Measuring the Relationship between Price and Yield

Robert S. Thompson - North Carolina State University - rsthomp2@ncsu.edu

Selected Paper prepared for presentation at the 2020 Agricultural & Applied Economics Association Annual Meeting, Kansas City, MO
July 26-28, 2020

Measuring the Relationship between Price and Yield

Robert S. Thompson

Abstract

There is an inverse relationship between price and yield for field crops. The correlation between historical deviations from expected prices and yields is commonly used as a measure of that relationship. In this paper, I use a simple system of supply and demand to show that the correlation between price and yield deviates depends on the level of demand side uncertainty which changes over time. Thus, correlation estimated with historical prices and yields is unlikely to be representative of the relationship in future years of interest. Using the system of supply and demand I propose a new measure of the relationship between price and yield and derive their joint distribution which provide several insights. I identify a new way to estimate the elasticity of demand for these agricultural commodities, provide a new method to derive the joint distribution of price and yield for revenue insurance rating, and show that prices may not be log-normally distributed.

1 Introduction

In the short run there is an inverse relationship between price and yield for agricultural commodities. Yields higher (lower) than expected are supply shocks that result in prices lower (higher) than expected. Of course, yield is only one part of what determines price. Yield, total number of acres, and demand jointly determine price. At planting time, harvest time prices are uncertain and uncertainty about price is determined by uncertainty about yield, total number of acres, and demand.

The relationship between price and yield is important to measure for several reasons. It is important in general for price forecasting purposes. Also, revenue per bushel is the product of price and yield. Because price and yield are both uncertain at planting time, the joint distribution of price and yield at harvest is necessary to determine the distribution of revenue.

One possible measure of this relationship is the correlation coefficient. This is the method adopted by the USDA Risk Management Agency (RMA) for purposes of rating revenue insurance (Coble et al. [2010]). More specifically, the RMA estimates the correlation between historical log price and yield deviates from expectations. They use the correlation in order to model the joint distribution of price and yield in the upcoming year. A point of concern in using this measure is that the variance of price changes from year to year. Although the best measure of the variance of price is a subject of controversy (Goodwin et al. [2017]), it is generally accepted that the variance of price changes over time. The correlation between log price and yield deviates is given by

$$\mathrm{Corr}[\log(P_t) - \mu_{\log(P_t)}, Y_t - \mu_{Y_t}] = \frac{\mathrm{Cov}[\log(P_t) - \mu_{\log(P_t)}, Y_t - \mu_{Y_t}]}{\sqrt{\mathrm{Var}[\log(P_t) - \mu_{P_t}]}\sqrt{\mathrm{Var}[Y_t - \mu_{Y_t}]}}$$

There are three terms that comprise this measure: the variance of price, of

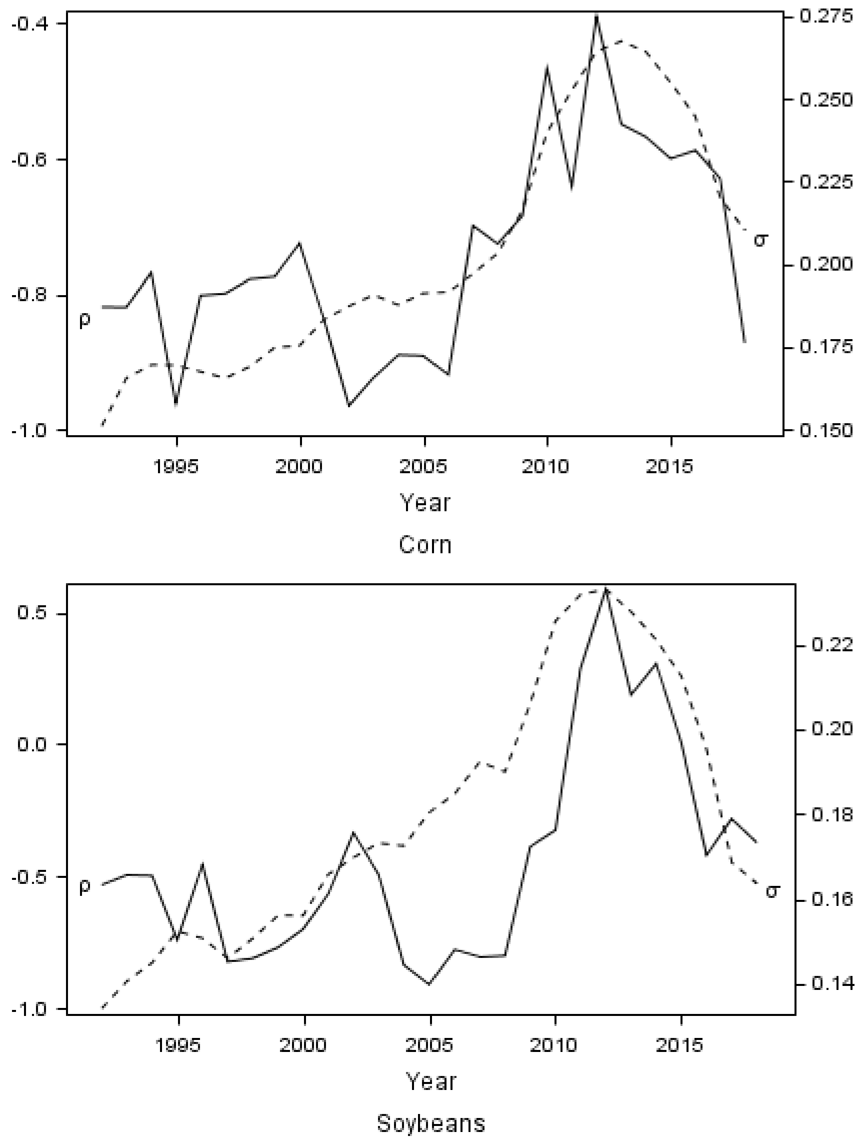

Figure 1: 7 year rolling average of the standard deviation of price (σ) and correlation between price and yield deviates (ρ) for corn and soybeans.

yield, and the covariance between the two. It should be clear this measure is only constant across time if each term that comprises this measure is constant across time – or the terms in the denominator change by the same proportion as the term in the numerator. Only when this measure is constant across time can it 1) be estimated with historical prices and yields, and 2) provide an accurate forecast of the relationship in the future. If the variance of price changes over time, it is likely the correlation changes as well.

Figure 1 shows a seven year rolling average of the variance of price[1] and a seven year rolling estimate of the correlation between price and yield deviates over time. This shows directly the effect of changes in price variance on the correlation between price and yield. As the variance of price increases, the absolute value of correlation decreases. That is, as the variance of price increases, less of the variations in price are explained by variations in yield. Figure 1 suggests that the reason the correlation changes over time is in part due to changes in the variance of price.

Paulson and Babcock [2008] also note that the correlation between price and yield deviations vary over time. Specifically, they note that for corn, the correlation estimated using the years 1975-2005 is -0.66, and using the years 1990-2005 is -0.81. The reasons for this, they explain, are changing agricultural policies over time. This suggests that correlation estimated by historical prices and yields may be a poor measure of the relationship between price and yield and a further investigation into that relationship is warranted.

