Utah State University

DigitalCommons@USU

Undergraduate Honors Capstone Projects

Honors Program

Spring 2017

Regime Switching in Cointegrated Time Series

Bradley David Zynda II
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/honors

Part of the Mathematics Commons, and the Statistics and Probability Commons

Utah State University
MERRILL-CAZIER LIBRARY

REGIME SWITCHING IN COINTEGRATED TIME SERIES

by

Bradley David Zynda II

Thesis submitted in partial fulfillment
of the requirements for the degree

of

University Honors

in

Statistics
in the Department of
Mathematics and Statistics

Approved:

_____   _____
Thesis/Project Advisor   **Department Honors Advisor**
Dr. Tyler J Brough   Dr. David Brown

_____
Director of Honors Program
Dr. Kristine Miller

Utah State University
Logan, Utah

Spring 2017

Abstract

Volatile commodities and markets can often be difficult to model and forecast given significant breaks in trends through time. To account for such breaks, regime switching methods allow for models to accommodate abrupt changes in behavior of the data. However, the difficulty often arises in beginning the process of choosing a model and its associated parameters with which to represent the data and the objects of interest. To improve model selection for these volatile markets, this research examines time series with regime switching components and argues that a synthesis of vector error correction models with regime switching models will ameliorate financial modeling. Using futures prices from dairy markets as the chief data of interest, it will be shown that the traditional methods applied to these kind of series are not consistent and the need for a synthesis of models is needed.

Acknowledgements

Section I: Introduction

I.1

As econometrician John Gewecke notes, undoubtedly taking inspiration from the esteemed statistician George Box, "All econometric models are wrong, but some are useful." [1] A model, then, is never the truth, but something that seeks to best capture and approximate the underlying reality and phenomenon of interest. The task at hand, consequently, comes in identifying and capturing the useful from the useless. Despite expert admission that no model ever gets it perfectly right, models govern, guide, and influence profoundly important decisions that affect countless numbers of people. The focus of this capstone research will then be to advance methods in modeling concerning financial and econometric data. The course markets take hold high stakes for not only investors, hedge funders, and speculators, but also the millions of other individuals who find themselves caught in the economic crossfire which shapes day to day livelihoods. The choices these individuals make can be both captured and influenced by the models which act as a guide and map for all players in the economy. Good modeling, then, holds a profound effect for both private industry and public policy, and bad modeling can entail disaster. In this capstone, cointegration techniques developed by Engle-Granger will be used to explore the relationship between dairy futures by fitting an error correction model. From there, structural breaks will be explored using the Bai Perron test. Such test will act as a Bayesian prior to set up the possibility of modeling using vector error correction

---

[1] Geweke, John and William McCausland, *Bayesian Specification Analysis in Econometrics*, Oxford University Press, Agriculture and Applied Economics Association, Vol. 83, No. 5, Proceedings Issue (Dec, 2001): 1.

models with regime switching methods, should the presence of regimes changes or structural breaks be detected.

I.2

One particular market of interest for the sake of this paper will be dairy commodities, among others. The dairy market, specifically as it pertains to milk and cheese, constitutes complex relations characterized by high volatility. Cheese is, of course, made with milk. Furthermore, cheddar cheese, sold in 40-lbs blocks or 500-lbs barrells, is the most widely tracked dairy product on the market.[2] In other words, what happens with cheddar cheese affects many other markets as well. Not only is cheddar cheese used in raw form, but is also derived to make a host of other products as well. Consequently, information concerning the relationship of futures prices, such as that between cheese and the milk it is made from, can provide ample opportunity for speculators and hedgers alike.

The relationship between milk and cheese in the dairy market is elusive, however. Firstly, there have been no cheese futures until recently.[3] Since cheese is derived from milk, there clearly exists a relationship between the milk market and the cheese market. Whereas cheese did not have long term futures until recently in 2010, the Chicago Mercantile Exchange's Class III Milk Market the kind of milk used to make cheese is only sold in futures contracts.[4] One might suggest, then, that modeling could take into account these two markets and use them to aid in price discovery so as to facilitate financial planning and policy. However, a second problem comes into play. Both cheese spot and

---

[2] Bozic, Marin and T. Randall Fortenbery, *Creating Synthetic Cheese Futures: A Method for Matching Cash and Futures Prices in Dairy*, Journal of Agribusiness, Vol. 30, No. 2 (2012): 187.
[3] Ibid.
[4] Ibid.

futures prices are largely dictated by market action whereas the Class III milk futures

prices are set by the Federal Milk Marketing Order (FMMO). Governmental regulation of

milk futures allow for a subsidy of dairy farmers, effectively setting a price floor. These

prices are determined by market prices of other cheese products.[5] That is, officials from the

FMMO announce the following month's futures prices for milk futures based on the market

prices of dairy goods from the month prior.

Specifically, Class III Milk prices are determined by a formula created by the FMMO

in the following way[6]:

Class III Price = (Class III skim milk price x 0.965) + (Butterfat price x 3.5).
Class III Skim Milk Price = (Protein price x 3.1) + (Other solids price x 5.9).
Protein Price = ((Cheese price - 0.1702) x 1.405) + ((((Cheese price - 0.1702) x
1.582) - Butterfat price) x 1.28).
Other Solids Price = (Dry whey price - 0.137) divided by 0.968.
Butterfat Price = (Butter price - 0.114) divided by 0.82.

Class III Milk prices, as can be seen, are set by the prices of cheese, and cheese is used to

make milk. Consequently, these two markets may mutually affect one another. Vector

autoregression and vector error correction models using cointegration may help uncover

the relationship of these commodities.

I.3

Principally, this research will test for cointegration between these milk and cheese

futures. Such testing will be done through the Engle-Granger two-step method for which

Clive Granger and Robert Engle won the Nobel Prize[7]. Cointegration may provide key

---

[5] Ibid.

[6] Gould, Brian W, *Pricing Formulas*, University of Wisconsin Dairy Marketing and Risk Management Program.

[7] Engle, Robert F. and Clive Granger, *Co-Integration and Error Correction: Representation, Estimation, and Testing*,Econometrica, Vol. 55, No. 2 (Mar., 1987).

economic information for these two markets as it explains and predicts the common stochastic path two random walk processes follow.[8] If cheese and milk futures are not cointegrated, there may be opportunity for arbitrage. For example, if the price of cheese futures significantly exceed that of milk futures, the opportunity to buy more milk at a lower price and sell it to make cheese at a higher price presents itself as an opportunity for profit.

From the pricing information supplied by the USDA above and created by the FMMO, cheese prices on the market are partly responsible for the next month's milk futures prices. Consequently, it will be expected that these two commodities influence one another's prices. When one hits a low, it seems reasonable to expect the other to go low as well and vice versa. Though they both may follow stochastic trends as a random walk process, one might speculate that they always maintain the same distance apart. That is, they move randomly in the same direction. It seems reasonable then to fit these data using a VAR and VECM model with an eye turned towards cointegration.

At its core, cointegration seeks uncover the possible common trends of two random walk models. Although two or more time series processes may be following a purely stochastic trend, they all may be following roughly the same stochastic trend, as alluded to above. The path of these prices or observations through time can be analogous with a drunk and her dog.[9] Although an intoxicated person may wander and stumble in a seemingly random path, the pet dog will never stray too far from his drunken owner.

---

[8] Wooldridge, Jeffrey M., *Introductory Econometrics: A Modern Approach, Fourth Edition*, (Mason: South-Western, 2009), 623-667.

[9] Murray, Michael P., *A Drunk and Her Dog: An Illustration of Cointegration and Error Correction,* The American Statistician, Vol. 48, No. 1 (Feb., 1994): 37-39.

Cointegrating processes acts much in the same way. The prices or returns of two markets which are cointegrated never wander to far from one another despite taking random turns along the way.

Cointegration is a part of a greater part of modeling using vector error correction models. This model, abbreviated as VECM, is a special kind of vector autoregression model (VAR) with an error correction component. The VAR is a multivariate autoregressive model. The VECM adds an error correction component that accounts for short-term dynamics between multivariate time series. The parameters of this error component can be estimated through the Engle-Granger two step method. First, one tests the series for first order integration, or to see if it is *I(1)* model. In other words, stationarity is tested after differencing the random walk model once. By regressing one *I(1)* series against the other, one can test the residuals of this linear combination of two series for stationarity. If the residuals follow a stationary process, the series are cointegrated. These residuals account for the short term disequilibrium between the two series by capturing deviations from the long term equilibrium.[10] The error correction component of the VECM does exactly this. With the general VAR, the stable, long term relationship is modeled. The error correction component, formed by the residuals or deviations from the long run trend, act as a means to account for the short term departures from the long run equilibrium.

Modeling markets in this way, and particularly dairy markets, can also allow for simulation and forecasting. Because VAR and VECM models are, as the name implies, autoregressive, the underlying model with its associated parameters can be hit with

---

[10] Alexander, Carol, *Market Models: A Guide to Financial Data Analysis*, (West Sussex, 2001): 351.

various shocks and analyzed with regards to how the future prices and volatility measure out. In other words, since these autoregressive models are in part deterministic, one can roll the model forward in time given a set of initial parameters and conditions.

