska samband mellan information från Twitter och förändringar av priser respektive avkastningar (Bollen, Mao, & Zeng, 2011, s.7). Även om EMH inte bidrar med någon förklaring till detta fenomen så visar den sig vara en modell som erbjuder generella förklaringar för marknadens beteende.

De statistiska sambanden finner förklaring i teorin om socialt humör, vilket tyder på att marknaden tar hänsyn till sinnestillstånd (Nofsinger, 2005, s. 144-146). Hänsynen till det sociala humöret medför en ytterligare dimension av komplexitet då prissättning och prisjustering ska förklaras i teorin. Forskning på den prediktiva aspekten av Twitter söker efter en förklaring via permanenta nyckelord och marknaden. Då innehållet samt aktiviteten på Twitter återspeglar sin sociala kontext är det rimligt att förvänta sig förändringar av trender, språkbruk och användande, vilket renderar nyckelorden verkningslösa. Det är skälet för detta arbetes valda avgränsning av tidsperiod samt händelseförlopp och nyckelord, vilket medför en undersökning av en specifik kontext.

2.3 Sammanfattning

EMH

- Aktörer på marknaden tar direkt hänsyn till all ny publik information och agerar sedan rationellt samt korrekt på den.
- Marknaderna antas vara halvstarka i sin marknadseffektivitet.

Finansiell sociologi

- Investerare har preferenser som grundar sig i sinnestillstånd och fattar beslut baserat på det kollektiva humöret i samhället.
- Under osäkerhet låter sig investerare ”vallas” av andra investerare.

Twitter som prediktor

- Innehållet av tweets och retweets har visat sig prediktera framtida pris- och avkastningförändringar.
- Statistiska samband mellan inläggsfrekvensen på Twitter för specifika nyckelord och marknadens framtida priser samt avkastningar har påvisats.

3 Metod

3.1 Material

Detta arbete har undersökt material som avser tidsintervallet 2011-01-19 till 2012-11-06 och förekommer i två olika former. Den första av dessa utgörs av spot-priser för råolja av kvalitén Western Texas Intermediate (WTI), vilket har hämtats från eia.gov. Den andra formen av datamaterial som samlades in, är kardinal data över frekvensen för antalet gånger nyckelorden nämns per dag via Tweets (inlägg på Twitter). Detta datamaterial samlades in via Topsy.com, vilket är ett företag som säljer analystjänster för Twitter.com. Informationen som Topsy.com säljer samt analyserar hämtas via en webcrawler, vilket endast samlar in publik information från Tweets. Informationen är globalt representerad då alla Tweets som webcrawlern kan komma åt samlas in. Materialet ämnar undersöka ett internationellt perspektiv vilket medför att nyckelorden är formulerade på engelska och inte arabiska. De nyckelord som undersöks i detta arbete är de som räknas upp nedan.

- Egypt
- Yemen
- Syria
- Kurdistan
- Pakistan

Den fullständiga listan över alla nyckelord där material inhämtades finns i Appendix B. De nyckelord som inte granskas i uppsatsen har för få observationer, vilket orsakar en singular matris. Följaktligen påverkas storleken för undersökningens omfattning.

Valet av nyckelord bygger på händelseförloppet av den ”Arabiska våren” som den skildras av Wikipedia.com samt bbc.co.uk. Alla nyckelorden bygger på ländernas engelska namn och de länder där endast protester uppstod specificerades ytterligare med ordet protest. Nyckelorden som representerar protester har för få observationer för att användas vid beräkning.

Målet med materialet att undersöka Tweets är att nå publik information som inte har institutionell grund. Informationen har möjlighet att påverka marknadspriset via den ökade risken för tillgången samt information om försämrade eller förbättrade drift för producenter av tillgången (Fama, 1970, s. 290).

Datamaterialet över spot-kurserna utgår från stängningspriserna för den amerikanska dagshandeln, vilket medför att det inte finns några noteringar under helger och amerikanska helgdagar. Materialet för inläggsfrekvensen på Tweetsen tar inte någon hänsyn till helger och helgdagar. För att fånga upp hela Twitteraktiviteten har spot-kursernas stängningspris för den sista handelsdagen fått utgöra prisnivån under helger och helgdagar.

Ett annat alternativ hade varit att endast ha med noteringar från Twitteraktivitet under de amerikanska handelsdagarna, vilket inte hade tagit hänsyn till de kulturella skillnaderna mellan USA och mellanösternområdet. Tidsserierna samt enklare förklaringar finns i Appendix B.

3.2 Metodval

Arbetet ämnar undersöka om det finns ett Granger-kausalt samband mellan tidsserierna för inläggsfrekvensen och de framtida avkastningarna för indexet. För att undersöka detta samband valdes att tillämpa ett Granger-non-causality test (GNC) på en Vector Autoregression (VAR). Nedan finns den generella formeln för en VAR-modell beräknad på stationära tidsserier (Zivot & Wang, 2002, s. 384).

En av anledningarna att detta arbete har valt att tillämpa en VAR-modell är dess egenskap att ta hänsyn till de förklarande variabelernas historiska utveckling samt den historiska utvecklingen av den beroende variabeln. Denna egenskap att ställa prognoser är även en lämplig grund för GNC-testet. GNC-testet har förmågan att undersöka om det Granger-kausala sambandet mellan en variabel och en annan även råder under konträra förhållanden, vilket medför att GNC-testet valdes till detta arbete.

Varje variabel i en VAR har sin egen ekvation som förklarar dess utveckling. Grunden för den beroende variabelns utveckling ligger i de laggade förklarande variabelerna, där en av dessa utgörs av de laggade versionerna av den beroende variabeln. GNC-testet utgår från följande förenklade definition. Den förklarande variabeln X sägs Granger-orska den beroende variabeln Y om de laggade värdena av både X och Y tillsammans predikterar Y bättre än de laggade värdena av Y allena (Zivot & Wang, 2002, s. 405). Y är i detta fall de log-normala avkastningarna för WTI-indexet och X är tidsserierna för respektive nyckelord.

$$\text{Formel 1.} \quad \vec{Y}_t = \vec{c} + \sum_{i=1}^P \Pi Y_{t-i} + \vec{\epsilon}_t$$

$$\vec{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \vec{c} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \quad \Pi = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1n} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \cdots & \beta_{nn} \end{pmatrix} \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad t = 1, \dots, T$$

Y är en (n×1) vektor för de beroende variabelerna och Y_{t-i} är en (n×1) vektor för de laggade värdena av Y vektorn, vilket utgör de förklarande variabelerna. \vec{c} är en (n×1) vektor för intercepten (α) samt Π är en (n×n) matris för koefficienterna (β) samt ϵ utgör en (n×1) vektor för residualerna och t utgör tiden. P är antalet tidsperioder som VAR-modellen ska laggas, vilket kommer att räknas fram via Akaike Information Criteria (AIC) och Final Prediction Error (FPE). Dessa mått är relativa på och testar modellens passform (Zivot & Wang, 2002, s. 388).