Recent literature suggests copulas may be appropriate to use to determine the relationship between price and yield (Goodwin and Hungerford [2014]; Ramsey et al. [2019]). Different types of copulas allow different types of nonlinear relationships whereas correlation can only model linear relationships. There are a wide variety of measures one could use to represent the relationship between price and yield, but which of those most appropriately reflects the true relationship between price and yield is unknown. Moreover, that question can only be answered when one knows the true relationship between price and yield.

[1] as given by the Black-Scholes implied volatility of the harvest contract at planting time.

In the following section, I derive the true relationship given a specific system of supply and demand equations. With a structural model for prices based on supply and demand, a more rigorous evaluation of what measures might be appropriate to estimate that relationship is performed. Although they do not consider yields explicitly, the following analysis is similar in spirit to that of Piggott [1978] and Myers and Runge [1985], which describe sources of price instability in terms of a system of supply and demand equations.

2 A Structural Model of Price and Yield

Consider a market for grain characterized by the following system of supply and demand.

$$Q_{D,t} = \alpha_t P_t^{-\eta} \tag{1}$$

$$Q_{S,t} = A_t Y_t \tag{2}$$

$$A_t = f_t(E[P_t|I_{\text{Planting}}]) \tag{3}$$

where $Q_{D,t}$ and $Q_{S,t}$ are quantity demanded and supplied in year t. In this system quantity demanded and supplied are units of grain, so in the following I assume they are measured in bushels. Quantity demanded is given by an exponential function with constant elasticity $-\eta$. The demand shifter, α_t, can be thought of as a function of all the factors that drive demand, including preferences, income, substitute good prices, complement good prices, domestic and international policy, and any other factors that drive demand. The timing of quantity supplied is different. Farmers do not directly select the amount of bushels to produce, rather, they select the amount of acres (A_t) to allocate to this grain. The amount of acres they select is a function of *expected* harvest time prices given the available information at planting time. The quantity of acres supplied is given by equation 3, which for the purposes of this paper suffices to be an arbitrary supply function – meaning no specific functional form need be imposed. Once the quantity of acres supplied is chosen, yields[2] (Y_t) are determined by nature and revealed to the market at harvest. The product of yield and acres gives the total quantity supplied, $Q_{S,t}$. The demand shifter, α_t is allowed to vary from year to year, but I assume the elasticity of demand is constant over time. The reason this assumption is needed will become apparent later.

It is important to note the results that follow crucially depend on the specific functional form of the demand equation I have chosen. The exponential demand function is, however, mathematically convenient. The reason is that quantity supplied is log linear in yields – no matter the specification of the supply of acres. Any specification of demand that is log linear in prices will result in a log linear relationship between price and yield. Thus the fundamental assumptions of this system are 1) quantity demanded is log linear in prices and 2) the elasticity of demand is constant over time. If these two assumptions are satisfied, the following results will hold.

Market equilibrium is met at harvest where $Q_{D,t} = Q_{S,t}$. Solving for market equilibrium prices at harvest gives,

$$P_t = \left(\frac{\alpha_t}{f_t(E[P_t|I_{\text{Planting}}])Y_t} \right)^{\frac{1}{\eta}}, \tag{4}$$

or in log terms,

$$\log(P_t) = \frac{1}{\eta}\Bigg(\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}])) - \log(Y_t) \Bigg). \tag{5}$$

[2] measured in bushels per acre

This gives the structural relationship between price and yield. Increases in supply serve to decrease prices. Increases in demand, α_t, serve to increase prices.

At this point it is important to identify which components of equation 5 are observable and when they are observed. Yields and prices are observable but only so at harvest. At planting time, prices and yield are random variables. Expected harvest prices given the information at planting time are observable from futures prices at planting time for contracts that call for delivery at harvest time, and the variance may be inferred from options prices. Since acres are determined at planting time, $f_t(E[P_t|I_{\text{Planting}}])$ is constant at planting time although it may not be known with certainty. The demand elasticity, η, is an unobservable constant. The demand shifter, $\log(\alpha_t)$, is unobservable and only realized at harvest, when the grain is brought to market. Prior to harvest the demand shifter is a random variable.

Given equation 5, the uncertainty about price can be attributed to the uncertainty about each individual factor that comprises price. Assuming the three random variables in the RHS of equation 5 are independent[3], the variance of harvest price at planting time is given by,

$$\text{Var}[\log(P_t)] = \frac{1}{\eta^2}\Bigg(\text{Var}[\log(\alpha_t)] + \text{Var}[\log(f_t(E[P_t|I_{\text{Planting}}]))] + \text{Var}[\log(Y_t)]\Bigg).$$

This shows that the uncertainty about price depends on the uncertainty about demand, the nnumber of acres, and yields. Also, this uncertainty is scaled by the demand elasticity. All else equal, an inelastic demand would have a higher price uncertainty than an elastic one. This characteristic was also noted by Lence and Hayes [2002], who showed that the level of price uncertainty was responsive to demand elasticity.

[3]This assumption is not necessarily trivial, as there is some evidence that yield is responsive

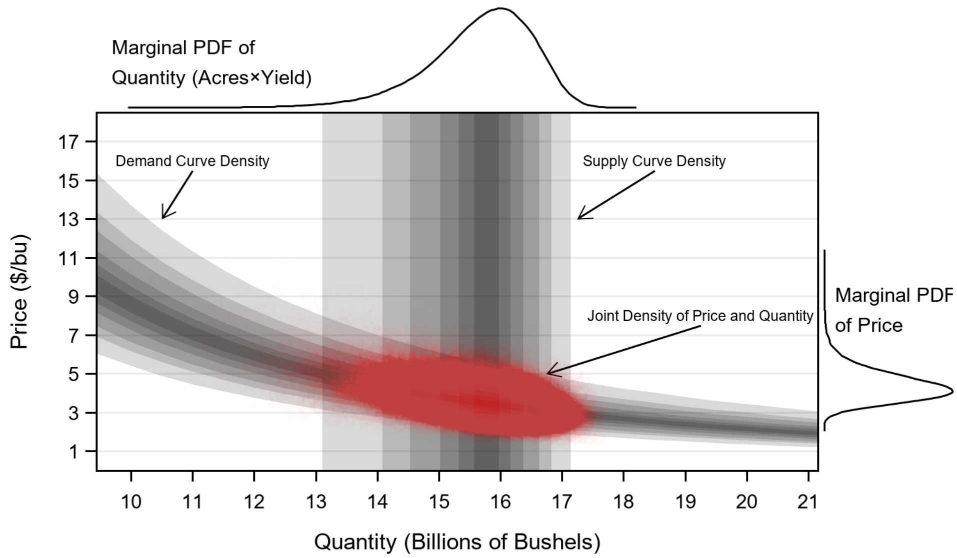

Figure 2: Example System of Supply and Demand for Corn

Figure 2 shows a depiction of this system and the different sources of uncertainty. After planting time, the supply curve is vertical because the number of acres has been chosen, but there is a range of uncertainty due to the uncertainty about yields. The demand curve also has uncertainty around it, such that both sources of uncertainty, supply and demand, contribute to price uncertainty, whose distribution is shown in the right hand margin.