The VECM model is also desirable given its allowance for learning effects to take place. In Austrian literature, and specifically the thought of Hayek, prices represent information concerning people's values which take time disperse through the economy. He writes:

> " The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess."[11]

In this way, the VECM allows for the "dispersed bits of knowledge" time to come together into full equilibrium. It is the adjustment for the short term deviations from the long term equilibrium. These short term deviations can be understood as the time it takes knowledge in the form of prices to be harnessed and shared.

Among other things, this cointegration process utilizes the Dickey-Fuller test for unit roots. However, presence of changes in regimes may obfuscate this clear-cut two-step modeling method pioneered by Engle-Granger. Although the Dickey-Fuller test may yield a small p-value and thereby reject the null hypothesis that a unit root is present and consequently indicate that the series are cointegrated, regime switches may turn this result into a Type I error.

---

[11] Hayek, Fredrich A, *The Use of Knowledge in Society*, The Library of Economics and Liberty, (1945).

These regime switches, detected endogenously by the Bai Perron Test[12], may correspond with many different factors affecting a market. In particular, policy changes, insofar as they pertain to dairy subsidies by the USDA, may be of interest as a particular source of volatility and regime breaks in the market of this research. The Bai Perron test will act as a Bayesian specification prior where it will be assumed with high confidence that there are breaks in regimes for this dairy data given the government's establishment of price floors in this market. Regime changes often correlate with changes in regulation and policy.[13] If this is the case, a simple linear cointegration may not be well suited for future forecasting and simulation of this particular data. That is, the presence of regime changes may be indicative of the need of better modeling methods or the need to improve the VECM. If such a model can be later built that incorporates regime switching into the vector error correction model using, among other things, Markov chain monte carlo methods, then the effect of the government's policy in the volatility of the dairy market may be better ascertained, not to mention other forms of forecasting and simulation as well.

Section II: Methods

Data for this project is gathered from real prices of Class III Milk exchange futures and Cash Settled Cheese futures from September 3rd, 2010 until March 15th, 2017 sold on the Chicago Mercantile Exchange. Prices taken at opening are used. VECM and VAR models use differencing between observations, and consequently the day-to-day returns are used.

---

[12] Bai, Jushan and Pierre Perron, *Estimating and Testing Linear Models with Multiple Structural Changes*, Econometrica, Vol. 66, No. 1 (Jan., 1998): 47-78.

[13] Ang, Andrew and Allan Timmermann, *Regime Changes and Financial Markets*, Annual Review of Financial Economics, Vol. 4 (2012): 314.

Furthermore, using a natural log transformation of the price differences tends to be more algebraically useful, simple, and elegant.

Time Series analysis constitutes a very different game than traditional statistics using controlled experiments or observational studies. Whereas the latter can afford many observations at one time, the former only gets one observation at one period of time and no more. Such is the case with the prices of the market of interest in this research. Typically time series begins with assuming that a particular model may follow some kind of deterministic trend with a stochastic component:

$$y_t = a + b_t + e_t$$

From there, the model will have to be tested for stationarity and differencing. In the context of this analysis, it is standard in practice to assume that prices, rate, and yield data are not stationary but rather integrated of order 1 or *I(1)* when implementing the Engle-Granger two-step cointegration model.[14] Consequently, it will be useful if both time series of milk and cheese futures can be found to be a random walk where each new observation at time *t* is only a result of the previous observation with the current error term e_t such that

$$y_t = y_{t-1} + \epsilon_t$$

For both the milk and cheese futures, model estimates will be made using Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). This will be facilitated through an R package `forecast` using the `auto.arima()` function. An ARIMA(0,1,0)

---

[14] Alexander, Carol, Market Models: A Guide to Financial Data Analysis, (West Sussex, 2001): 348.

process is indeed a random walk. Should the function favor this model using AIC/BIC methods above, it will then be assumed that both milk and cheese futures prices follow a random walk.

Differencing the random walks to form an *I(1)* model will help in beginning the Engle-Granger two step procedure since it constitutes the first step. Forming the two *I(1)* series, as described above, as a linear combination will be the final step in the cointegration process. This linear combination is formed as an OLS regression model using R's function `lm()` taking the form:

$$x_t = c + \alpha y_t + \epsilon_t$$

Using the augmented Dickey-Fuller test, with help of the R function `adf.test()` in the tseries package will test this regression model of *I(1)* series for stationarity. There is a cointegrating relationship between two series only if the residuals of the linear model are a stationary process. This stationary linear combination is like glue which keeps codependency between the two series.[15]

After testing for first order differencing, examining the relationship between the milk and cheese returns will be key. That is, it will be important to see to what extent previous returns of milk or cheese influence the return at some point in time *t*. Testing for Granger causality will allow for a statistical evaluation of the significance of the influence of milk on cheese returns and cheese on milk returns. Granger causality seeks to test whether there is a lead-lag relationship between variables in a multivariate time series.[16] So, if milk returns Granger cause cheese returns, then the returns of cheese today are determined in

[15] Alexander, Carol, *Market Models: A Guide to Financial Data Analysis*, (West Sussex, 2001): 351.
[16] Alexander, Carol, *Market Models: A Guide to Financial Data Analysis,* (West Sussex, 2001): 344.

part by the returns of milk yesterday or another time in the past. This idea is closely related to the concept of cointegration, but the latter signifies that the series follow a common stochastic trend as well.

A VAR and VECM model will be fitted to the milk and cheese futures returns to compare each's ability to capture the behavior of the series. This can be done again with the help of the `lm()` model in R as the difference of one series can be formed in a linear combination of the lag of the difference of the other series and itself with a constant and error term:

$$\Delta y_t = \alpha_1 + \Delta y_{t-1} + \Delta x_{t-1} + \epsilon_{1t}$$
$$\Delta x_t = \alpha_2 + \Delta x_{t-1} + \Delta y_{t-1} + \epsilon_{2t}$$

Using OLS estimates, the parameters can be found and the significance evaluated. The VECM looks much the same, but with the added error correction component:

$$\Delta y_t = \alpha_1 + \Delta y_{t-1} + \Delta x_{t-1} + \gamma_1 z_{t-1} + \epsilon_{1t}$$
$$\Delta x_t = \alpha_2 + \Delta x_{t-1} + \Delta y_{t-1} + \gamma_2 z_{t-1} + \epsilon_{2t}$$

To facilitate coding, the `tsDyn` package was used to estimate and fit a VECM using OLS again and the Engle-Granger two step method. Using significance tests, it can be determined if each parameter and predictor are relevant and hold explanatory power for each series.

Once the models are fitted and evaluated, attention will be turned to potential structural breaks in regime. These regimes are periods where the series is characterized by completely new parameters. In the presence of regime switching, the regime of one set of

13

estimated parameters characterizing the model break and jump into a different regime with different parameters. Such breaks will be endogenously estimated using the Bai-Perron Test. Bai and Perron were among the first to develop techniques allowing for analysis of multiple structural breaks in time series. Such breaks are treated as unknown variables which are detected by minimizing the sum of square residuals in the overall model. Consequently, these breaks are learned from the data endogenously as opposed to traditional exogenous methods.[17] Using the R package `strucchange`, structural breaks in the dairy data can be estimated using the methods developed by Bai and Perron.

The break points given by the Bai-Perron Test will act as a Bayesian prior for future modeling. In other words, it will constitute part of the specification analysis element of Bayesian modeling in the future. Specification analysis in this way will seek to use the speculated model to map predictions for data before any data is actually observed. Afterwards, the predictions can be compared with real data to assess its accuracy. In a similar way, after a model is fitted with real data, one could compare a simulation replication something close to the original data to evaluate performance.[18]

If the breakpoints exist in these time series and are substantial, then linear-Gaussian assumptions used in the OLS regression modeling may yield inaccurate results and lose its explanatory power. Forming a VECM using Bayesian methods with a prior distribution of regime breaks will be no easy process, but may yield better predictive

[17] Perron, Pierre, Dealing with Structural Breaks. Palgrave Handbook of Econometrics, Vol. 1, (2005): 8.
[18] Geweke, John and William McCausland, *Bayesian Specification Analysis in Econometrics*, Oxford University Press, Agriculture and Applied Economics Association, Vol. 83, No. 5, Proceedings Issue (Dec, 2001): 1.

accuracies. Using MCMC methods developed by Koop et al[19], it is possible to do and will be

implemented in future work that this research begins.