För att undersöka de Granger-kausala förhållandet mellan variabler utförs GNC-testet i form av ett Wald-test, där nollhypoteserna formuleras som följer. Den förklarande variabeln y_2 Granger-orsakar inte y_1 om alla koefficienterna för de laggade värdena av y_2 är lika med noll i ekvationen för y_1 . Detta säger att y_1 inte Granger-orsakar y_2 när det omvända sambandet råder. Implikationen är att matrisen för koefficienterna Π har nollhypoteser som inte förkastas för de element som tillhör variabeln y_1 i respektive regression y_2, y_3, y_4, y_5, y_6 , för ytterligare förklaringar se Appendix C. Variabeln y_1 , är de log-normala avkastningarna och y_2, y_3, y_4, y_5, y_6 , är tidsserierna för respektive nyckelord.

Då den valda modellen för detta arbete är en VAR-modell, kommer arbetet endast att undersöka om det råder ett linjärt samband mellan aktiviteten på Twitter och avkastningarna för WTI-indexet. Modellen är avsedd för det insamlade datamaterialet och undersöker sambandet som råder under tidsintervallet som nämns ovan. Det innebär att modellen inte nödvändigtvis kommer ha ett samband för andra tidsserier eller andra tidsintervall med samma nyckelord. Det beror på de kriser och revolutioner som uppstod under den ”Arabiska våren”.

4 Empiri

4.1 Korsvalidering

Korsvalideringen har till syfte att granska ifall aktiviteten för tweetsen i tidsserierna har något att göra med protesterna, d.v.s. om tidpunkterna sammanfaller med aktiviteten i tidsserierna för nyckelorden. Händelserna som granskas för varje tidsserie är vid förekommande fall start för protesten, störtandet av regering samt den amerikanska film som skildrade profeten Mohammed. Valet av händelser baserar sig på en lista från Wikipedia över viktiga händelser i mellanöstern under ”Arabiska våren”. Detta är ett sätt att säkerställa att det är den ämnade informationen för händelserna som undersöks och därmed att korrelationen motsvarar informationens samband med WTI-indexet. Nedan presenteras en tabell över de kontrollhändelser som har använts vid denna korsvalidering.

Tabell 1.

Händelse:	Datum:	Inläggsfrekvens för nyckelord:	Genomsnittlig Inläggsfrekvens:
Egypten			20211,46
”Friday of anger” – Protesterna startar	2011-01-28	339852	
“Friday of departure” – Regeringen störtas	2011-02-11	404439	
“Anti-islam film” – film som avbildar mohammed	2012-09-13	90067	
Jemen			3533,95
Protesterna startar	2011-01-27	8652	
Regeringen störtas	2012-02-27	4960	
”Anti-islam film”	2012-09-13	49543	
Syrien			14969,27
Protester startar	2011-03-25	19166	
Homs massaker	2012-02-04	85114	
”Anti-islam film”	2012-09-13	13233	
Kurdistan			730,10
Politiska problem i Kurdistan	2012-05-19	7819	
Syrien angriper syriska delen av Kurdistan	2012-07-28	7841	
”Anti-islam film”	2012-09-13	1101	
Pakistan			18830,69
Osama Bin Laden mördas	2011-05-02	242991	
”Anti-islam film”	2012-09-13	22464	
USA mördar Pakistans krigsherre	2012-10-23	64988	

Alla utvalda kontrollhändelser är av betydande art för det specifika landet utom ”anti-islam film” som fungerar som en kontrollhändelse för alla tidsserierna. Den genomgående kontrollhändelsen har till syfte att fungera som en måttstock för värdena. Den ger också en bild över den inhemska representationen av Tweets, då ett inbördeskrig eller en störtad regering anses alvarligare än en film.

För nyckelordet Egypten motsvarade inläggsfrekvensens storlek vikten av händelserna samt dess tidpunkt. Inläggsfrekvensen för nyckelordet Jemen tenderar att avvika från kontrollhändelserna. Syrien är ytterligare ett nyckelord där representationen av händelserna präglas av inkonsekventa inslag. Inläggsfrekvensen för Kurdistan och Pakistan representerar kontrollhändelserna samt övriga händelser genom ökad aktivitet vid de specifika tidpunkterna. Hela inläggsfrekvensen för samtliga nyckelords tidsserier med ovanstående händelser utmärkta finns i Appendix B.

Vid granskningen av datamaterialet framgår det att Jemens representation av datamaterialet för ”anti-islam filmen” främst återspeglar internationella omnämningar av nyckelordet Jemen till följd av våldsamheter vid Amerikanska ambassaden. Detta medför att aktiviteten för datamaterialet inte är en god representation för händelsen. Aktiviteten antar extrema värden på grund av variationer av det socio-kulturella avståndet, vilket är en implikation av internationell överrepresentation (Lucy & Zhang, 2009, s. 76).

För att beskriva sambandet mellan inläggsfrekvensen och avkastningarna av WTI-index beräknades korrelationerna för tidsserierna. Dessa korrelationer beräknades enligt följande ekvation och presenteras i nedanstående tabell.

$$\text{Formel 2.} \quad \rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

Tabell 2.

Nyckelord:	ρ :	t-statistika:
Egypten	-0,198583	-4,820453
Jemen	0,109679	-4,299622
Syrien	-0,05537	1,651151
Kurdistan	-0,17785	-1,319251
Pakistan	0,069237	2,625180

I tabellen ovan går det att avläsa att tidsserien för nyckelorden Jemen och Pakistan visar på en positiv korrelation med WTI-indexet. Tidsserien för nyckelordet Pakistan skildrar främst händelser med positiva sinnessillstånd vilket visar på en positiv korrelation medan de övriga tidsserierna främst skildrar händelser med negativa sinnessillstånd vilket leder till negativ korrelation. Dessa resultat överensstämmer med resultaten från Bollen, Mao och Zeng (2011) respektive Zhang, Fuehres och Gloor (2011). Tidsserien för nyckelordet Jemens avvikelse från detta mönster kan ses som en implikation av inläggsfrekvensens missrepresentation. I tabellen framgår det även att korrelationskoefficienterna för Syrien och Pakistan inte är signifikant skilda från noll, vilket säger att effekterna är försmå i förhållande till antalet observationer för att uppfylla kraven av en konfidensgrad på 95%.