2.1 Measures of the Relationship between Price and Yield

With equilibrium prices defined, the true relationship between price and yield under this framework is known, which can be used to evaluate which measures of the relationship are most appropriate. I begin by evaluating the commonly used correlation coefficient.

to price (Miao et al. [2016]). Thus yield and demand may be related.

2.1.1 Correlation

As mentioned previously, correlation is currently used by the RMA as a measure of the relationship between price and yield for purposes of rating revenue insurance. The correlation between price and yield deviates within this framework is given by

$$
\begin{aligned}
&\mathrm{Corr}[\log(P_t) - \mu_{P_t}, \log(Y_t) - \mu_{\log(Y_t)}] && (6)\\
&\qquad = \frac{\mathrm{Cov}[\log(P_t) - \mu_{P_t}, \log(Y_t) - \mu_{\log(Y_t)}]}{\sqrt{\mathrm{Var}[\log(P_t)]}\sqrt{\mathrm{Var}[\log(Y_t)]}} && (7)\\
&\qquad = -\frac{\sqrt{\mathrm{Var}[\log(Y_t)]}}{\eta\sqrt{\mathrm{Var}[\log(P_t)]}} && (8)\\
&\qquad = -\frac{\sqrt{\mathrm{Var}[\log(Y_t)]}}{\sqrt{\mathrm{Var}[\log(\alpha_t) - \log(f_t(E[P_t|I_{\mathrm{Planting}}]))] + \mathrm{Var}[\log(Y_t)]}}. && (9)
\end{aligned}
$$

This shows that the correlation may change over time because it depends on the relative variance of $\log(P_t)$ and $\log(Y_t)$[4]. Only when $\frac{\mathrm{Var}[\log(Y_t)]}{\mathrm{Var}[\log(\alpha_t)-\log(f_t(E[P_t|I_{\mathrm{Planting}}]))]}$ is constant over time is correlation constant across time. This condition would need to hold in order for an estimate of correlation from a sample of historical prices and yields to be representative of the relationship between price and yield in the future. Also, note that this result does not depend on the assumption of a constant elasticity of demand – only the assumption that quantity demanded is log-linear in price.

This condition is unlikely to hold in most markets for field crops. Even if the acreage uncertainty is constant, there are changes in demand side uncertainty across time. Perhaps the best example of this is the ethanol mandate implemented in the United States which increased corn demand – and corn demand uncertainty – to a high level between around 2005 and 2010. There are many

[4]The RMA uses yield deviates, not log yield deviates in estimating the correlation. The same general result – the absolute value of correlation decreases as the variance of price increases – applies in this case as well. However, there is no closed form solution available because of the non-linear relationship between $\log(P_t)$ and Y_t under this system.

sources of changes in demand side uncertainty, many of which stem from domestic and international policy changes. The recent US trade war with China and the COVID-19 pandemic are other examples of this. The evidence shown in figure 1 is also consistent with these results. During periods of high price variance – likely due to increases in demand side uncertainty – the absolute value of correlation decreases.

The important point is that correlation measures the *relative* amount of variation in one variable explained by variation in the other variable. In this system, any changes in uncertainty of prices due to changes in demand side uncertainty inherently changes the *relative* amount of variation of prices explained by variations in yield. Therefore, correlation is unlikely to be an appropriate measure of the relationship between price and yield across time under this system of supply and demand.

Ultimately if the sample correlation is estimated in this fashion with historical price and yield deviates, the result according to this framework will be the correlation at the average historical price variance. It follows that if the variance of price in the year of interest is higher (lower) than average, then the dependence between price and yield will be overestimated (underestimated).

2.1.2 Covariance

Given the potential problems with correlation I now propose an alternative measure of the relationship between price and yield. Ultimately what is needed is a measure that is constant over time so that it can be consistently estimated with a sample of historical prices and yields.

Because the expected values of price and yield change each year it is important to center the variables over their means such that they are comparable across time. Note that I assume expected price can be obtained from futures markets and expected yield can be obtained from historical yield[5]. Since $\log(P_t)$ is linear in $\log(Y_t)$, covariance, a linear measure, is appropriate to describe their relationship. Using equation 5 to derive the covariance between the two gives;

$$\mathrm{Cov}[\log(P_t) - \mu_{P_t}, \log(Y_t) - \mu_{Y_t}] = -\frac{1}{\eta}\mathrm{Var}[\log(Y_t) - \mu_{Y_t}]. \tag{10}$$

This measure is not necessarily constant across time because it includes the variance of yield which may change over time[6]. In order to estimate this with a sample of historical prices and yields the measure must be constant across time. To obtain a time invariant measure, I divide the log yield deviates by their variance and then calculate the covariance,

$$\mathrm{Cov}\left[\log(P_t) - \mu_{P_t}, \frac{\log(Y_t) - \mu_{Y_t}}{\mathrm{Var}[\log(Y_t) - \mu_{Y_t}]}\right] = -\frac{1}{\eta}, \tag{11}$$

which only depends on the elasticity of demand[7]. This is intuitive, as yield deviates are supply shocks that are realized after acres supplied has been chosen. Changes in yield from expectations manifest themselves as movements up or down the demand curve, hence the role of the elasticity of demand. More importantly for practical purposes, this measure is constant across time as long as the elasticity of demand is constant across time. Using this, the elasticity of demand can be calculated as well,

$$\eta = -\mathrm{Cov}\left[\log(P_t) - \mu_{P_t}, \frac{\log(Y_t) - \mu_{Y_t}}{\mathrm{Var}[\log(Y_t) - \mu_{Y_t}]}\right]^{-1}. \tag{12}$$

This will be useful in the derivation of the joint distribution of price and yield later on. Also, it should now be apparent the need for the assumption that the

[5]This is usually done assuming a linear trend in time.

[6]If in fact the variance of yield is constant over time then this measure is constant across time. To avoid unececcary assumptions, I do not assume this is the case.

[7]This could be estimated with a regression using equation 5 as well.

elasticity of demand is constant across time. With this assumption the elasticity of demand can be estimated using a sample of historical prices and yields. This is the metric that defines the relationship between historical prices and yields that will carry through to future prices and yields. Important in its own right, I have identified a novel way in which the elasticity of demand for field crops may be estimated.

According to this framework, the covariance as shown in equation 11 is an appropriate measure of the relationship between price and yield. Important for practical purposes, this measure can be estimated using historical prices and yields. The two fundamental assumptions that lead to this result are 1) quantity demanded is log linear in prices and 2) the elasticity of demand is constant over time. I do not wish to argue these assumptions are completely satisfied in the markets of interest. I do argue, however, that these assumptions are 1) more reasonable and 2) less limiting than those that would be required for correlation to be an appropriate measure of the relationship between price and yield. They are more reasonable in the sense that there isn't evidence to suggest they are violated in these markets, while there is evidence to suggest the assumptions that would be required for correlation to be an appropriate measure are violated – primarily the evidence that the variance of price changes over time due to changes in demand side uncertainty. They are less limiting in the sense that they allow correlation to change over time for which there is evidence as shown in figure 1.