Section III: Analysis

Figure 1: Times Series Plot of Milk Futures Prices

Milk Futures Prices

---
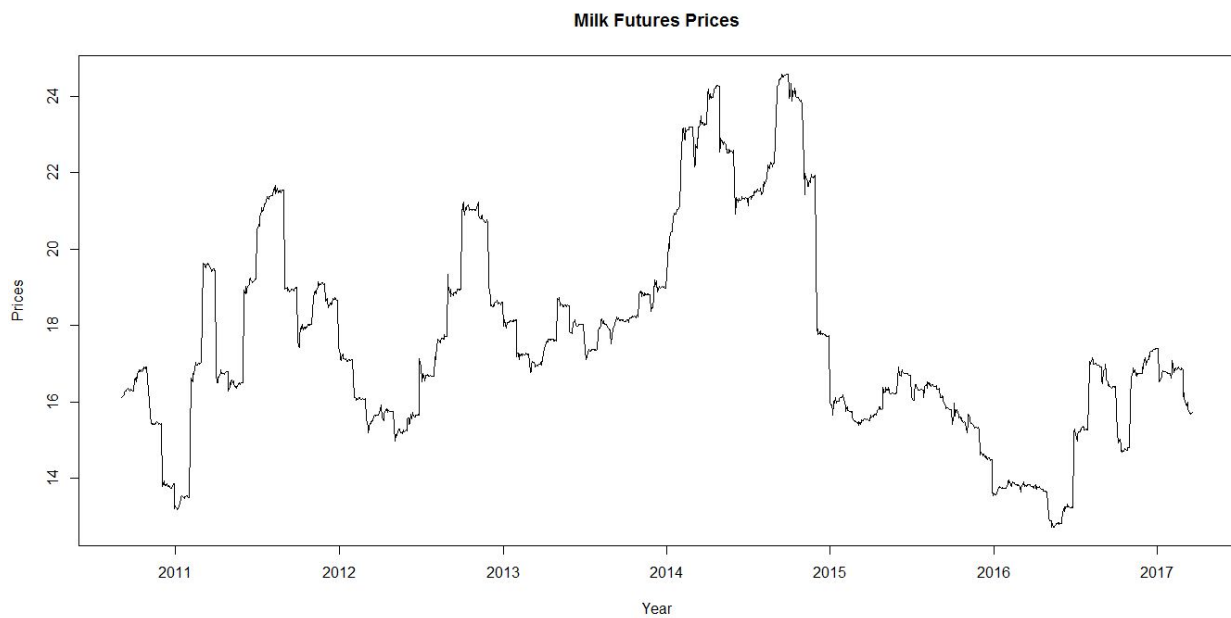
[19] Jochmann, Markus and Gary Koop, *Regime-switching cointegration*, Studies of Nonlinear Dynamics and Econometrics, Vol. 19, No. 1, (Feb., 2015).

Figure 2: Times Series Plot of Cheese Futures Prices

Cheese Futures Prices
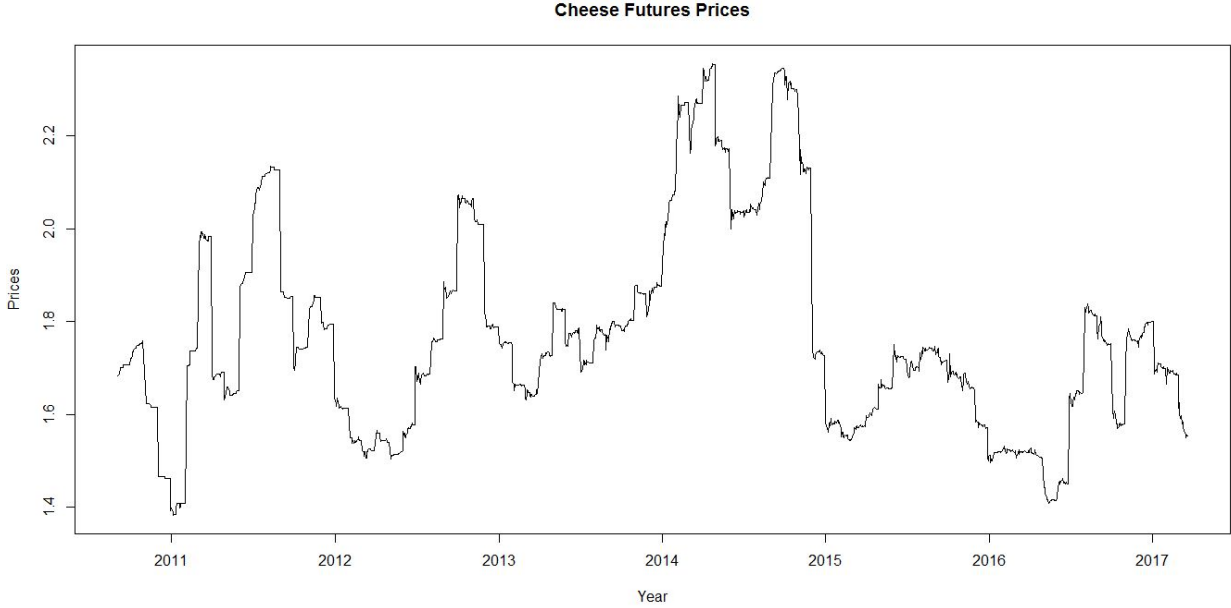
Random Walk Models

When executing the Engle Granger two step methods, one of the initial considerations is that both time series processes follow a random walk and are hence non-stationary. Consequently, they ought to be integrated to order 1 such that they are stationary. First, the PACF and ACF plots are displayed:

Figure 3: ACF and PACF Plot of Milk Futures
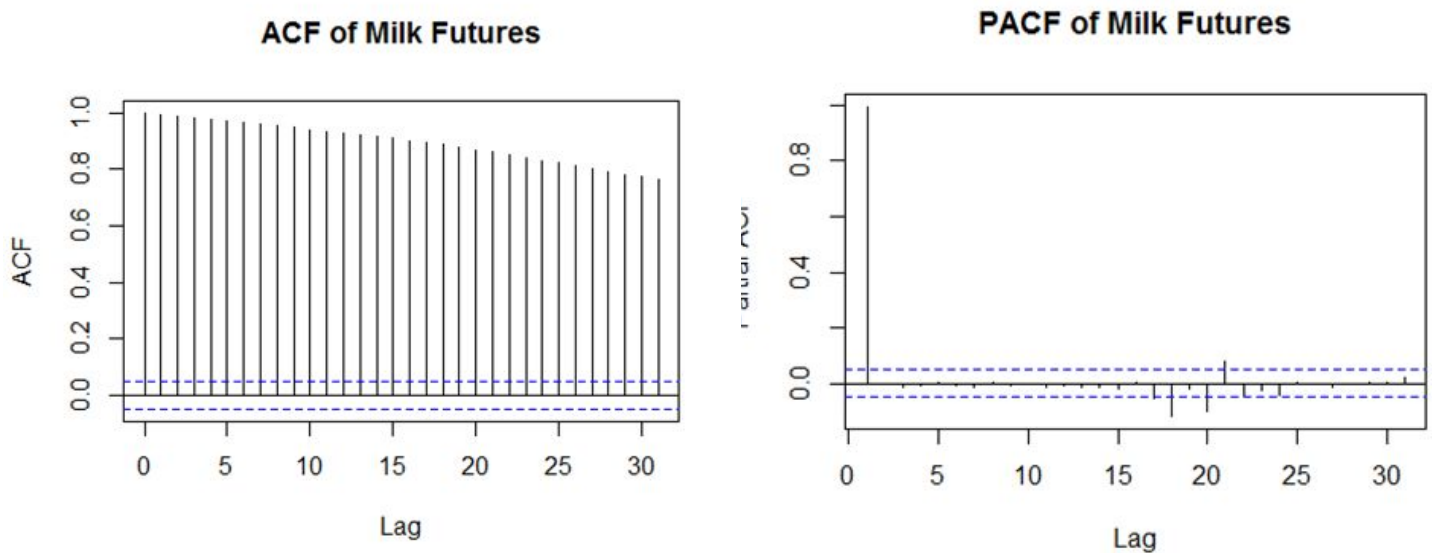
As can be seen, the PACF suggests that there is one lag and that an AR(1) model may be the best way to plot this time series for milk. It is *AR* signature, so to speak. Furthermore, this idea may suggest why so many significant lags showed up in the ACF - they were really all explained by one significant lag in the PACF..

With help of the `auto.arima()` function in the `forecast` package, a model estimate based on AICc can be generated. The milk price data from September 2010 to

March 2017 is contained in the time series object `milk_ts`, and `cheese_ts` for the cheese prices.

```
auto.arima(milk_ts)
## Series: milk_ts
## ARIMA(0,1,0) with drift
##
## Coefficients:
##          drift
##         -0.0002
## s.e.     0.0071
##
## sigma^2 estimated as 0.08035:  log likelihood=-248.96
## AIC=501.91    AICc=501.92    BIC=512.64
```

As can be seen, this functions recommends an ARIMA(0,1,0) model, and not an AR(1) model. This effectively suggests that the milk data be modeled by:

$$y_t = \alpha y_{t-1} + \epsilon_t$$

where

$$\alpha = 1$$

Which is to say that each new observation is solely the result of an error term or white noise. In other words, the milk data constitutes a random walk.