4.2 Behandling av data materialet

Syftet med undersökningen är att undersöka de log-normala avkastningarna för WTI indexet beräknade på spot-priserna som genomfördes genom att tillämpa standard formeln.

$$Y_t = \ln\left(\frac{p_t}{p_{t-1}}\right) = \ln(p_t) - \ln(p_{t-1})$$

Där Y_t är de log-normala avkastningarna, p_t är spotpriset idag och p_{t-1} är gårdagens spotpris. För att kunna tillämpa formel 1 på tidsserierna krävs det att dessa är stationära. En stationär variabel rör sig kring nollan på Y-axeln vid en visuell granskning, se Appendix A. Tidsserierna har även testats via ett Dickey-Fuller enhetsrots test med en konfidensgrad på 95%, vilket påvisade att tidsserierna för nyckelorden samt för log-avkastningarna var stationära (Westerlund, 2005, s. 207-209). Tidsserierna testades därefter för heteroskedasticitet med ett White's test (Westerlund, 2005, s. 181-182). Testet visade att tidsserierna för nyckelorden led av heteroskedastiska residualer, vilket kunde mildras genom att tillämpa den naturliga logaritmen (Gujarati & Porter, 2010, s. 297-298).

4.3 Granger-non-causality test

Tidsserierna för den skattade VAR-modellen testas för kointegration mellan tidsserierna genom att använda ett Johansen's kointegrationstest med en konfidensgrad på 95%, vilket påvisade att de inte råder någon kointegration. Därefter testas VAR-modellen med AIC och Final Prediction Error (FPE) för att beräkna antalet laggar för estimatet som ska testas med GNC. Det maximala antalet laggar som testades av AIC och FPE sattes till 12. I nedanstående tabell presenteras resultaten av AIC och FPE testen.

Tabell 3.

Lag	FPE	AIC
0	1.19e-06	3.387219
1	7.57e-09	-1.671521
2	6.79e-09	-1.780162
3	6.69e-09	-1.795309
4	6.48e-09*	-1.828026*
5	6.63e-09	-1.805321
6	6.64e-09	-1.803693
7	6.89e-09	-1.766478
8	7.24e-09	-1.718565
9	7.74e-09	-1.652271
10	8.28e-09	-1.585983
11	8.33e-09	-1.580091
12	8.74e-09	-1.533739

Både AIC och FPE är relativa tester där lägsta värdet eftersträvas för att visa hur många tidsperioder VAR-modellen ska laggas. I tabellen ovan är det möjligt att avläsa att estimatet som ska testas med GNC enligt AIC och FPE ska laggas med 4 tidsperioder (P = 4). Hädanefter skildrar detta arbete endast estimat och resultat från följande regression.

$$\text{Formel 3.} \quad y_{1t} = \alpha_1 + \sum_{i=1}^P \vec{B} \vec{Y}_{t-i} + \epsilon_{1t}$$

$$\vec{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \vec{c} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \quad \vec{B} = (\beta_{11} \quad \dots \quad \beta_{1n}) \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Regressionen har de log-normala avkastningarna för WTI-indexet som den beroende variabeln y_1 . Interceptet α_1 är det första elementet i c-vektorn och P är antalet laggar i detta fall 4. B är den första radvektorn från Π -matrisen som innehåller VAR-modellens koefficienter. Där v är kolonnvektorn för variablerna med $n=6$ och $y_1, y_2, y_3, y_4, y_5, y_6$, representerar de laggade variablerna, respektive log-avkastningarna för WTI och log-aktiviteten för Egypten, Jemen, Syrien, Kurdistan, Pakistan. ϵ_1 är regressionens residualer och det första elementet från vektorn ϵ som innehåller residualerna. I nedanstående tabell presenteras koefficienterna från estimatet på regressionen.

Tabell 4.

Parameter (t-4)	Koefficient	Standard error	t-statistika
Log-avkastning WTI (y_1)	0.006573	0.03916	0.16784
Egypten (y_2)	-0.005667	0.00228	-2.49040
Jemen (y_3)	-0.002776	0.00136	-2.03457
Syrien (y_4)	0.001297	0.00197	0.65688
Kurdistan (y_5)	-0.000566	0.00188	-0.30042
Pakistan (y_6)	0.003658	0.00182	2.01285

Regressionens koefficienter beskriver hur många procentenheter som WTI-indexets log-avkastningen förändras i förhållande till ett nyckelords log-frekvens. Exempelvis om log-frekvensen för alla variablerna är $\ln(1000) = 6,9078$ vilket enligt modellen leder till att log-avkastningen för indexet blir $-0,034$, d.v.s. indexet sjunker med 3,4 procentenheter. De genomsnittliga inläggsfrekvenserna ges i tabell 1.

Av estimatet framgår att det finns en variabel vars t-statistika inte förkastar nollhypotesen vid någon av laggarna av den regression som ska undersökas och det är variabeln för tidsserien Syrien. En variabel vars t-statistika inte förkastar nollhypotesen bör inte vara med i regressionen då nollhypotesen för variabeln motsvarar att koefficienten är noll. Konfidensgraden för detta arbete är satt till 0,95 som motsvarar en t-statistika på 1,96. Variabeln Syrien utesluts trots det inte från modellen eftersom den har en teoretisk grund att vara med samt att det annars förekommer en risk för omitterad variabel bias. Konsekvensen av att ha med variabeln påverkar varken väntevärdesriktigheten eller konfidensintervallet vid testning eller dess varians, vilket är konsekvenserna av att felaktigt utesluta variabeln Syrien (Gujarati & Porter, 2010, s. 221-227).