I proceed to evaluate what the joint distribution of price and yield may be under this framework and compare it to the distribution derived from a correlation based framework. This joint distribution may be of interest in several applications, one of which is rating revenue insurance.

2.2 The Joint Distribution of Price and Yield

There are a number of ways to proceed to evaluate the joint distribution of price and yield given this framework. In this section, I explore a few possible methods and explain the benefits and drawbacks of each.

2.2.1 Log-Normal Prices

The first and the simplest is to use the method the RMA uses currently to derive the joint distribution of price and yield – with one exception. Instead of estimating the sample correlation between historical price and yield in the usual way, estimate the sample covariance between historical price and yield as given in equation 6. Next, estimate the variance of yield. Then, obtain the variance of price from options markets in the year of interest via the Black-Scholes (BS) option pricing model. With these three components, the correlation in the year of interest can be calculated. With this correlation estimate, the rest of the methods the RMA currently uses to derive the joint distribution can be used as is.

The main benefit of this approach is its simplicity and consistency with common practice. There is one main drawback of this approach. Prices are assumed to follow a log-normal distribution – in Monte Carlo simulations and the BS model to obtain the variance of price. The problem with this specification is that there is empirical evidence that yields – and subsequently log yields – are not necessarily normally distributed. Specifically, the distribution of yields is negatively skewed, such that the tail of the distribution is fatter on the left than the right[8]. This empirical evidence along with the framework presented above imply that prices are not necessarily log-normally distributed. Recall equation

[8] Log yield will be even more strongly skewed than yield.

5,

$$\log(P_t) = \frac{1}{\eta}\Bigg(\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}])) - \log(Y_t)\Bigg).$$

This shows that if $\log(Y_t)$ and $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$ are both normally distributed, then $\log(P_t)$ will be normally distributed. If $\log(Y_t)$ is negatively skewed, for which there is empirical evidence, and $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$ is symmetrically distributed, then the distribution of $\log(P_t)$ will be positively skewed. This implies that the distribution of P_t will be even more positively skewed than the log-normal distribution. This in itself has implications for option pricing, as the BS model is based on the assumption that prices are log-normally distributed. If prices are not log-normally distributed analyses derived using the BS model will be invalid.

2.2.2 Non-Log-Normal Prices

To derive the joint distribution of price and yield that is completely consistent with this framework, the log-normal assumption must be relaxed. In doing this the BS model must be abandoned and a new option pricing model consistent with this framework must be developed. This is a surprisingly simple feat since the equation relating price and yield is completely specified. Suppose for a moment that the first two moments of log price are known. The first two moments of the distribution of yield can be estimated from historical yields. This leaves only two pieces of the equilibrium price that must be accounted for, namely α_t and $f_t(E[P_t|I_{\text{Planting}}])$. For our purposes it suffices to model the log difference of these two variables rather than individually. The first two moments of the difference of these two variables can be inferred from the first two moments of $\log(P_t)$ and $\log(Y_t)$. Rearranging equation 5,

$$\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}])) = \eta\log(P_t) + \log(Y_t). \tag{13}$$

and taking expectations,

$$E[\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))] = E[\eta \log(P_t)] + E[\log(Y_t)] \tag{14}$$

$$= \eta E[\log(P_t)] + E[\log(Y_t)]. \tag{15}$$

The variance can be calculated in a similar manner[9], again centering the variable over its mean,

$$\text{Var}[\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}])) - \eta E[\log(P_t)] + E[\log(Y_t)]] =$$

$$\eta^2 \text{Var}[\log(P_t)] - \text{Var}[\log(Y_t)] \tag{16}$$

Notice that these moments depend on several parameters. The first two moments of log yield can be obtained from historical yields and η can be obtained from historical prices and yields. Also, continue to suppose the first two moments of log price are known. Although it is uncertain what the distribution of $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$ should be, the first two moments of it can be matched given the first two moments of $\log(P_t)$ and $\log(Y_t)$.

Next, I discuss how to obtain the first two moments of $\log(P_t)$. These moments are not directly observable but can be inferred from market prices. What are observed are futures and options prices. I assume that futures prices represent the expected value of price, $E[P_t]$[10]. Also, I assume that options prices represent the expected value of the option payout, $E[(P_t - k)_+]$ for a call and $E[(k - P_t)_+]$ for a put. Given some values for the first two moments of $\log(P_t)$, and distributional specifications for $\log(Y_t)$ and $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$, a Monte Carlo analysis of the distribution of price can be conducted by using

[9] here I assume that $\log(Y_t)$ is uncorrelated with $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$.

[10] Note that $\log(E[P_t]) \neq E[\log(P_t)]$.

equation 5. Theoretical futures and options prices can be calculated with this price distribution that depend on those values chosen for the first two moments of $\log(P_t)$. The values chosen can be varied until the theoretical futures and options prices match the market futures and options prices. This additional step is needed to obtain the first two moments of $\log(P_t)$ since we have relaxed the log-normal assumption and abandoned the use of the BS model.

Of course, to conduct this Monte Carlo analysis distributions must be specified for both $\log(Y_t)$ and $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$. There is a large literature on the specifications for the distribution of yields, but none on what the distribution of $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$ may be.

This more complicated method is more consistent with the theory developed in this paper. However, departures from the BS model should at the least be taken with caution and differences in the final result – the joint distribution of price and yield – may be small when using this method relative to the method that uses the log-normal assumption.

3 Data

The data used in the following analysis consist of yearly US national level corn and soybean yield and corn and soybean futures prices between 1960 and 2019. Corn and soybean yields are obtained from the USDA National Agricultural Statistics Service (NASS [2020]). Futures prices for corn and soybeans are obtained from Barchart.com (Barchart [2019]). Specifically, the futures prices used are planting time and harvest time prices of the harvest contract. March is planting time for corn and soybeans, and October is harvest time for corn and soybeans. The harvest contract for corn is December and the harvest contract for soybeans is November. A monthly average of daily settlement prices is used in the analysis.

4 Empirical Analysis

In this section, I conduct an empirical analysis of the new framework I have developed. First, I show by how much the correlation between price and yield changes over time, and how it relates to the framework I have developed. Then, I derive the joint distribution of price and yield using a Monte Carlo analysis and compare the results to the methods currently used by the RMA. To simplify the following empirical analysis, I assume that the expected value of log yield follows a quadratic trend in time, and the variance of the mean centered log yields are constant over time.

4.1 Changes in Correlation over Time

As shown previously, there is empirical evidence that correlation between price and yield tends to change over time. According to the theoretical framework in this paper, the change in correlation is soley due to changes in the variance of price. This is of course assuming the elasticity of demand is constant over time. It is important to first show how much correlation responds to changes in the variance of price. This will in turn show by how much the joint distribution of price and yield might be misspecified by using a constant correlation measure. Figure 3 shows two measures of the correlation between price and yield deviates over time for corn and soybeans. The first, $(\bar{\rho})$, is the correlation estimated assuming the correlation is constant across time. The second, ρ, is the correlation calculated in a manner consistent with the theoretical framework in this paper. That is, by first estimating the covariance then using the variance of log price and log yield to calculate the correlation specific to each year. The changes in ρ shown here are soley due to changes in the variance of price. This shows that correlation is quite responsive to changes in the variance of price, varying from around -0.3 to -0.8 for both commodities.