Random walk models are clearly not stationary as there is a unit root present. However, differencing by one will result in the model becoming stationary as a white noise process. That it is a white noise process can be evidenced again by `auto.arima()`:

```
auto.arima(diff(milk_ts, 1))
## Series: diff(milk_ts, 1)
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##          intercept
##           -0.0002
## s.e.       0.0071
```

18

```
##
## sigma^2 estimated as 0.08035:  log likelihood=-249.11
## AIC=502.23   AICc=502.24   BIC=512.96
```

And an ARIMA(0,0,0) model is a white noise model:

$$y_t - y_{t-1} = \epsilon_t$$

To confirm that differencing by 1 is sufficient, the `ndiffs()` function estimates how

many differencing terms are needed. This function can be set to use the Augmented Dickey

Fuller test to yield a differencing term, among other methods. For the milk data,

differencing by 1, or since it is a random walk integrating by order 1, is sufficient according

to `ndiffs()`. To further confirm that this is an I(1) process, the ADF test will be

performed outright:

```
ndiffs(milk_ts)
## [1] 1
adf.test(diff(milk_ts, 1), alternative = "stationary")
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(milk_ts, 1)
## Dickey-Fuller = -10.8844, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```

The null hypothesis that the series is non-stationary is rejected given the significant

p-value. It can then be concluded that the the milk futures constitute a random walk *I(1)*

process.

The cheese futures constitute the exact same process as the milk. These futures

prices, too, are a *I(1)* random walk process.

```
auto.arima(cheese_ts)
## Series: cheese_ts
```

```
## ARIMA(0,1,0) with drift
##
## Coefficients:
##         drift
##         -1e-04
## s.e.    7e-04
##
## sigma^2 estimated as 0.0007492:  log likelihood=3435.1
## AIC=-6866.19    AICc=-6866.18    BIC=-6855.46
adf.test(diff(cheese_ts, 1))
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(cheese_ts, 1)
## Dickey-Fuller = -10.7963, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```

As mentioned above, the cheese futures also are best modeled with an ARIMA(0,1,0) or a

random walk after estimation. Furthermore, differencing once results in stationarity. Hence

the cheese futures are also an *I(1)* process.

Differencing by order 1 in the two futures above is that same as setting the time

series in terms of returns instead of prices. In finance, doing logarithmic transformations

on returns tends to be the norm. When such log transform is applied below, the outcomes

outlined above for both cheese and milk are the same.

```
auto.arima(log(milk_ts))
## Series: log(milk_ts)
## ARIMA(0,1,0) with drift
##
## Coefficients:
##         drift
##         0e+00
## s.e.    4e-04
##
## sigma^2 estimated as 0.0002481:  log likelihood=4306.12
## AIC=-8608.23    AICc=-8608.22    BIC=-8597.5
adf.test(diff(log(milk_ts), 1), alternative = "stationary")
## Warning in adf.test(diff(log(milk_ts), 1), alternative =
## "stationary"): p-value smaller than printed p-value
```

```
## 
##  Augmented Dickey-Fuller Test
## 
## data:  diff(log(milk_ts), 1)
## Dickey-Fuller = -10.9225, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```

Figure 4: Log Milk Future Returns from 2011 until 2017
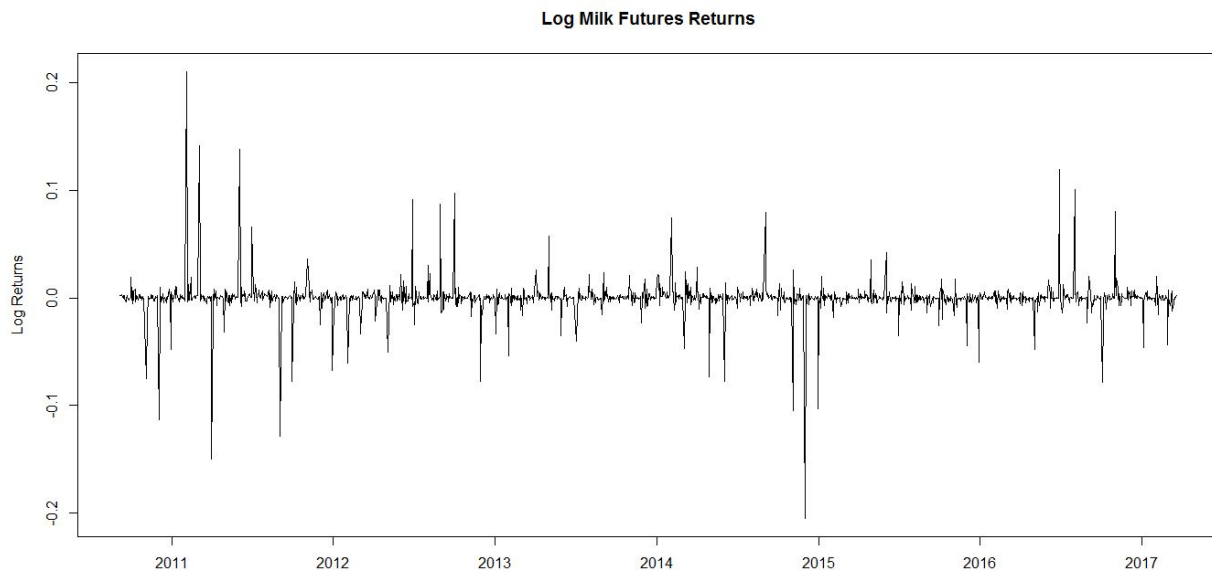

Log Milk Futures Returns

```
auto.arima(log(cheese_ts))
## Series: log(cheese_ts)
## ARIMA(0,1,0) with drift
## 
## Coefficients:
##        drift
##        -1e-04
## s.e.   4e-04
## 
## sigma^2 estimated as 0.0002277:  log likelihood=4373.5
## AIC=-8743.01   AICc=-8743   BIC=-8732.28
adf.test(diff(log(cheese_ts), 1), alternative =
"stationary")
## Warning in adf.test(diff(log(cheese_ts), 1), alternative
= "stationary"):
## p-value smaller than printed p-value
## 
##  Augmented Dickey-Fuller Test
```

21

```
##
## data:  diff(log(cheese_ts), 1)
## Dickey-Fuller = -10.8382, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```
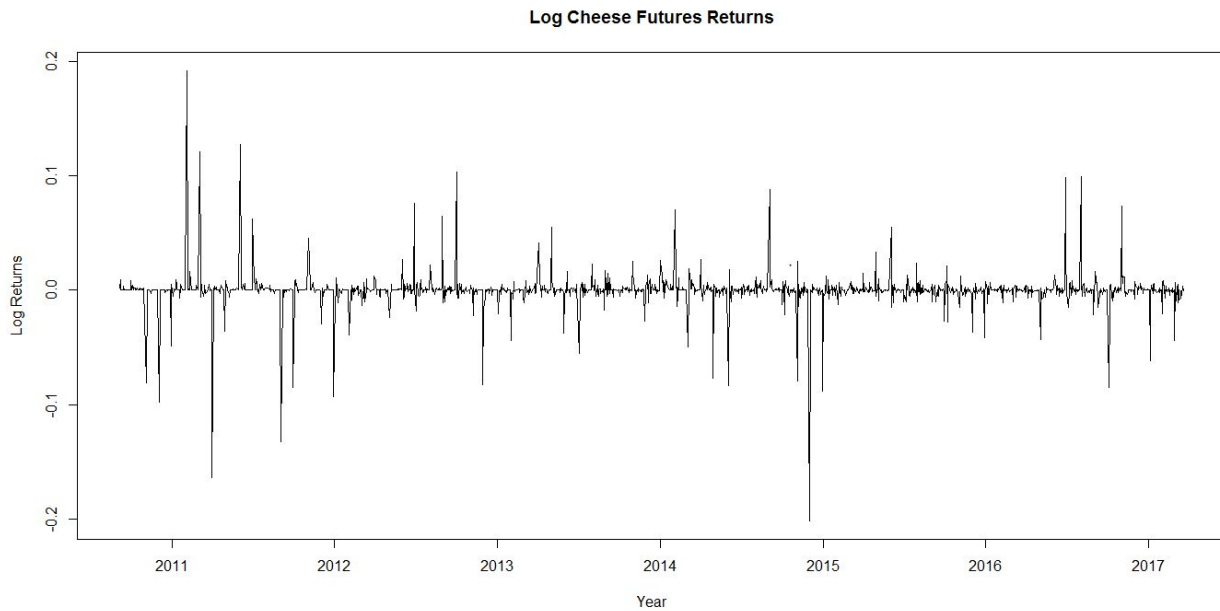
Figure 5: Log Cheese Future Returns from 2011 until 2017

In both cases, the log transforms of the futures returns for both milk and cheese follow a random walk and are *I(1)*. Based on AICc values, an ARIMA(0,1,0) process was recommended just as it was before. Furthermore, plotting both the milk and cheese series after differencing as in Figures 4 and 5 shows a stationary trend, albeit a white noise process.

Cointegration

The Engle-Granger two step method for cointegration first assumes that both series in question follow a random walk model. This has been shown above. The models can be then be formed into a linear combination of one another and regressed on one another such that:

$$x_t = c + \alpha y_t + \epsilon_t$$

Where x_t and y_t are the two random walks. Hence the milk and cheese futures below are formed into a linear combination such that the error terms then become mean reverting. If the error terms, or residuals, of this regression of cheese on milk returns fit normality assumptions and can be shown to constitute a stationary model (that is the model is mean-reverting) then the two series can be said to be cointegrated. One hint of cointegration comes in examining how closely the log milk and cheese prices follow a common stochastic trend. The plot below is indicative of a near identical path.