GNC-testet genomförs på hela VAR(t-4)-modellen då det tillåter en att granska relationerna mellan variablerna. Råder ett Granger-kausalt samband mellan en variabel och en annan bör inte ett Granger-kausalt samband råda under motsatta förhållanden. GNC-testet på VAR-modellen kommer att genomföras via ett Wald-test, vilket har en Wald-statistika som jämförs mot χ^2 -fördelningen och har en konfidensgrad på 95%, se Appendix C. GNC-testet genomförs enligt ekvationerna 4 och 5, vilket motsvarar ett tidsintervall på 653 observationer/dagar. Nedan presenteras ekvation 4 och 5 vilket motsvarar GNC-testet för regression 3.

$$\text{Formel 4.} \quad y_{1t} = \alpha_1 + \sum_{i=1}^P \beta_i y_{1t-i} + \epsilon_{1t}$$

$$\text{Formel 5.} \quad y_{1t} = \alpha_1 + \sum_{i=1}^P \beta_i y_{1t-i} + \sum_{i=1}^P \vec{\gamma}_i \cdot \vec{x}_{t-i} + \epsilon_{1t}$$

$$\vec{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \vec{c} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \quad \vec{x} = \begin{pmatrix} y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad \vec{\gamma} = (\beta_{12} \quad \cdots \quad \beta_{1n})$$

$$\vec{B} = (\beta_{11} \quad \cdots \quad \beta_{1n})$$

Där y_{1t} är de log-normala avkastningarna för WTI-indexet, d.v.s. y_1 från vektorn \mathbf{Y} och α är första elementet från vektorn \mathbf{c} . β (β_{11}) är koefficienten för WTI-indexets log-avkastningar och P är 4 och motsvarar laggen för estimatet. ϵ (ϵ_1) är första elementet ur vektorn ϵ som innehåller residualerna. γ är en radvektor (1×5) för koefficienterna och \mathbf{x} är en kolonnvektor (5×1) för y_2, y_3, y_4, y_5, y_6 som representerar de laggade variablerna, respektive log-aktiviteten för Egypten, Jemen, Syrien, Kurdistan, Pakistan. Nedan presenteras resultaten för GNC-testet.

Tabell 5.

Beroende variabel: WTI			
Namn:	χ^2 -Statistika:	Frihetsgrader:	P-värde:
Egypten	9.234062	4	0.0555
Jemen	6.634285	4	0.1565
Syrien	3.655228	4	0.4547
Kurdistan	11.57905	4	0.0208
Pakistan	10.03376	4	0.0399
Alla	39.91594	20	0.0051

Den valda konfidensgraden för GNC-testet är 0,95. Detta medför att endast Kurdistan och Pakistan som enskilda variabler visar på ett prediktivt Granger-kausalt samband med log-avkastningarna för WTI. Det råder även ett Granger-kausalt samband mellan samtliga variabler och log-avkastningarna för WTI.

Det Granger-kausala sambandet mellan log-avkastningarna från WTI och nyckelorden Kurdistan samt Pakistan motsvaras inte under omvända förhållanden. Ingen av de ovannämnda nyckelorden visade på Granger-kausala samband under konträra förhållanden. Detta beror på att nollhypoteserna för koefficienterna som rörvariabeln y_1 samt de laggade variablerna av y_1 inte kan förkastas för regressionerna y_2, y_3, y_4, y_5, y_6 , vilket motsvara log-frekvenserna för nyckelorden Egypten, Jemen, Syrien, Kurdistan, Pakistan och y_1 är log-avkastningarna från WTI-indexet.

I tabell 4 är det möjligt att avläsa att koefficienten för Pakistans variabel är positiv medan variabeln för Kurdistan har en koefficient med ett negativt värde. Pakistans tidsserie innehåller två stora toppar som beror på cricket och en för Usama Bin Ladens död. Dessa återspeglar positiva nyheter medan Kurdistans tidsserie snarare återspeglar negativa nyheter från politisk oro och krig se Appendix A. Utfallen för Pakistans samt Kurdistans koefficienter överensstämmer med respektive Bollen, Zheng och Mao's (2011) och Zhang, Fuehres och Gloor's (2011) utfall.

Utöver koefficienterna kan det även avläsas i tabell 2, att tidsserien för nyckelorden för Pakistan visar på en negativ korrelation med tidsserien över de log-normala avkastningarna från WTI-indexet. Detta tyder på att det råder korrelation mellan representation av optimism samt pessimism på Twitter och den framtida riktningen för avkastningarna på WTI-index.

5 Diskussion

Detta arbete har undersökt ifall det råder någon korrelation mellan inläggsaktiviteten på Twitter och avkastningarna från WTI-index. Resultaten visar på att det råder en korrelation mellan inläggsaktiviteten och avkastningarna men även ett prediktivt samband. Utav de 5 valda nyckelorden visade sig 2 stycken ha ett enskilt Granger-kausalt samband med WTI. Aktiviteten för dessa nyckelord återspeglas cirka 4 dagar efteråt av förändringar i avkastningarna för WTI-indexet. Resultaten visar också att aktiviteten för alla nyckelorden tillsammans har en prediktiv förmåga för avkastningarna som är cirka 4 dagar. Uppsatsen demonstrerar att ny publik information via Twitter inte omedelbart återspeglas av marknaden. Detta tyder på vikten av de sociala processerna vid beslutsfattande samt för investeringar.

En implikation av arbetets resultat är att hänsyn till ny information inte omedelbart tas av oljeindexet, vilket tyder på att aktörerna på marknaden antingen missar den nya informationen eller inte vet hur de ska ställa sig till den. Detta bör inte uppstå enligt EMH under förutsättningar att halvstark eller stark marknadseffektivitet råder. Det kan ses som att marknadseffektiviteten för WTI-indexet är svag men det är en osannolik slutsats då resultaten visar att WTI-indexet tar hänsyn till information som inte uppfyller de tre kraven som Fama (1970) nämner. Resultaten visar på det finns ytterligare informationssamband som EMH inte fullt beaktar.