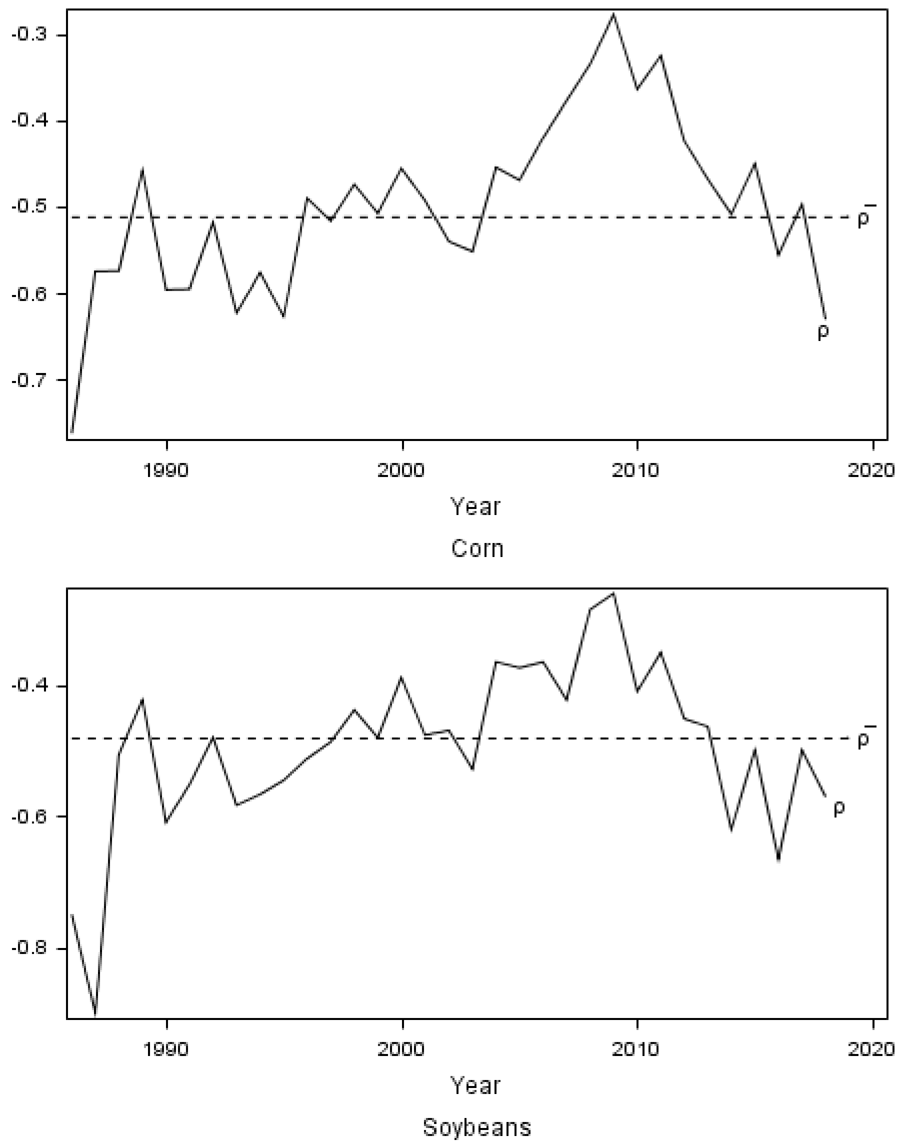

Figure 3: Correlation estimated using historical price and yield deviates

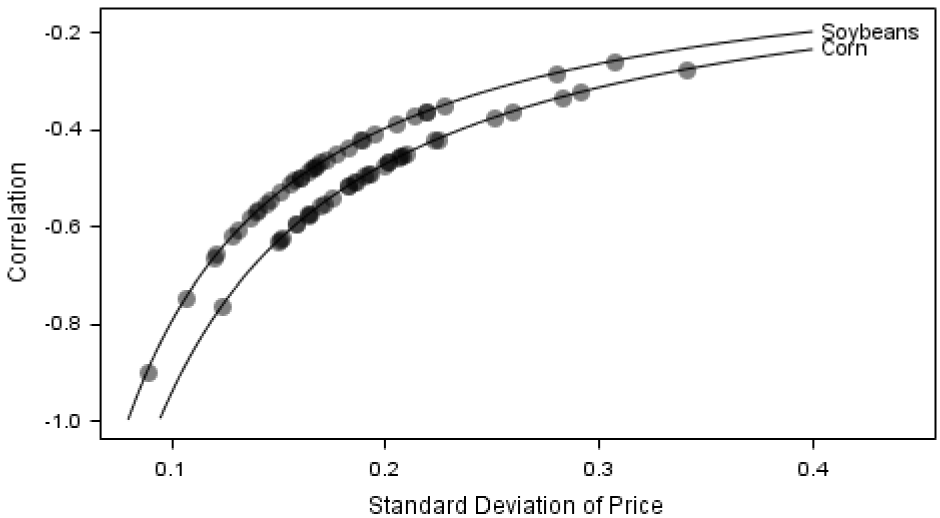

Figure 4: Response of correlation to changes in the standard deviation of price. Note: Each dot represents one year's observed standard deviation.

Another illustration of how sensitive correlation is to the variance of price is shown in figure 4, where the correlation is plotted against the standard deviation of price again for corn and soybeans. This shows the nonlinear relationship between the correlation and the variance of price in this framework. In the figure, each dot represents the standard deviation of price in one year. Larger price variance implies there is a larger amount of uncertainty about the number of acres or demand, which implies that a smaller amount of the variation in prices will be explained by variations in yield. Also interesting beause correlation may be no less than -1, a lower limit for the standard deviation of price given a constant variance of yields is implied, about 0.095 for corn and 0.08 for soybeans. This can be thought of as the variance of price if there was no uncertainty about the number of acres and the level of demand. Note also that for standard deviations of price less than around 0.15, the slope of the curve is very steep. This means that the correlation is very sensitive at lower price variances.

The larger the price variance, the less sensitive correlation is to changes in the variance of price.

Since correlation is quite sensitive to changes in the standard deviation of price, the joint distribution between price and yield will also be sensitive to these changes. If the year of interest has a relatively low price variance, then correlation as currently used by the RMA will be show too weak a relationship between price and yield. Conversely if the variance of price is relatively high, it will show too strong a relationship between price and yield. This shows that it is important to account for changes in correlation across time when determining the joint distribution of price and yield.

4.2 The Joint Distribution of Price and Yield

In this section, I evaluate the differences in the joint distributions of price and yield that are derived using the three methods that have been discussed. Those methods are 1) the current methods used by the RMA 2) the methods developed in this paper assuming prices are log-normally distributed and 3) the methods developed in this paper that do not assume prices are log-normally distributed, denoted RMA, LN, and non-LN, respectively.