Figure 6: Milk and Cheese Futures Prices

Log Prices of Dairy Commodities
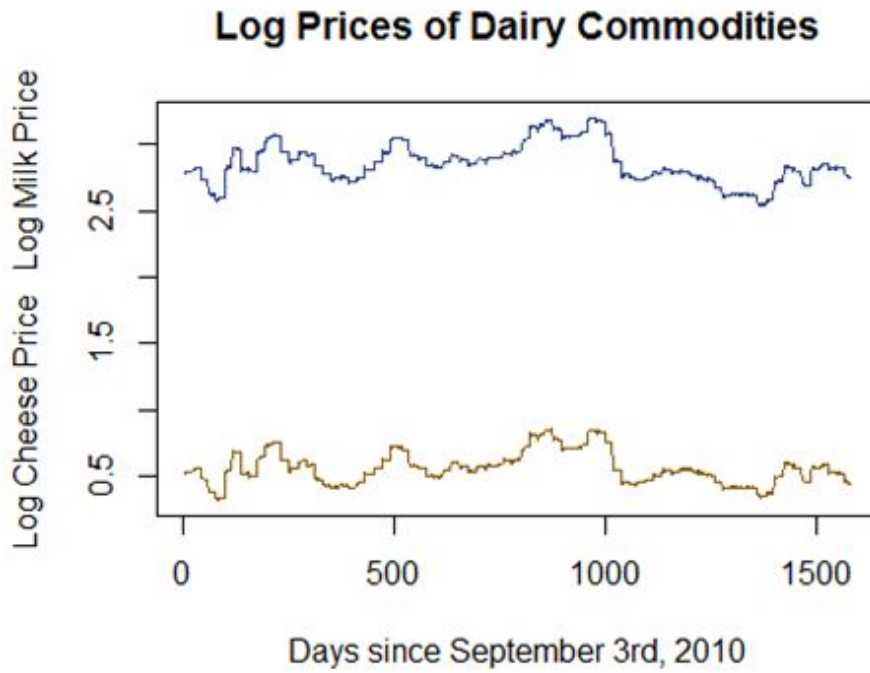
Regressing log milk returns against log cheese returns gives the following results below.

Figure 7: Regression of log Milk and Cheese returns
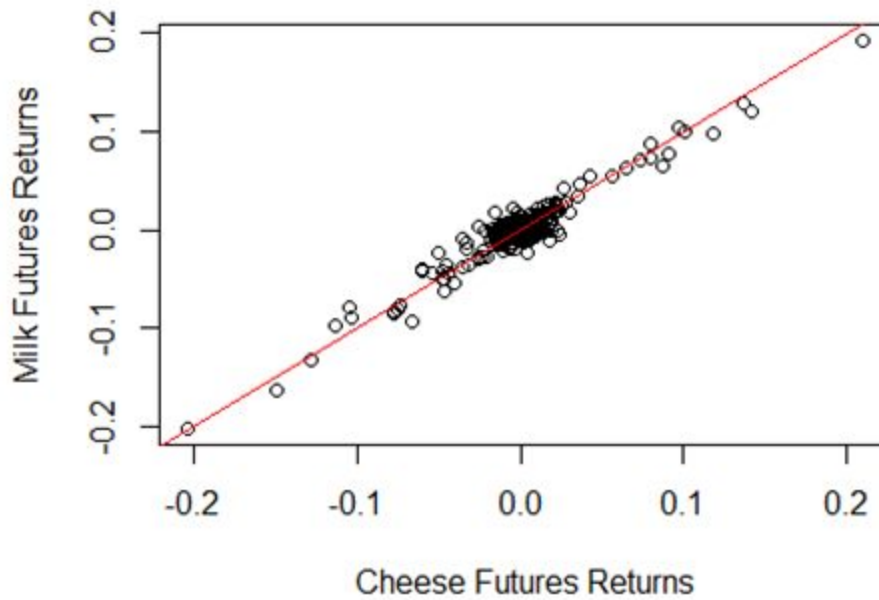
There is good visual evidence that this regression maybe a very good fit. However, most of

the data are accumulated around 0.0 instead of following a long, smooth trend. This may be

indicative of misleading results.

```
## Call:
## lm(formula = d1.log.milk ~ d1.log.cheese)
##
## Residuals:
##     Min      1Q     Median      3Q     Max
## -0.032850 -0.001813 -0.000035  0.001860  0.028710
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.462e-05  1.237e-04     0.28    0.78
## d1.log.cheese 9.917e-01  8.198e-03   120.96  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004912 on 1575 degrees of freedom
## Multiple R-squared:  0.9028, Adjusted R-squared:  0.9028
## F-statistic: 1.463e+04 on 1 and 1575 DF,  p-value: < 2.2e-16
```

Based on the output above, OLS estimates gives a highly significant β_1 value. Furthermore,

R-squared value is quite high at .9028, indicating a good fit. A Normal Q-Q Plot tells a bit of

a different story, however:

Figure 8: Normal Q-Q Plot of Residuals
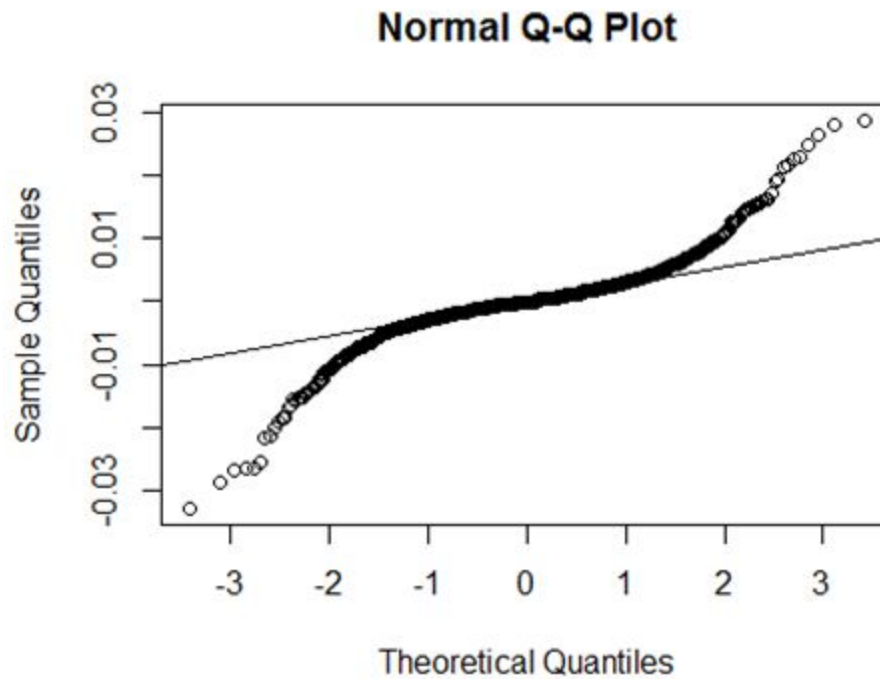
Although skewness is relatively low at -0.1713, the kurtosis is quite high at 8.7518, as evidenced by the fat tails in the plot above. This could affect some of the underlying assumptions of this linear model.

The residuals of this plot can be defined as:

$$z_t = \beta_1 x_t + y_t$$

Since β_0 is not significant it is left out. The residuals can then be defined as a time series process and tested for stationarity:

$$z_t = z_{t-1} + \epsilon_t$$

This is done using, again, the ADF Test.

```
##  Augmented Dickey-Fuller Test
##
## data:  milk.lm$residuals
## Dickey-Fuller = -38.2617, Lag order = 1, p-value = 0.01
## alternative hypothesis: stationary
```

The highly significant p-value provides evidence against the null hypothesis of nonstationarity, and it is concluded that the residuals follow a stationary *I(0)* process. This is the second step of the Engle Granger method, and it can be concluded that the milk and cheese futures are cointegrated. That is, prices of one commodity may be in part determined by previous prices and trends of the other commodity.

Granger Causality

Using the R package `lmtest`, the function `grangertest()` can help discover the presence of Granger causality between two or more time series. The results can be found in the table below:

Table 1: Results of test for Granger Causality

Granger Causality Test			
Dependent	Independent	F-Statistic	P-value
Cheese	Milk	0.5734	0.449
Milk	Cheese	23.031	<0.0001

General VAR

To demonstrate the importance of cointegration, a general vector autoregressive

(VAR) model will be fit to the log milk and cheese returns and later compared to that of a

vector error correction model (VECM). As described before, a VECM is a special kind of VAR

with an added error correction component which is formed by the error terms of the

regression of one future on another. This VAR model below will assume no significant

deviations from long term trends, whereas the VECM would help correct those short term

deviations.