Resultaten tyder på att avkastningarna för WTI-indexet sjunker till följd av konflikter och protesterna i länder som inte har någon större oljeproduktion. Detta ger det stöd åt teorin om att människor fattar investeringsbeslut baserat på sinnessillstånd som Lowenstein (2000) presenterar i sin artikel. Det kan också ses som en följd av det sociala aggregerade humöret då både arbetets korsvalidering av materialet samt koefficienter tyder på tidsseriernas inflytande på WTI-indexet beror på om sinnessillstånden är positiva eller negativa (Nofsinger, 2005). Ytterligare en aspekt som ger stöd åt det resonemanget är kontexten inom vilken missnöjet spred sig mellan de nordafrikanska och arabiska länderna under den ”Arabiska våren”, vilket tyder på ett socialt humör i samhället men också inslag av ”vallning” (Pretcher & Parker 2007). ”Vallning” kan även vara en bakomliggande faktor att de negativa korrelationerna är starka jämfört med de positiva men det krävs ytterligare tester för att påvisa någon koppling. Via korsvalideringen framgår det att tidsserier med överrepresentation av tweets från andra kulturer leder till missvisande material, vilken kan vara orsak till den positiva korrelationen för Jemen och WTI-indexet samt att det inte finns ett Granger-kausalt samband för både Jemen och Syrien (Lucy och Zhang, 2009).

Likt resultaten från Bollen, Mao och Zeng (2011) samt Zhang, Fuehres och Gloor (2011) studier visar detta arbete på att det finns ett Granger-kausalt samband mellan Twitter och marknaden. Till skillnad från de tidigare nämnda studierna undersöker detta arbete Twitter-aktivitet som baserar sig på samhällskonflikter. Dessa konflikter har även en makroekonomisk påverkan för de drabbade länderna då regeringarna för Egypten och Jemen störtades samt att ett pågående inbördeskrig bröt ut i Syrien. En implikation av dessa händelser är att det tydligt framgår vilka institutionella nyheter som släpps i landet och att det är möjligt att särskilja effekterna av Tweetsen från dessa.

En begränsning för denna uppsats är att den inte har tagit hänsyn till de geografiska eller kulturella avstånden för de undersökta tweetsen. För att ta hänsyn till dessa faktorer krävs en mer extensiv undersökning än denna uppsats tillåter. Denna uppsats pekar dock på att resultaten kan förbättras då hänsyn till dessa faktorer tas. Ett förslag är att anamma uppdelningen av information efter geografiska samt kulturella avstånd som tas upp av Lucy och Zhang (2009) och tillämpa den på tweets. Ett sätt att utveckla undersökning så att den tar hänsyn till det geografiska avståndet är att tillämpa nyckelord på arabiska då det kan fungera som en form av geografiskt filter. En annan begränsning för denna undersökning är att den endast undersöker det linjära sambandet. Det starkaste sambandet mellan ny information på Twitter och avkastningarna för WTI behöver således inte vara linjärt. Det finns utrymme för vidare forskning även från denna infallsvinkel.

Uppsatsens tillvägagångssätt för val av nyckelord har ytterligare en svaghet då det är svårt att hitta samt urskilja vid ett tidigt stadium vilka händelser som är lämpliga för prediktion via denna metod. Det är även svårt att avläsa eller hämta den aggregerade informationen i real-tid. Ett sätt att lösa dessa problem kan vara att via programmering använda sig av Twitters öppna programmeringsgränssnitt (API) för sällning. Möjligheterna med Twitters API har inte undersökts i denna uppsats, detta område bör därav undersökas vidare av framtida forskning.

Genom att endast undersöka ny information via Twitter med anknytning till ”Arabiska våren” under det valda tidsintervallet återspeglar arbetet inte de normala sociala förutsättningarna. De indikationer som detta arbete har fått av materialet angående normala förutsättningar säger att det hitintills är svårt att få tillräckligt med data från Twitter för den valda ekonometriska modellen om händelsen inte är ett extremfall. Att undersöka om liknande resultat åstadkoms vid normala förutsättningar eller vid andra extremfall återstår därför att undersökas.

Valet att enbart undersöka nyckelord knutna till ”Arabiska våren” för med sig ytterligare en begränsning, nämligen att dess prediktiva förmåga är problematisk att granska. Ett tillvägagångssätt för att undersöka dess prediktiva förmåga är genom in-sample och out-sample metoden. Den metoden lämpar sig inte i detta fall då det är olje-indexets förmåga att ta hänsyn till ny information från Twitter med avseende på händelserna i ”Arabiska våren” som undersöks.

Resultaten från detta arbete visar att Twitter kan tillämpas som en indikator för psykologiska effekter på oljemarknaden. Implikationen av detta är att Twitter kan ge kompletterande information till beslutsunderlag för investeringar och riskbedömning. För att göra det möjligt krävs dock ytterligare modifiering av den ekonometriska metod som presenteras i detta arbete då detta arbete endast presenterar en ex post undersökning. Modifiering bör tillmötesgå de ovannämnda avgränsningarna och restriktionerna med denna metod.

6 Litteraturförteckning

Asgharian, Hossein & Nordén, Lars, Räntebärande instrument: värdering och riskhantering, Lund, Studentlitteratur, 2007.

Bollen, Johan & Mao, Huina & Zeng, Xiaojun, "Twitter mood predicts the stock market", Journal of Computer Science, Bloomington & Manchester: Indiana university, School of Informatics and Computing & University of Manchester, School of Computer Science, 2010

Cameron, A. Colin & Trivedi, Pravin K., Microeconometrics: Methods and Applications, Cambridge University Press, 2005.

Comunello, Francesca & Anzera, Giuseppe, Will the revolution be tweeted? A conceptual framework for understanding the social media and the Arab Spring, Rom: Sapienza University of Rome, Department of Communication and Social Research 2005.

Fama, Eugene F. & Laffer, Arthur B., "Information and Capital markets", Journal of Business, Chicago: University of Chicago Press, 1970.

Gujarati, Damodar N. & Porter, Dawn C., Essentials of Econometrics, New york: McGraw-Hill, 2010.

Lowenstein, George, "PREFERENCES, BEHAVIOR, AND WELFARE Emotions in Economic Theory and Economic Behavior", The American Economic Review, Vol. 90, No. 2, American Economic Association, 2002.

Lucey, Brian M. & Zhang, QiYu, "Does cultural distance matter in international stock market co-movement? Evidence from emerging economies around the world", Emerging Markets Review, Elsevier 2009.

Nofsinger, John R., "Social Mood and Financial Economics", The Journal of Behavioral Finance 2005, Vol. 6, No. 3, The Institute of Behavioral Finance, 2005.

Pretcher, Robert R. & Parker, Wayne D., "The Financial/Economic Dichotomy in Social Behavioral Dynamics: The Socioeconomic Perspective", The Journal of Behavioral Finance 2007, Vol.86, No.2, The Institute of Behavioral Finance, 2007.

Westerlund, Joakim, Introduktion till ekonometri, Lund: Studentlitteratur, 2005.