Again, to simplify the analysis, I assume the variance of log-yield is constant over time. The distribution of yields must be specified for all three methods, and for the non-LN method, a distribution must be specified for $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$. Following Goodwin and Hungerford [2014] I specify a Weibull distribution for the mean centered yields. Since there is no evidence or theory to motivate the use of any specific distribution for $\log(\alpha_t) - \log(f_t(E[P_t|I_{\text{Planting}}]))$, I specify it to follow a normal distribution.

In the following, I evaluate the joint distribution of price and yield for a variety of possible price variance scenarios. Here I present three possible standard

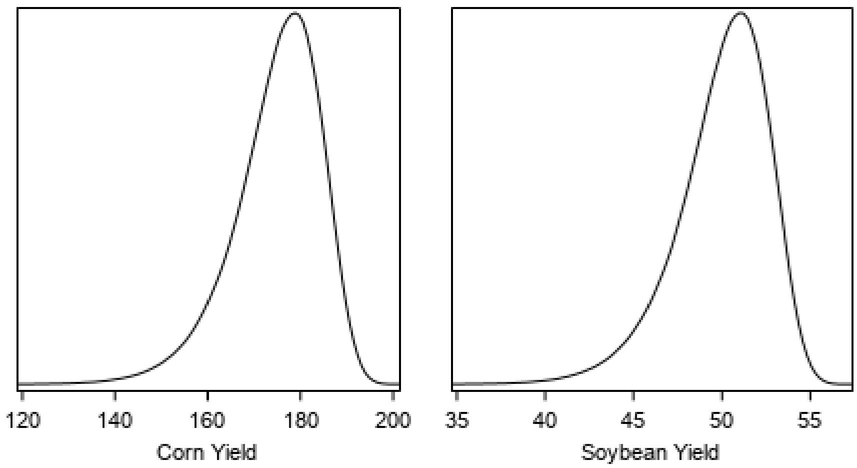

Figure 5: Density of Corn and Soybean Yield

deviations of price, a lower end scenario, the price volatility factor used by the RMA in 2020, and an upper end scenario. Figure 5 shows the marginal Weibull density of corn and soybean yields centered over the expected yield in 2020, detrended using a quadratic time trend in an OLS regression. This shows how the yields are negatively skewed. Also, it is important to note that the same density of yield is used in the following for each of the three methods.

Figure 6 shows the density of price evaluated using each of the three methods (RMA, LN, and non-LN) and for several possible standard deviations of price. This shows that the density of price is the same under the RMA and LN methods but not the non-LN method. Also, it shows that the difference in these two densities depend of the level of price variance. As the level of price variance increases, the density under the non-LN method converges to that of the RMA and LN method. This is because the distribution of yield, which is negatively skewed, influences the density of price by a larger amount when the overall price variance is low. When the variance of price is high the density of price is dominated by the term comprising the demand shifter and the number of acres

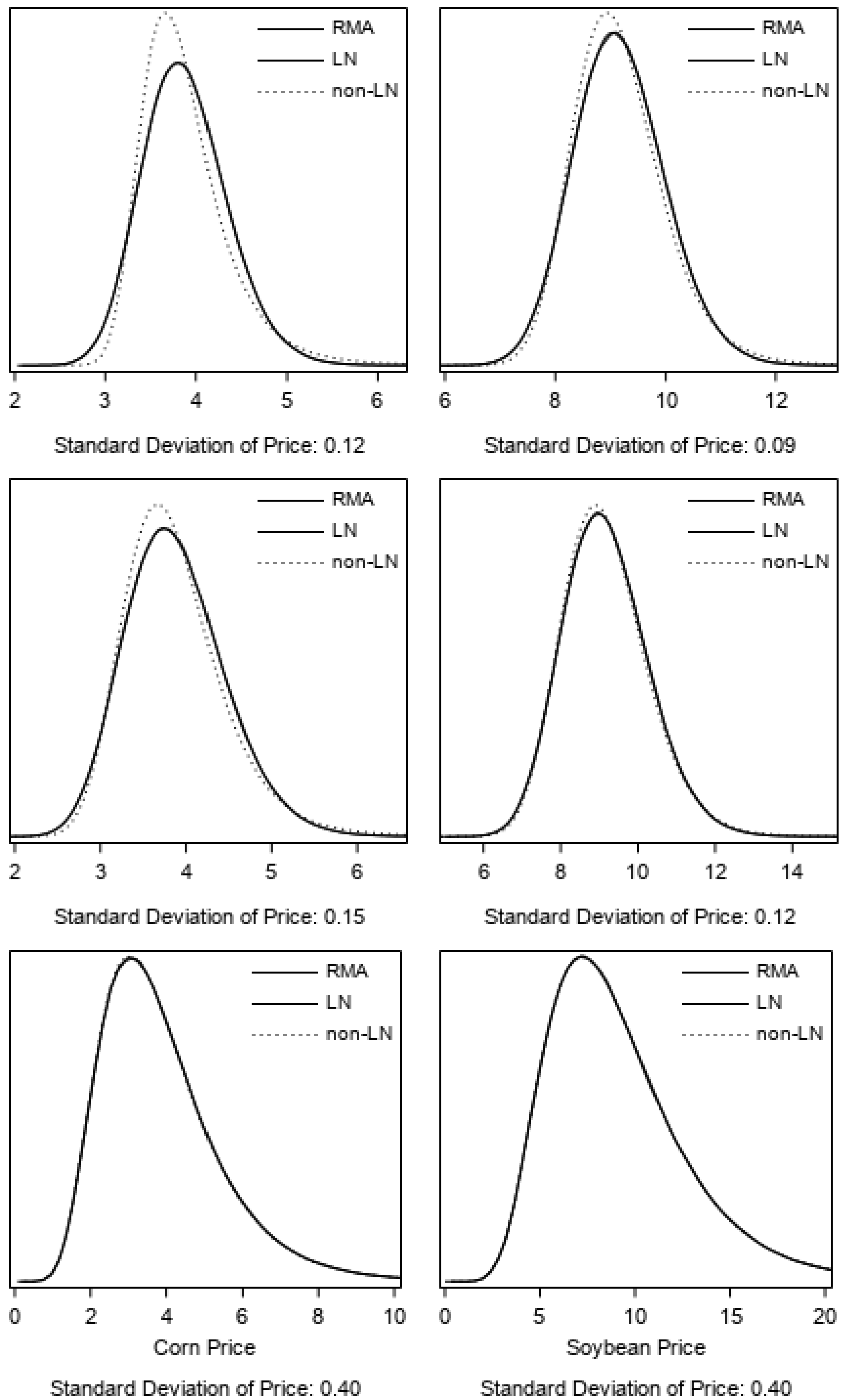

Figure 6: Density of Corn and Soybean Price under several possible standard deviations of price

which is assumed to follow a normal distribution. This implies that a lognormal specification for price is a good approximation only when the variance of price is relatively high.