Below, a linear model is formed by regressing log milk returns against lagged log

milk and cheese returns. Then, the same is done but this time switching the dependent

term from milk to cheese:

```
milk.var <- lm(lead.d1.log.milk ~ l1.d1.log.milk +
l1.d1.log.cheese)
cheese.var <- lm(lead.d1.log.cheese ~ l1.d1.log.cheese +
l1.d1.log.milk)
```

Next, the significance of each term for the milk return equation are evaluated:

```
summary(milk.var)
##
## Call:
## lm(formula = lead.d1.log.milk ~ l1.d1.log.milk +
l1.d1.log.cheese)
##
## Residuals:
##       Min          1Q    Median       3Q         Max
## -0.204068 -0.001520  0.000168  0.001891  0.208932
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.989e-06  3.942e-04  -0.008    0.994
## l1.d1.log.milk   -3.619e-01  8.030e-02  -4.507 7.05e-06 ***
```

28

```
## l1.d1.log.cheese  4.022e-01  8.381e-02   4.799 1.75e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1
##
## Residual standard error: 0.01565 on 1573 degrees of freedom
## Multiple R-squared:  0.01445,    Adjusted R-squared:
0.01319
## F-statistic: 11.53 on 2 and 1573 DF,  p-value: 1.069e-05
```

Based on the output above, although the intercept term is not significant, the lagged milk

and cheese terms are. This first part of this VAR model seems to signify that milk returns

are explained by previous prices of milk and cheese.

```
summary(cheese.var)
##
## Call:
## lm(formula = lead.d1.log.cheese ~ l1.d1.log.cheese +
l1.d1.log.milk)
##
## Residuals:
##        Min       1Q    Median       3Q        Max
## -0.201619 -0.001006  0.000053  0.001467  0.192167
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -5.067e-05  3.804e-04  -0.133 0.894
## l1.d1.log.cheese  1.995e-02  8.086e-02   0.247 0.805
## l1.d1.log.milk    3.800e-03  7.748e-02   0.049  0.961
##
## Residual standard error: 0.0151 on 1573 degrees of
freedom
## Multiple R-squared:  0.0005642,  Adjusted R-squared:
-0.0007066
## F-statistic: 0.444 on 2 and 1573 DF,  p-value: 0.6416
```

Unlike milk, cheese does not seem to have any significant predictors, and this model does

not seem to have much explanatory power at all in regards to cheese returns. It is also

important to point out that there does seem to be high correlation among the predictors:

Table 2:

Covariance of Matrix Residuals		
	Log Cheese Returns	Log Milk Returns
Log Cheese Returns	0.0002280	0.0002261
Log Milk Returns	0.0002261	0.0002449

Table 3:

Correlation Matrix of Residuals		
	Log Cheese Returns	Log Milk Returns
Log Cheese Returns	1.00	.9565
Log Milk Returns	.9565	1.00

VECM

Now, attention is turned to fitting a VECM. This process can be done in much the same way

as above by forming a linear model of each of the series. However, one now adds on the

residuals found in forming a linear combination of the milk and cheese futures in the

second part of the Engle-Granger Test. The R package `tsDyn` allows for ease in fitting a

VECM in this way using the Engle-Granger method and OLS estimates.

```
vecm <- (lineVar(data.frame(cbind(d1.log.cheese,
d1.log.milk)), lag=1, r=1, model="VECM", estim="2OLS"))

summary(vecm)
## #############
## ###Model VECM
## #############
## Full sample size: 1577    End sample size: 1575
## Number of variables: 2   Number of estimated slope
parameters 8
## AIC -29565.33    BIC -29517.07   SSR 1.082727
## Cointegrating vector (estimated by 2OLS):
##     d1.log.cheese d1.log.milk
## r1              1  -0.9103948
##
##
##                           ECT                  Intercept

## Equation d1.log.cheese -1.4611(0.1683)***
-5.7e-05(0.0005)
## Equation d1.log.milk    0.0875(0.1795)
2.4e-06(0.0005)
##                           d1.log.cheese -1   d1.log.milk -1
## Equation d1.log.cheese 0.2900(0.1028)**
-0.7107(0.0943)***
## Equation d1.log.milk    0.1583(0.1096)
-0.6440(0.1005)***
```

The output is a bit tricky to read. In the table below, the values of the estimated parameters

for each predictor as well as its corresponding p-value can be seen and compared with that

of the VAR model.

Table 4: VAR and VECM Model Results

Model Fit Comparison									
Mode	Dependent	Intercept	P-value	Lag Milk	P-value	Lag Cheese	P-value	Error Correction	P-value
VAR	Log Milk Returns	-2.989E-06	0.994	-0.3619	<.0001	0.4022	<.0001	N/A	N/A
	Log Cheese Returns	-0.00005067	0.894	0.0038	0.961	0.01995	0.805	N/A	N/A
VECM	Log Milk Returns	0.0000024	>.10	-0.644	<.0001	0.1583	>.10	0.0875	>.10
	Log Cheese Returns	-0.000057	>.10	-0.7107	<.0001	0.29	<.001	-1.4611	<.0001

The difference is rather striking. The VAR model indicates that milk returns are highly dependent on previous milk and cheese returns and the cheese has no significant predictors. In the VECM, it looks quite the opposite. It is the cheese returns which have the significant predictors of previous milk and cheese returns, whereas milk is explained in part by previous milk returns. Also noteworthy is the highly significant estimate for the error correction component on cheese. It seems, as indicated by initial testing, that a VECM is more appropriate than a standard VAR.

Regime Changes

Despite fitting a VECM model, there may be a better ways of modeling these two dairy futures. As can be seen in Figure 1 and Figure 2, the milk and cheese futures seems to go through distinct periods of highs and lows. As such, looking for breaks in regime may be sensible. This test will act as a factor in subsequent research where modeling a VECM will take into account the distinct changes in regimes.

Here, the original milk and cheese time series will be examined for structural breaks in regimes using the Bai-Perron Test in the R package `strucchange`. This method looks within the data and estimates breakpoints by using the Bayesian Information Criterion (BIC) and residual sum of squares (RSS). As it goes from point to point, it examines how

much the BIC and RSS can be minimized as it incorporates a potential break point in

regime. In doing this, it does not go to each point sequentially, but rather first incorporates

the break with the greatest lowering of BIC and RSS instead of the earliest date to occur as

can be seen in the code below in the case of milk:

```
bp.milk <- breakpoints(milk_ts~1); bp.milk
##
##   Optimal 5-segment partition:
##
## Call:
## breakpoints.formula(formula = milk_ts ~ 1)
##
## Breakpoints at observation number:
## 469 779 1015 1251
##
## Corresponding to breakdates:
## 469 779 1015 1251
summary(bp.milk)
##
##   Optimal (m+1)-segment partition:
##
## Call:
## breakpoints.formula(formula = milk_ts ~ 1)
##
## Breakpoints at observation number:
##
## m = 1                  1034
## m = 2              779 1015
## m = 3          469 779 1015
## m = 4          469 779 1015 1251
## m = 5    236 472 779 1015 1251
##
## Corresponding to breakdates:
##
## m = 1                  1034
## m = 2              779 1015
## m = 3          469 779 1015
## m = 4          469 779 1015 1251
## m = 5    236 472 779 1015 1251
##
## Fit:
##
```

```
## m    0   1     2    3    4     5
## RSS 11712  8008  4096  3766  3635  3635
## BIC  7656  7071  6028  5910  5869  5883
```

Table 5: Dates of Breaks in Regimes with 95% Confidence Limits

Milk Breakpoints			
	2.5%	Breakpoint	97.5%
1	8/20/2012	8/29/2012	10/8/2012
2	12/2/2013	12/4/2013	12/5/2013
3	11/26/2014	11/28/2014	12/03/2014
4	10/9/2014	11/20/2015	11/25/2015

Additionally, the plot below graphically demonstrates the process of the test as it minimizes BIC and RSS as it seeks breaks in regime:

Figure 9: BIC and RSS Comparisons for Milk

BIC and Residual Sum of Squares
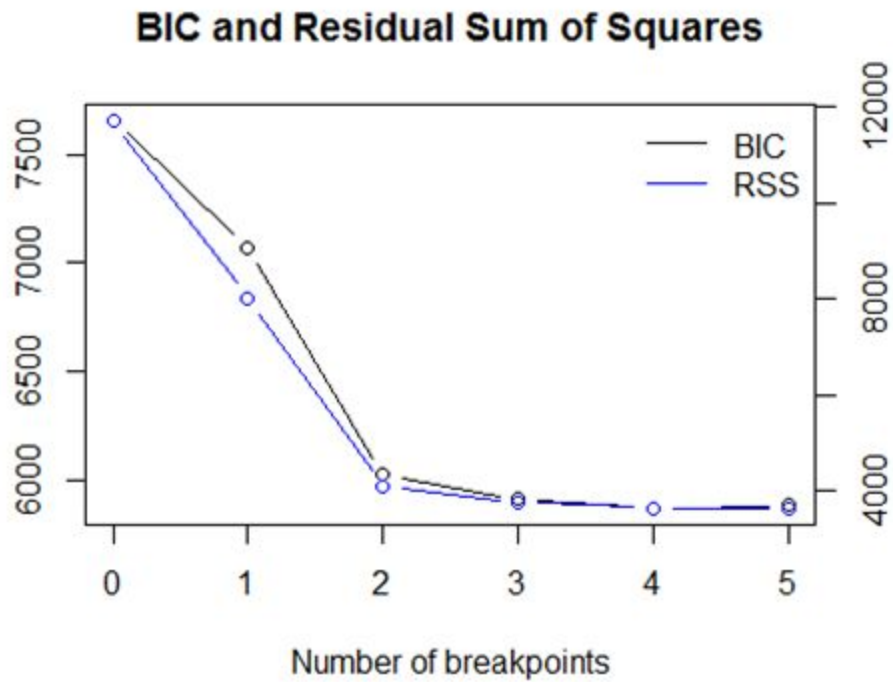
It can be clearly seen that the BIC and RSS favor four break points, or rendering a model with five regimes. Lastly, the breakpoints shown with blue lines can be seen below with the plotted milk futures prices alongside the 95% confidence intervals in red:

