commodity price prediction and research

Brandon Moretz

### Introduction

The purpose of this artifact is to memorialize the findings from the exploratory data analysis performed on the commodities historical price data. There will be several revisions to this document as the analysis unfolds and relationships are extracted, distributions are parameterized, and trading strategies are formalized.

### data processing and market STYLIZED facts

The dataset we are working with in this project is composed of price points on various energy commodities, specifically we have intraday open, high and low and the daily closing price of which we will be exploring in-depth for several commodities. Our initial commodities of interest include Crude (WTI and Brent), Heating Oil (Ultra low Sulfur), Gasoline and Natural Gas. The first step in financial data analysis is typically to take the pricing data and convert these to returns. This transformation serves multiple purposes; however, we are going to specifically callout the two main benefits of this approach.

The first benefit we get from using returns over prices is that this puts all the symbols on a level playing field by process of normalization: measuring all the variables in a comparable metric, thus enabling evaluation of analytical relationships amongst the variables despite originating from prices of unequal values. This is not only a nice mathematical property and convenience; it is a requirement for many of the statistical and mathematical models and procedures we are going to implore in later research.

The second benefit of this normalization process is a corollary from the first stylized fact of market prices, in that they follow an approximately log-normal distribution due to the prices having the mathematical property of being restricted to the domain of greater than or equal to zero. By converting our price series first into simple returns, and then taking the natural log of this series, we transform the data distribution from log-normal, to one that is approximately normal. We can prove this mathematically, however, for brevity we will instead rely here on the mathematical intuition derived from the Central Limit Theorem (CLT), which states that “the sum of a number of independent and identically distributed random variables (Ri) with finite variances will tend to a normal distribution as the number of variables grows.” Here, our price series are assumed to be independent and identically distributed, our returns are log-transformed which have the property of time-dependent additivity (Rt+1 = Rt + Rt-1), and thereby will be approximately normal. Later, our assumptions of normality will be revised and expanded to include maximum likelihood estimates for parametric versions of the normal distributions to further account for both tail risk and tail dependence, both of which are market stylized facts that will be explored in-depth.

### Relationships

### Research

The

### Conclusion

### Appendix