Figure 7 shows the density of revenues again evaluated using each of the three methods and for several possible standard deviations of price. In this figure there is a different density found by each of the three methods. Note that the revenue density found by the LN and non-LN method has a much smaller spread than that found by the RMA method when the standard deviation of price is low. This is because the LN and non-LN methods impose a larger correlation between yield and price when the standard deviation of price is low. The opposite is true when the standard deviation of price is high, but to a less extreme extent. This is due to the shape of the curve presented in figure 4. At lower standard deviations of price the sensitivity of correlation to changes in the standard deviation of price is high and at at higher standard deviations of price that sensitivity is low. Thus it would take some extremely large price variances for the spread of the density of revenues found under the LN and non-LN methods to diverge by a large amount from the density found using the RMA method.

Tables 1a and 1b show the actuarially fair revenue insurance premiums calculated by each of the three methods across a range of price variance scenarios and revenue coverage levels. This shows that for low price variance scenarios, the premium rates calculated using the LN and non-LN methods are lower than the premium rates calculated using the RMA method. The converse is true for high price variance scenarios. The difference between the RMA and LN rates are soley due to the different correlations imposed by the two methods. The difference in teh RMA and non-LN rates are due to the differnt correlations imposed *and* the difference in the difference in the shape of the distribution

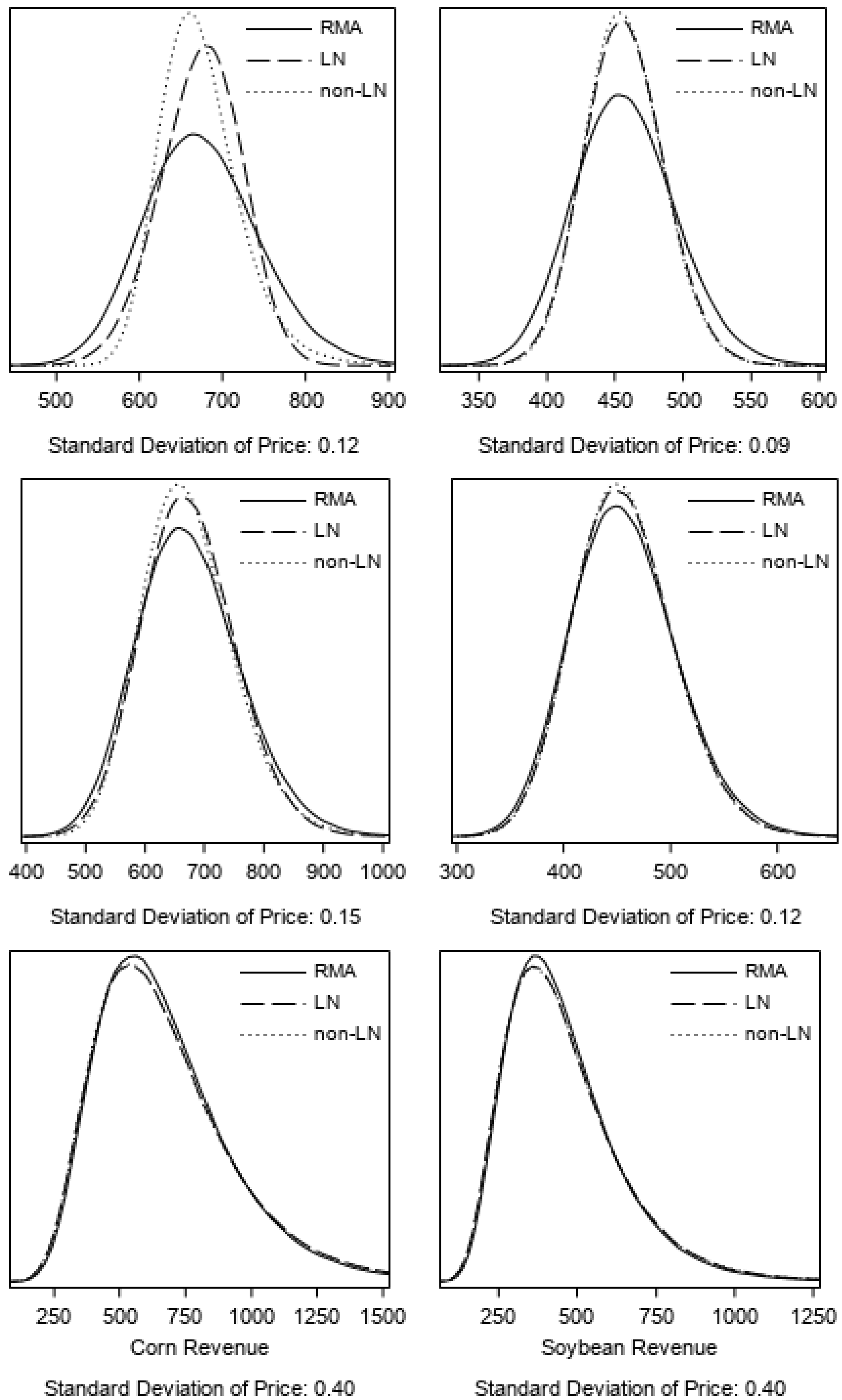

Figure 7: Density of Corn and Soybean Revenue under several possible standard deviations of price

Table 1a: Actuarially fair revenue insurance rates

Commodity	Std Dev of Price	Coverage Level	Method	Probability of Loss	Premium
Corn	0.12	95	RMA	0.313	12.47
			LN	0.240	7.45
			non-LN	0.249	5.54
		85	RMA	0.058	1.39
			LN	0.024	0.48
			non-LN	0.004	0.05
		75	RMA	0.002	0.02
			LN	0.000	0.00
			non-LN	0.000	0.00
	0.15	95	RMA	0.361	19.05
			LN	0.342	16.45
			non-LN	0.356	16.20
		85	RMA	0.113	3.89
			LN	0.092	2.89
			non-LN	0.087	2.38
		75	RMA	0.012	0.27
			LN	0.008	0.16
			non-LN	0.004	0.07
	0.20	95	RMA	0.416	31.08
			LN	0.415	30.97
			non-LN	0.421	30.99
		85	RMA	0.202	10.67
			LN	0.201	10.61
			non-LN	0.202	10.35
		75	RMA	0.055	1.97
			LN	0.055	1.95
			non-LN	0.052	1.78
	0.30	95	RMA	0.480	56.31
			LN	0.483	58.53
			non-LN	0.487	59.00
		85	RMA	0.326	29.31
			LN	0.334	31.15
			non-LN	0.337	31.37
		75	RMA	0.172	11.34
			LN	0.182	12.47
			non-LN	0.184	12.52