Zhang, Xue & Fuehres, Hauke & Gloor, Peter A., "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear"", Procedia - Social and Behavioral Sciences 26, Elsevier 2011.

Zhang, Xue & Fuehres, Hauke & Gloor, Peter A., Predicting Asset Value through Twitter Buzz, Berlin: Springer-Verlag, 2012.

Zivot, Eric and Wang, Jiahui (Jeffery), Modeling Financial Time Series with S-PLUS, Springer-Verlag, 2002.

6.1 Datakällor

www.eia.gov

topsy.com

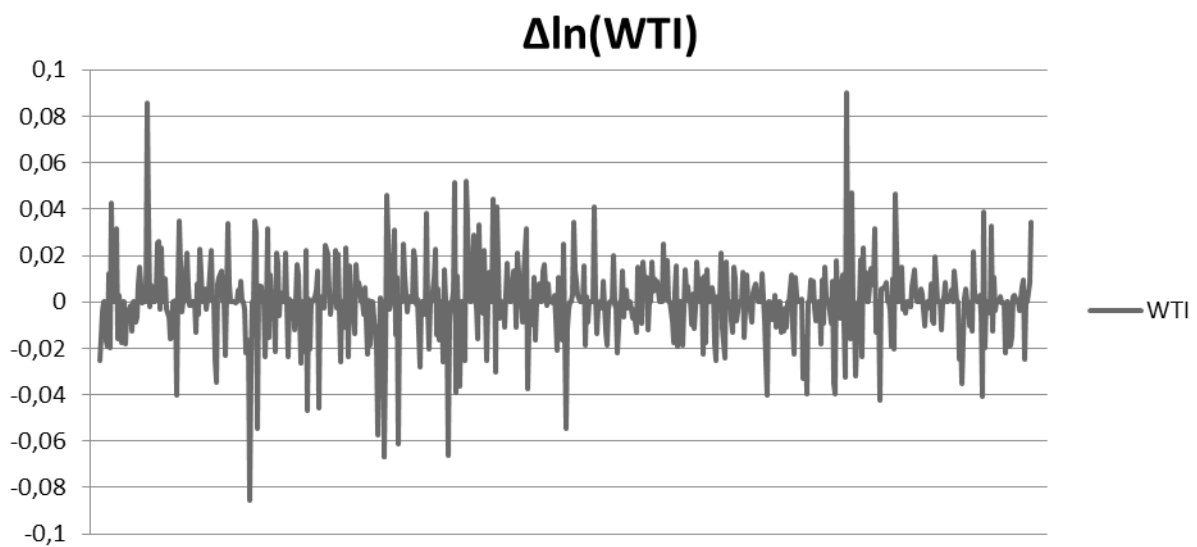
Appendix

Appendix A



Figur 1

Obehandlad tidsserie som uppenbart inte är stationär då den inte rör sig kring nollan för Y-axeln.



Figur 2

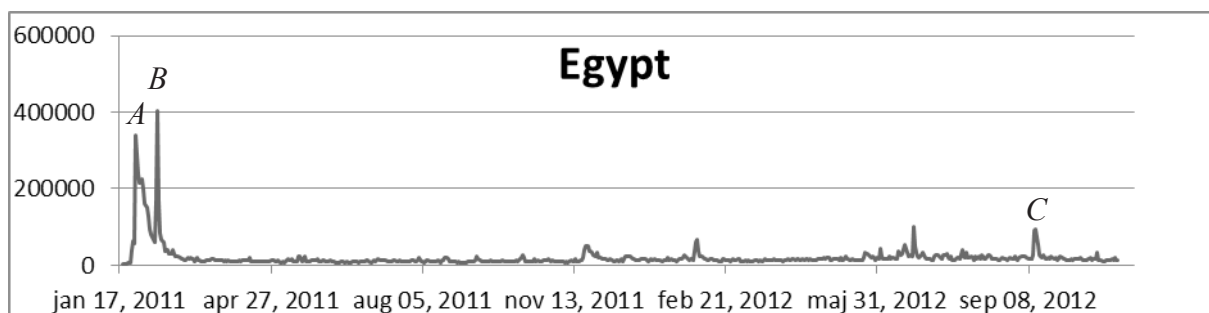
Tidsserien rör sig kring värdet noll för Y-axeln vilket är ett tecken på att den vid det här laget är stationär.

Appendix B

Nedan anges den fulla listan med nyckelord som samlades in till undersökningen i kronologisk ordning.

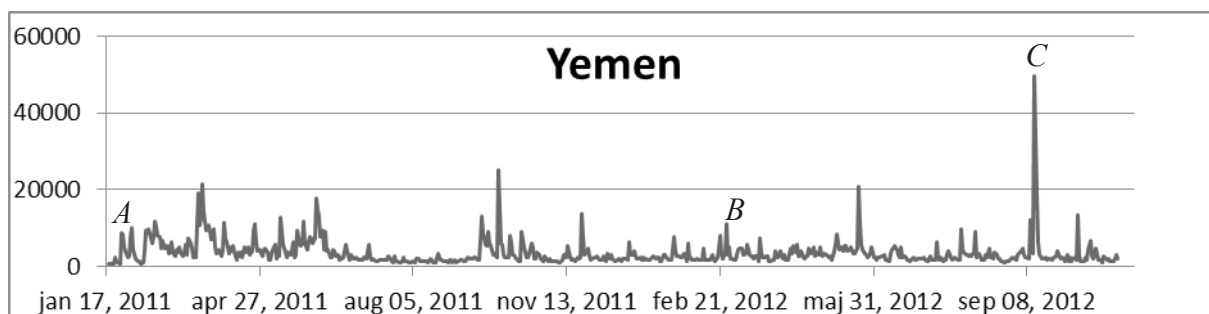
Tabell 6.