Figure 10: Milk Futures Prices with Breakpoints and Confidence Levels

Milk Futures
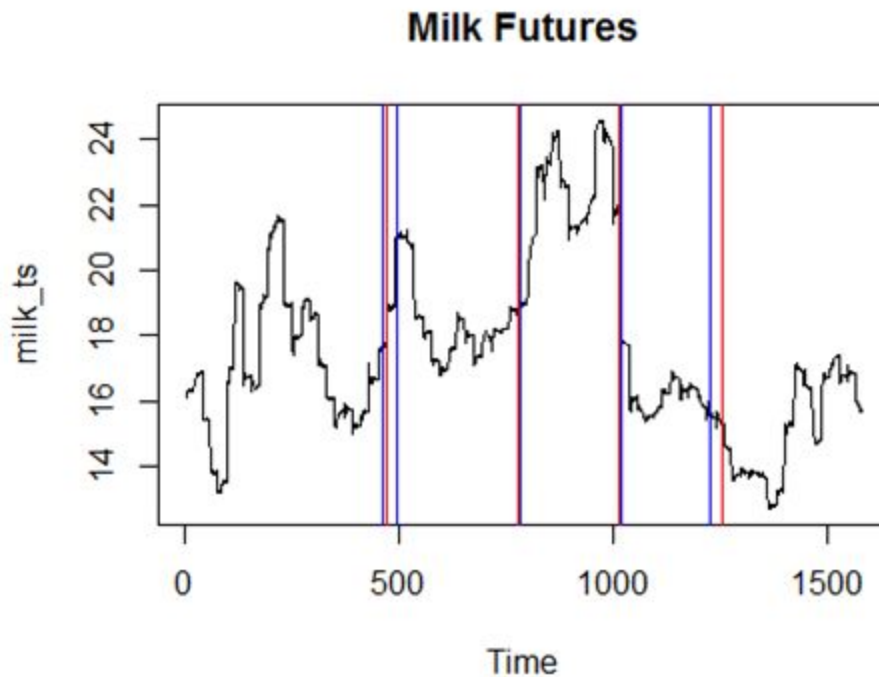
The test generates days since the starting price of the time series September 3, 2010. To facilitate interpretability, the actual dates corresponding to these days can be seen Table 3 above.

A similar story can be had with cheese. Note this time, however, that cheese experiences five distinct breakpoints creating a six segment partition whereas milk only had four break points and therefore a five segment partition.

```
bp.cheese <- breakpoints(cheese_ts~1); bp.cheese
##
##   Optimal 6-segment partition:
##
## Call:
## breakpoints.formula(formula = cheese_ts ~ 1)
##
## Breakpoints at observation number:
## 236 472 779 1015 1256
```

36

```
##
## Corresponding to breakdates:
## 236 472 779 1015 1256
summary(bp.cheese)
##
##   Optimal (m+1)-segment partition:
##
## Call:
## breakpoints.formula(formula = cheese_ts ~ 1)
##
## Breakpoints at observation number:
##
## m = 1                   1016
## m = 2               779 1015
## m = 3           450 779 1015
## m = 4   236 472 779 1015
## m = 5   236 472 779 1015 1256
```

Table 6: Cheese Breakpoints and 95% Confidence Limits

Cheese Breakpoints			
	2.5%	Breakpoint	97.5%
1	8/19/2011	9/12/2011	12/15/2011
2	8/23/2012	9/4/2012	9/14/2012
3	12/2/2013	12/4/2013	12/6/2013
4	11/26/2014	11/28/2014	12/3/2014
5	7/24/2015	11/30/2015	12/23/2015

Figure 11: Cheese Futures Prices with Breakpoints and 95% Confidence Levels

Cheese Futures
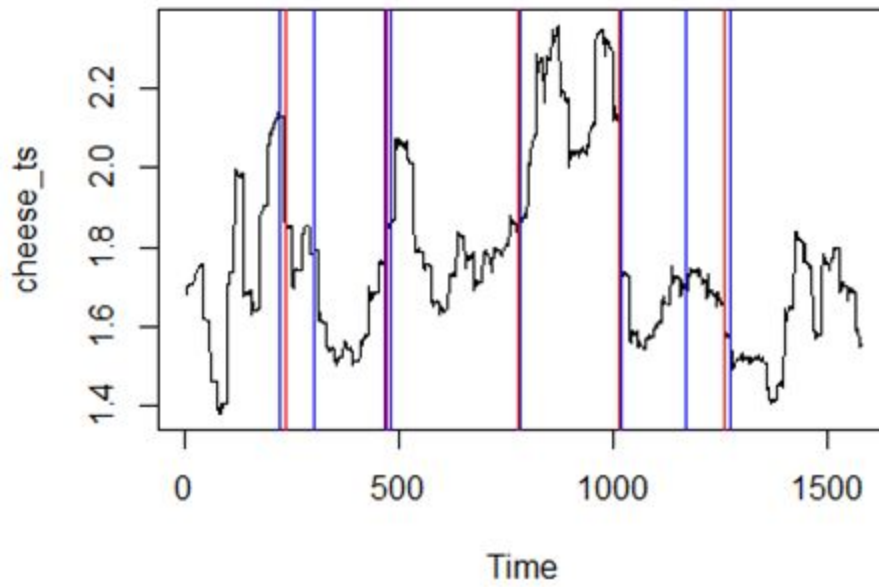
What is more interesting, however, is that milk and cheese share many exact breakpoints

and confidence limits with the dates below, or approximately so.

Table 7: Breakpoint Comparison of Milk and Cheese

Future	2.5%	Breakpoint	97.5%
Cheese	8/19/2011	9/12/2011	12/15/2011
Cheese	8/23/2012	9/4/2012	9/14/2012
Milk	8/20/2012	8/29/2012	10/8/2012
Cheese	12/2/2013	12/4/2013	12/6/2013
Milk	12/2/2013	12/4/2013	12/5/2013
Cheese	11/26/2014	11/28/2014	12/3/2014
Milk	11/26/2014	11/28/2014	12/3/2014
Cheese	7/24/2015	11/30/2015	12/23/2015
Milk	10/9/2014	11/20/2015	11/25/2015

Conclusions

Milk and Cheese Futures prices constitute a clear cointegrating relationship. This is strongly evidenced by the results of the Engle-Granger Two Step method using the Dickey-Fuller test for stationarity. This is indicative of a common stochastic trend among the prices of cheese futures and the prices of milk futures. Such result should not be surprising, however. Cheese is made from milk and milk price is set by the government using, among other factors, the market price of cheese. Consequently, it appears rather sensible that these two commodities should share a cointegrating relationship.

Upon finding evidence for cointegration, a VAR and VECM was fit. As noted before, a VECM is a special kind of VAR model. What is interesting, though, is that different explanatory variables were found to be significant in each model. In the case of the VAR,

present day milk returns were the result of previous milk and cheese returns. However, cheese returns themselves had no significant components.

This stands in striking contrast to the VECM. With the added error correction component, OLS estimates showed that for milk returns, only once lagged milk returns were significant. However, cheese returns significantly depended on both lagged milk returns and cheese returns. Additionally, it had a significant error correction term.

Compounding this story is the results of the test for Granger causality. It is evident that cheese returns Granger cause milk returns but not vice versa. Yet, the presence of cointegration is strongly statistically significant according to the methods developed by Engle-Granger. Taking into account the cointegration again necessarily brings vector error correction modeling to the forefront. And, the Granger causality, like the VAR, conflicts with the results of the VECM.

Intuition does not help to clarify initial considerations of these models. Because cheese is made from milk, it seems likely that the returns of cheese would be determined in part by the returns of milk from previous points in time. Additionally, federal pricing of milk futures incorporates prior prices of cheese into the price formulas. Consequently, it also seems likely that milk returns would be dependent on that of cheese.