Note: Coverage level is in terms of percentage of expected yield

Table 1b: Actuarially fair revenue insurance rates

Commodity	Std Dev of Price	Coverage Level	Method	Probability of Loss	Premium
Soybeans	0.09	95	RMA	0.236	4.74
			LN	0.181	2.69
			non-LN	0.176	2.47
		85	RMA	0.022	0.26
			LN	0.006	0.05
			non-LN	0.004	0.03
		75	RMA	0.000	0.00
			LN	0.000	0.00
			non-LN	0.000	0.00
	0.12	95	RMA	0.303	8.48
			LN	0.294	7.73
			non-LN	0.294	7.64
		85	RMA	0.065	1.13
			LN	0.055	0.90
			non-LN	0.054	0.85
		75	RMA	0.003	0.04
			LN	0.002	0.02
			non-LN	0.002	0.02
	0.20	95	RMA	0.407	20.79
			LN	0.412	21.86
			non-LN	0.413	21.85
		85	RMA	0.205	7.49
			LN	0.215	8.19
			non-LN	0.214	8.17
		75	RMA	0.065	1.64
			LN	0.072	1.93
			non-LN	0.072	1.92
	0.30	95	RMA	0.474	37.81
			LN	0.476	39.35
			non-LN	0.477	39.43
		85	RMA	0.327	20.18
			LN	0.334	21.51
			non-LN	0.335	21.55
		75	RMA	0.184	8.52
			LN	0.195	9.43
			non-LN	0.195	9.43

Note: Coverage level is in terms of percentage of expected yield of revenue. The differences in premiums across the LN and non-LN are soley due to the different shapes of the revenue distributions because they both have the same correlations imposed between price and yield. Also note that those differences are more pronounced at lower coverage levels because the differences in the distributions are more pronounced in the tails. These results suggest that the new methods proposed in this paper have the potential to improve on the current methods used by the RMA in revenue insurance rating.

5 Conclusion

In previous literature, the correlation between price and yield deviates is typically used as the measure of the relationship between price and yield. In this paper, I develop a framework that allows a rigorous investigation of that relationship based on two primary assumptions. Under this framework I have shown that the correlation is a poor measure of the relationship between price and yield. A more appropriate measure of that relationship is the covariance. Also, this framework has implications for the joint distribution of price and yield. Prices are typically assumed to follow a log-normal distribution, but under this framework I show that is not necessarily the case. Also, there are several opportunities for future research topics suggested by this paper.

Although important for several reasons, there has been little previous research on the measurement of the relationship between price and yield. Of that research, only empirical techniques have been used in evaluations of the relationship between price and yield (Goodwin and Hungerford [2014];Paulson and Babcock [2008]; and Ramsey et al. [2019]). In a case such as this one where data are scarce and the theory is tractable, it is especially useful to use the economic theory to inform the empirical methods. Fortunately, the structure of these markets have certain characteristics that – with some assumptions – allow for a simple and tractable structural model of price and yield. This structural model facilitates the estimation of the joint distribution of price and yield that has a distinct advantage over current practice in that it allows the correlation of price and yield to change depending on the overall level of price variance. I also identify a new way to estimate the elasticity of demand for field crops which may be of interest to agricultural economists in general. The validity of the results presented here rely on these limiting assumptions. I do not wish to argue the assumptions hold true in these markets, but that it is important to create a framework that may be used to precisely identify what assumptions are needed to make meaningful inference.

It is important to emphasize the unique characteristic of the market for field crops that allow for such a simple model. Quantity supplied is the product of acres and yield – or quantity supplied is log linear in yields. Thus, if quantity demanded is log linear in prices, then yield must be log linear in price as well. The specification of quantity supplied is known – the only specification required is that for quantity demanded. This may not be the most appropriate specification of demand for these markets. Future research should explore the implications of different demand specifications on the relationship between price and yield and what assumptions may be needed to measure that relationship. An investigation into different specifications of demand may lead to assumptions that are more reasonable than those needed in this paper.

This paper focuses on national level price and yield which ignores state and county level variations in yields and prices. It will be important for future research to explore ways to incorporate state or county level yield variations into this framework, and it should be noted that this additional layer of complexity is not trivial. There is an extensive literature on spatial dependence in yields and finding a way to incorporate these methods into the framework developed in this paper could prove useful.

Although not its specific focus, this paper has implications for the specification of the marginal distribution of price that goes against common practice. It provides a combination of theoretical and empirical evidence that prices for field crops are not log-normally distributed due to the skewed nature of the distribution of yields. This is important in its own right because the BS option pricing model is founded upon the assumption of log-normality. Again, there have been empirical investigations into whether this assumption holds in agricultural commodity markets as well as proposed alternative distributions, but no theoretical justifications have been provided to date. A thorough investigation of this is conducted in Goodwin et al. [2017], where they argue that although the underlying log-normal assumption is often violated, the BS model is still preferred relative to competing alternatives. Perhaps most notably, Warren Buffett [2009] argues that "The Black-Scholes formula has approached the status of holy writ in finance... the formula represents conventional wisdom and any substitute that I might offer would engender extreme skepticism." The theoretical evidence presented in this paper should certainly be met with skepticism for its departure from the assumptions of the BS model, but it is very reasonable that the distribution of yield should inform the distribution of price. Future research should explore whether the distribution of price implied by the economic theory in this paper is a better approximation of the true distribution than log-normal.

References

Barchart. Powergen software. `Barchart.com`, 2019.

Warren Buffett. Letter to shareholders of berkshire hathaway. *Inc."(annual re*, 2009.

Keith H Coble, Thomas O Knight, Barry K Goodwin, Mary Frances Miller, Roderick M Rejesus, and George Duffield. A comprehensive review of the rma aph and combo rating methodology: Final report. *Report prepared for the USDA Risk Management Agency by Sumaria Systems, Inc., US Department of Agriculture, Risk Management Agency, Washington, DC*, 2010.

Barry K Goodwin and Ashley Hungerford. Copula-based models of systemic risk in us agriculture: implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, 97(3):879–896, 2014.

Barry K Goodwin, Ardian Harri, Roderick M Rejesus, and Keith H Coble. Measuring price risk in rating revenue coverage: Bs or no bs? *American Journal of Agricultural Economics*, 100(2):456–478, 2017.

Sergio H Lence and Dermot J Hayes. Us farm policy and the volatility of commodity prices and farm revenues. *American Journal of Agricultural Economics*, 84(2):335–351, 2002.

Ruiqing Miao, Madhu Khanna, and Haixiao Huang. Responsiveness of crop yield and acreage to prices and climate. *American Journal of Agricultural Economics*, 98(1):191–211, 2016.

Robert J Myers and C Ford Runge. The relative contribution of supply and demand to instability in the us corn market. *North Central Journal of Agricultural Economics*, pages 70–78, 1985.

USDA National Agricultural Statistics Service NASS. Nass - quick stats. `https://data.nal.usda.gov/dataset/nass-quick-stats`, 2020.

Nicholas D Paulson and Bruce A Babcock. Get a grip: should area revenue coverage be offered through the farm bill or as a crop insurance program? *Journal of Agricultural and Resource Economics*, pages 137–153, 2008.

Roley R Piggott. Decomposing the variance of gross revenue into demand and supply components. *Australian Journal of Agricultural Economics*, 22(2-3): 145–157, 1978.

A Ford Ramsey, Barry K Goodwin, and Sujit K Ghosh. How high the hedge: Relationships between prices and yields in the federal crop insurance program. *Journal of Agricultural and Resource Economics*, 44(1835-2019-1539): 227–245, 2019.