Nyckelord:		
Jordan protest	Iraq protest	Saudi Arabi protest
Oman protest	Bahrain protest	Syria
Egypt	Kuwait protest	Kurdistan
Yemen	Marocco protest	Pakistan
Djibouti protest	Mauritania protest	Israeli protest
Sudan protest	Lebanon protest	



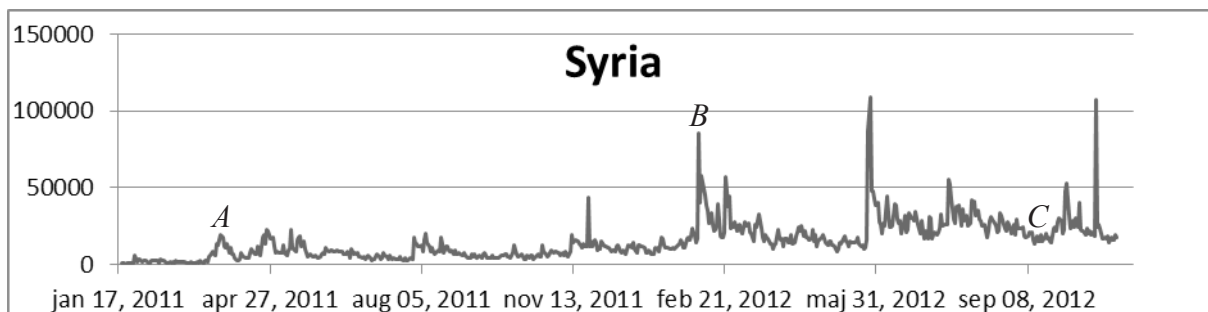
Tabell Eygypen

Hänvisning:	Händelse:	Tidpunkt:
A	”Friday of anger” – Protesterna startar	2011-01-28
B	“Friday of departure” – Regeringen störtas	2011-02-11
C	“Anti-islam film”	2012-09-13



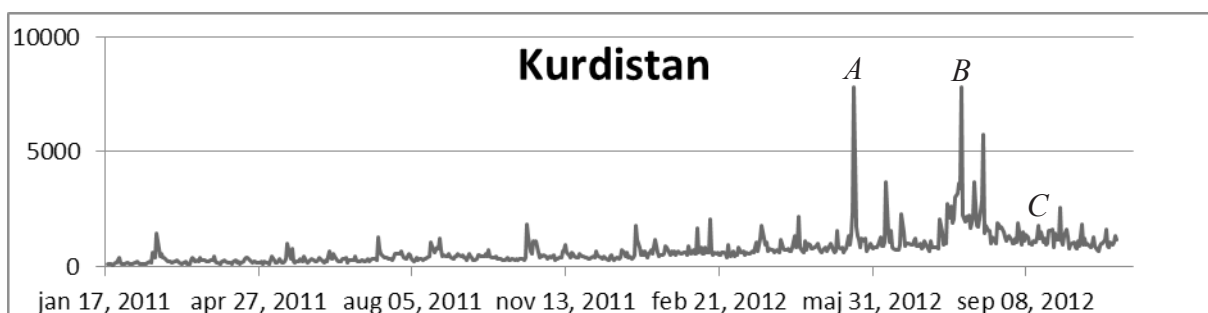
Tabell Jemen

Hänvisning:	Händelse:	Tidpunkt:
A	Protesterna startar	2011-01-27
B	Regeringen störtas	2012-02-27
C	“Anti-islam film”	2012-09-13



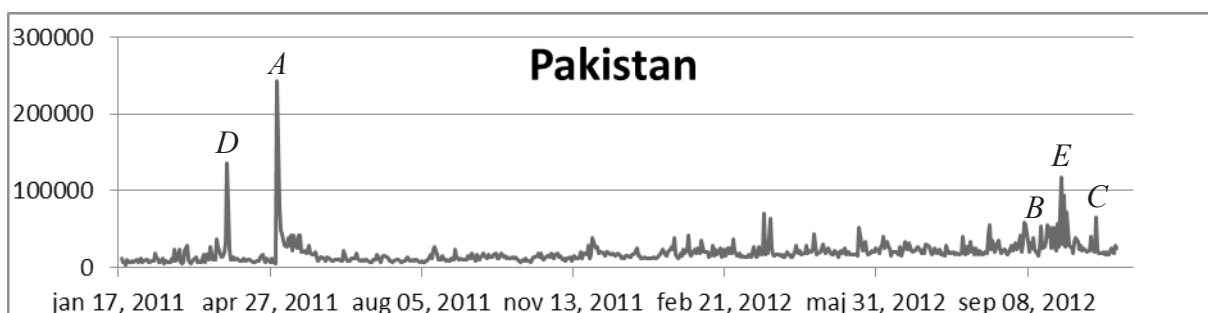
Tabell Syrien

Hänvisning:	Händelse:	Tidpunkt:
A	Protester startar	2011-03-25
B	Homs massaker	2012-02-04
C	“Anti-islam film”	2012-09-13



Tabell Kurdistan

Hänvisning:	Händelse:	Tidpunkt:
A	Politiska problem i Kurdistan	2012-05-19
B	Syrien angriper syriska delen av Kurdistan	2012-07-28
C	“Anti-islam film”	2012-09-13



Tabell Pakistan

Hänvisning:	Händelse:	Tidpunkt:
A	Osama Bin Laden mördas	2011-02-11
B	“Anti-islam film”	2012-09-13
C	USA mördar Pakistans krigsherre	2012-10-23
D	Cricket VM semifinal mot Indien	2011-03-30
E	Cricket mot Indien	2012-09-30

Appendix C

Nollhypoteser bivariat VAR

Nedan ges exempel på formuleringar av nollhypoteser för Granger-kausala nollhypoteser beroende på samband.

$$\begin{array}{cccc}
 H_0: \begin{matrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{matrix} & H_0: \begin{matrix} \beta_{11} & 0 \\ \beta_{21} & \beta_{22} \end{matrix} & H_0: \begin{matrix} \beta_{11} & \beta_{12} \\ 0 & \beta_{22} \end{matrix} & H_0: \begin{matrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{matrix} \\
 x \leftrightarrow y & x \Rightarrow y & y \Rightarrow x & x \perp y
 \end{array}$$

Den första av de ovanstående nollhypoteserna testar ett Granger-kausalt samband oberoende av riktning mellan variablerna. Den andra nollhypotesen testar om y Granger-orsakar x men x Granger-orsakar inte y. Den tredje nollhypotesen testar motsatsen till den andra nollhypotesen och slutligen den fjärde nollhypotesen testar det variablerna är oberoende av varandra genom att ingen variabel Granger-orsakar den andra (Zivot & Wang, 2002, s. 405). Ovan motsvaras y av de log-normala avkastningarna från WTI-indexet samt x är log-frekvensen för nyckelordet Egypten och $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$ är deras respektive koefficienter under en VAR(1)-modell.

Waldtest

För att kunna testa VAR(P) modellen med ett Wald-test bör VAR(P) modellen omskrivas till nedanstående uttryck.