Structural breaks are clearly present in the model. Not only are there statistically significant breaks, but also that they by and large are the same between both milk and cheese futures. This suggest the presence of nonlinearities in the underlying pieces of these models. That is, there exists many different regimes among these series, and different regimes entail different parameter estimates for each regime. What is interesting, however,

is that despite having a clearly cointegrated relation where both milk and cheese futures follow the same stochastic trend, there was one break point more in cheese that was not found in milk. Overall, the time series plots look the same, and prices and returns move in almost the same direction. Yet, BIC and RSS were minimized by the time the Bai-Perron method had detected four break points, whereas it took five for cheese.

Discussion

Often times, investigation into multivariate time series of this sort does not initially pair up vector autoregression, stationarity, and the like with regime switching. However, the inclusion of all these methods help elucidate the contradictory results introduced above. Better modeling may need to incorporate the reality of the changes in regime. From there, not only forecasting, but the potential effect of regulation in this industry may become more apparent.

As was seen from the methods modeling and estimating the VAR and VECM representations of this data, milk and cheese returns were regressed against one another along with lagged terms of their own returns. Furthermore, cointegration was set up as a linear regression. Consequently, there are all the statistical assumptions that accompany that kind of modeling. For example, one potential violation of these assumptions is the Normal Q-Q Plot in Figure where there are fat tails. The numerical value of the kurtosis which makes those tails fat was at a very large 8.75. This may be indicative of variance changing due to volatility, among other things.

By far the most concerning result for these assumptions, however, is the presence of regime switching. The methods used for the models in this paper presuppose a certain linearity across the data. However, the regime breaks threaten this underlying assumption. These conditional nonlinearities confound the underlying statistical assumptions that make sound results of these models possible.

Methods developed incorporating regime-switching and cointegration developed by Markus Jochmann and Gary Koop[20] may hold the key to applying the methods of this paper in a way that accounts for the nonlinearities. Such modeling requires deployment of Markov switches and Bayesian inference. This methods allows and account for parameter changes in regime switches and can also model changes in cointegrating relationships during these regime switches. Development of his methods may provide a cogent way to further explore the how cheese and milk futures relate to one another and uncover the reasons behind the changes in regimes.

All in all, this derivative market will require more time and attention in order to better understand the relationships and attribute causes to the shocks. The first steps have been established, though. Once the methods behind the models for this data have been rendered, it will be all the easier to apply it to other commodities and markets as well, such as cattle, oil, and many more.

---

[20] Jochmann, Markus and Gary Koop, *Regime-switching cointegration*, Studies of Nonlinear Dynamics and Econometrics, Vol. 19, No. 1, (Feb., 2015): 1.

References

Alexander, Carol, *Market Models: A Guide to Financial Data Analysis,* West Sussex: John
        Wiley & Sons Ltd, 2001

Ang, Andrew and Allan Timmermann. "Regime Changes and Financial Markets." *Annual
        Review of Financial Economics* Vol. 4 (2012)

Bai, Jushan and Pierre Perron. "Estimating and Testing Linear Models with Multiple
        Structural Changes." *Econometrica* Vol. 66, No. 1 (Jan., 1998)
        http://www.econ.nyu.edu/user/baij/econometrica98.pdf, pg 47

Bai, Jushan and Pierre Perron. "Estimating and Testing Linear Models with Multiple
        Structural Changes." *Econometrica* Vol. 66, No. 1 (Jan., 1998)

Bozic, Marin and T. Randall Fortenbery. "Creating Synthetic Cheese Futures: A Method for
        Matching Cash and Futures Prices in Dairy." *Journal of Agribusiness* Vol. 30, No. 2
        (2012)

Engle, Robert F. and Clive Granger. "Co-Integration and Error Correction: Representation,
        Estimation, and Testing." *Econometrica* Vol. 55, No. 2 (Mar., 1987)
        http://www.jstor.org/stable/1913236?seq=1#page_scan_tab_contents

Geweke, John and William McCausland. "Bayesian Specification Analysis in Econometrics*."
        Oxford University Press, Agriculture and Applied Economics Association* Vol. 83,
        Proceedings No. 5 (Dec, 2001)

Gould, Brian W. University of Wisconsin Dairy Marketing and Risk Management Program.
        Pricing Formulas. https://future.aae.wisc.edu/collection/fmmo/pricing_f.pdf

Hayek, Fredrich A. "*The Use of Knowledge in Society."* The Library of Economics and Liberty
        (1945)

Jochmann, Markus and Gary Koop. "Regime-switching cointegration." *Studies of Nonlinear
        Dynamics and Econometrics* Vol. 19, No. 1, (Feb., 2015)

Murray, Michael P. "A Drunk and Her Dog: An Illustration of Cointegration and Error
        Correction." *The American Statistician* Vol. 48, No. 1 (Feb., 1994)

Perron, Pierre. "Dealing with Structural Breaks." *Palgrave Handbook of Econometrics* Vol. 1
        (2005) http://people.bu.edu/perron/papers/dealing.pdf

Perron, Pierre. "Dealing with Structural Breaks." *Palgrave Handbook of Econometrics* Vol. 1
        (2005)

Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach, Fourth Edition.* Mason: South-Western, 2009.

Reflection

I came to Utah State University expecting much of a similar experience which I had in high school. I didn't think I could eat in a place that wasn't a cafeteria, I expected stringent grading procedures with high expectations, and I placed strict studying regiments on myself. However, it became clear after my first semester that the early college experience was in fact easier than my honors experience in high school. Shortly thereafter, I applied and was accepted into the honors program here at Utah State University so that I could challenge myself and go deeper into the disciplines I loved.

This honors capstone, as the capstone proper of my undergraduate collegiate education, constituted in many ways the single most difficult project I have embarked on at Utah State. My early honors experience was formed mostly by deeper readings into philosophy. At the time, I had envisioned becoming an academic, although a clerical one, in that field. After passing through the Koch Scholar program, I met Dr. Tyler Brough. Despite not having any financial experience, I did have statistical training which Dr. Brough took note of. The normal sequence of the finance and economics major does not entail any math above introductory calculus and statistics. By virtue of having statistical skills, I was poised to be better adept to engaging in deeper applications of finance which necessitates strong quantitative skills. As such, Dr. Brough was very willing to work with me and help me in my own endeavors.

I chose finance to do my capstone research because it seemed to have a good tie between my majors of philosophy and statistics. Financial modeling relied heavily on the statistics, and the policy questions which were informed by the modeling sometimes crossed into philosophical territory. Additionally, statistical methods often times rely on some philosophical assumptions as to what constitutes good science. This can be chiefly shown in the Bayesian vs. Frequentist debate in statistics. Yet, the prior exposure of some of these themes and tools did not always ease the challenges ahead.

My statistical training has largely been outside of time series analysis, whereas this project was entirely time series analysis. I had to build everything from the ground up. This proved to be an enormous difficulty as I often did not understand what I was doing. I spent countless hours programming in R, reading academic journals, and talking with Dr. Brough over and over again about the same questions. It was not until about March that core concepts finally began to sink in.

The other challenge I faced was the evolving nature of the project. At first, we envisioned taking cheese spot prices and testing for cointegration with milk futures. We thought we could hit the ground running. Additionally, we imagined that there would be clear correlations with changes in government policy. However, the first change was caught up in my inexperience and Dr. Brough's demands in other parts of his working career. We then imagined dropping the question of the government's effect on policy. Additionally, our data did not have clear cheese spot prices, so we had to change to using cheese future prices. However, cheese futures are a recent creation, and there are periods in times with missing observations. We had hoped to use specification analysis techniques to model regime changes.

I became worried, though, that these adaptations would not satisfy the original vision. Yet, to fulfill that original vision would mean to incorporate far more advance methods as outlined in the conclusions. That kind of programming and modeling would take a lot more practice, reading, and training which would make the project much more adept for a master's or even PhD thesis/dissertation - which it may become. Consequently, the idea became a synthesis of modeling techniques to lay the groundwork and show that traditional methods for these data may not be appropriate for this data set. And, this certainly was the case. There will be many more questions and interesting insights to be had as such.

Despite the challenges, I owe everything to this project. I learned how to work to teach myself new, hard material. Furthermore, it led to my eventual employment with Strata, which not only offered new research opportunities, but also funded my upcoming master's program in financial economics.

And, it is because of this experience that I have pursued that very program. The skills in time series analysis I have gained in this project directly apply and correspond to the work I will be doing with Strata and, eventually, my master's thesis. It is amazing to think back and see how one opportunity led to another. For the chance to be a part of this Honors program and what it has given me, I am incredibly thankful.

David Zynda

May 2, 2017

Bradley David Zynda is a graduate of Utah State University with degrees in Statistics and Philosophy. Born and raised in Ohio, he came to Utah State to originally pursue a degree in Electrical Engineering. After taking his first philosophy and statistics courses, however, the path forward was unwaveringly set. Philosophy, as Wittgenstein prescribed, acted as therapy through the myriad experiences and challenges David faced which come with being 1000 miles away from home and discovering oneself. Statistics and its foundations was never too removed from philosophy either, especially concerning its theoretical underpinnings and understanding of knowledge. In pursuit of continuing to grow in knowledge and research, David will continue his studies as Utah State University through funding provided by the Institute of Political Economy. As he pursues a Masters in Financial Economics, he will continue his work with Dr. Tyler Brough that began with this project.