Formel 6. $\vec{y}_i = \mathbf{Z} \vec{\beta}_i + \vec{\epsilon}_i \quad i=1, \dots, n \quad n = \text{antal regressioner}$

$$\vec{y}_i = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} \quad \mathbf{Z} = \begin{pmatrix} 1 & \cdots & 1 \\ y_{11} & \cdots & y_{1T-P} \\ \vdots & \ddots & \vdots \\ y_{k1} & \cdots & y_{kT-P} \end{pmatrix} \quad \vec{\beta}_i = \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad \vec{\epsilon}_i = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

Där y_i är en $(T \times 1)$ vektor med observationer motsvarande ekvation i, Z är en $(T \times k)$ matris där raden t ges av variablerna och de laggade värdena av variablerna upp till värdet P, vilket beskrivs av nedanstående uttryck.

$$\mathbf{Z}_t' = (1, y'_{t-1}, \dots, y'_{t-p}) \quad k = np + 1$$

β är en $(k \times 1)$ vektor för parametrarna där k ges av det ovanstående uttrycket och ϵ är en $(T \times 1)$ vektor där T är tiden för sista observationen för tidsserien. Genom att vektormultiplicera y med raden av värdet t och därefter invertera produkten av kolumnen t och raden t, fås $(k \times n)$ matrisen för koefficienterna. Ekvationen för ovan nämnda beräkningar ges nedan (Zivot & Wang, 2002, s. 386).

Formel 7. $\hat{\mathbf{\Pi}} = \mathbf{y} \mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1}$

För att göra inferens för VAR(P)-modellen möjlig via ett Wald-test krävs det att en Wald-statistika beräknas. Denna Wald-statistika är approximativt $\chi^2(h)$ -fördelad med h restriktioner, vilket motsvarar den full rank matrisen för \mathbf{R} och tar sig uttrycket nedan (Zivot & Wang, 2002, s. 388).

$$\text{Formel 8.} \quad W = (\mathbf{R} \times \text{vec}(\hat{\boldsymbol{\Pi}}) - \mathbf{r})' \left\{ \mathbf{R} [\widehat{\text{cov}}(\text{vec}(\hat{\boldsymbol{\Pi}}))] \mathbf{R}' \right\} (\mathbf{R} \times \text{vec}(\hat{\boldsymbol{\Pi}}) - \mathbf{r})$$

\mathbf{R} är en $(h \times m)$ matris av full rank, $\text{vec}(\hat{\boldsymbol{\Pi}})$ är en $(m \times 1)$ vektor med koefficienterna för VAR(P) och \mathbf{r} är en $(h \times 1)$ vektor med restriktioner. $\mathbf{R} \times \text{vec}(\hat{\boldsymbol{\Pi}}) = \mathbf{r}$ är sambandet som tillåter Wald-testet att undersöka koefficienter via flera olika regressioner. I följande ordning kommer de olika beståndsdelarna av den ovanstående ekvationen att beskrivas (Zivot & Wang, 2002, s. 388).

1. \mathbf{R}
2. $\text{vec}(\hat{\boldsymbol{\Pi}})$
3. Den estimerade kovariansen för $\text{vec}(\hat{\boldsymbol{\Pi}})$

Detta arbete har en \mathbf{r} -vektor som präglas av tal skilda från noll för koefficienterna för de förklarande variablerna samt alla de laggade värdena av y_5, y_6 d.v.s. Kurdistan och Pakistan för första regressionen y_1 . För regressionerna y_2, y_3, y_4, y_5, y_6 har \mathbf{r} -vektorn värdena noll för koefficienten för variabeln samt alla de laggade värdena av y_1 , m.a.o. log-avkastningarna för WTI-indexet.

R-matrisen

\mathbf{R} är en matris där parametrarna som ska undersökas bestäms och som för denna undersökningen endast innehåller ettor och nollor. Detta är en följd av att Granger-kausalteten mellan variablerna testas om koefficienterna för de förklarande variablerna i en viss regression är noll för alla laggade värden eller ej. När \mathbf{R} -matrisen multipliceras med koefficient vektorn $\text{vec}(\hat{\boldsymbol{\Pi}})$ blir produkten de parametrar som ska undersökas d.v.s. restriktionerna, vilket tillåter detta arbete att undersöka de olika nollhypoteserna. Fördelen med \mathbf{R} -matrisen är att den tillåter tester av olika samband och parametrar från de skilda regressionerna i VAR-modellen. \mathbf{R} -matrisen är en matris av full rank av antalet restriktioner h (Cameron & Trivedi, 2005, s. 224).

$\hat{\boldsymbol{\Pi}}$ -vektorn

VAR-modellen består av n antal regressioner som i sin tur har k antal parametrar var. För att kunna ta full hänsyn till alla parametrarna oberoende av vilken regression de tillhör samt att möjliggöra multiplikation med \mathbf{R} -matrisen fordras en vektorisering av $\boldsymbol{\beta}$ -vektorn för samtliga regressioner. Detta skapar en $(kn \times 1)$ vektor av den tidigare $(k \times n)$ matrisen, hädanefter kommer storleken på vektorn att benämnas $(kn=m)$ (Zivot & Wang, 2002, s. 387).

$$\text{Formel 9.} \quad \text{vec}(\hat{\boldsymbol{\Pi}}) = \begin{pmatrix} \vec{\beta}_1 \\ \vdots \\ \vec{\beta}_n \end{pmatrix}$$

Kovariansmatrisen för $\hat{\Pi}$

Den estimerade kovariansmatrisen för $vec(\hat{\Pi})$ har diagonal bestående av varianser för varje element av värdet i och ges av uttrycket nedan.

$$\text{Formel 10.} \quad \widehat{cov}(vec(\hat{\Pi})) = (\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\Sigma}$$

$$\text{Formel 11.} \quad \hat{\Sigma} = \frac{1}{T-k} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t' \quad \epsilon_t = y - \hat{\Pi}' \mathbf{Z}_t$$

\mathbf{Z} -matrisen är matrisen för variablerna och dess laggade värden vilket multipliceras med en transponerad \mathbf{Z} -matris som därefter inverteras. Den estimerade kovariansmatrisen är därav Kronecker produkten av den inverterade \mathbf{Z} produkten och OLS-estimatoren (Formel 11) (Zivot & Wang, 2002, s. 387).