# Agribusiness Analysis and Forecasting

## Introduction to Forecasting and Simulation

Henry Bryant

Texas A&M University

Spring, 2020

# Course is about Decision Making

How do you make decisions?

- Do you list all of the possible outcomes and their consequences?
- Do you think about the probabilities for each possible outcome? $P(win) = 25\%$
- Do you fixate on the outcome you want? $P(win) = 100\%$
- Do you just go through life without planning?

# Risk and Decision Making

- An agent is subject to *risk* when the future realization of some random variable affects their utility, and the probability distribution of that random variable can be reasonably quantified. For example, the amount of rainfall tomorrow, or the S&P 500 closing value next week.
- You are constantly faced with decisions and risk.
- In business, simulation is the tool used to analyze risky decisions.
- You already use simulation when you think through possible outcomes for a decision, and the likelihoods of those various outcomes.

# Relationship Between AGEC-622 and Your Previous Courses

- Statistics and AGEC 621
  - Probabilities, probability distributions
  - Mean, Variance, Covariance, Correlation
  - Statistical tests and significance
  - Multiple regression analysis
- Accounting, Finance, Business
- Hopefully you know how to use some of these tools.
- We will be using these tools in a systematic way to forecast risky variables and those forecasts to make better decisions.

# Simulation and Ag Economics

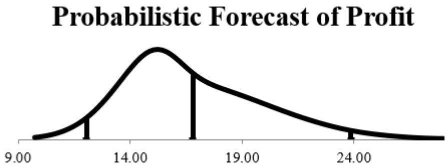
- Common theme is – **RISK** and **MONEY**.
- Example: simple profit equation for a single-product firm.
  - If no risk the equation produces one number.

$$Profit = (P * Y) – FC – (VC * Y)$$

  - In a risky world there are many outcomes.

$$\tilde{Profit} = (\tilde{P} * \tilde{Y}) – FC – (\tilde{VC} * \tilde{Y})$$

  - Here, the yield $(Y)$, price $(P)$, and variable cost $(VC)$ are stochastic.



# Application of Risk in a Decision

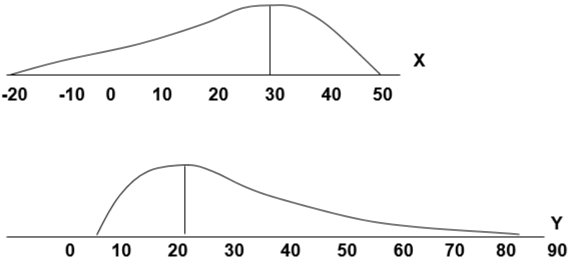
- Given two investments, $X$ and $Y$ with no risk.
  - Return for $X$ is 30%
  - Return for $Y$ is 20%

- Obviously we choose $X$

- If there is a risk involved, then the answer is not as simple anymore.

# Application of Risk in a Decision

- Given two investments, $X$ and $Y$ *with* risk.
  - *Expected* return for $X$ is 30%
  - *Expected* return for $Y$ is 20%
- What if the distributions of returns are known?




- Using simulation we can estimate the shape of the distribution for returns (or profits) for risky investments.

# Simulation Models and Decision Making

- Start with a problem that includes risk.
- Identify risky variables and control variables you can change (scenarios).
- Gather data, estimate regressions, develop equations for a model of the problem. Often, these are accounting identities. For example:

$$\text{Profit} = \text{Revenue} - \text{Variable Costs} - \text{Fixed Costs}$$

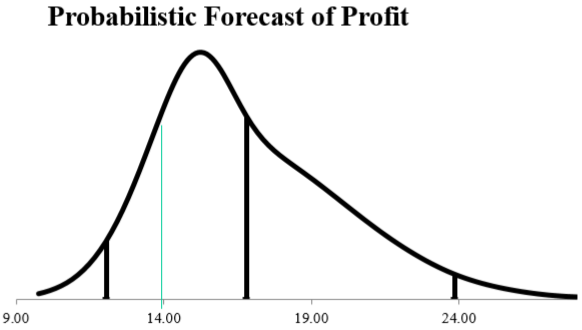
- Simulate model under risk.
- Analyze results and decide on your preferred values for the control variables.

# Purpose of Simulation

- To <u>approximate</u> probability distributions that we <u>can not observe</u> and apply them to economic analysis of risky alternatives (strategies) so the decision maker can make <u>better decisions</u>.

$$\tilde{Profit} = (\tilde{P} * \tilde{Y}) – FC – (\tilde{VC} * \tilde{Y})$$

$$Probability(P > 14)$$



# Forecasting and Simulation

- Forecasters generally give only a point estimate of a variable.
- Because we use simulation, we will develop and report probabilistic forecasts.
    - We will include risk in our forecasts for business decision analysis.
    - Simply this involves adding more than a simple confidence interval about the forecast.
    - Also, it is harder to be proven wrong if you give a range with a probability about the center point.

# Simulate a Forecast

Two components to a probabilistic forecast. AGEC 621 taught you how to develop a deterministic forecast based on a linear model.

1. **Deterministic Forecast**. For example,

$$\hat{Y}_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i$$

2. The **Stochastic Component** $\tilde{e}$ is generally ignored

$$\tilde{e}_i = Y_i - \hat{Y}_i$$

3. The **complete probabilistic forecast** model becomes

$$\tilde{Y}_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \tilde{e}_i$$

# Steps for Probabilistic Forecasting

Steps for probabilistic forecasting

1. Estimate the best econometric model to explain trend, seasonal, cyclical, structural variability to get $\hat{Y}_{T+1}$.
2. Residuals $\tilde{e}$ are the unexplained variability or risk.
3. We may often assume that the residuals are distributed normal. Simulate the risk in Simetar as: $\tilde{e} = NORM(0, \sigma_e)$
4. Probabilistic forecast is:
$$\tilde{Y}_{T+1} = \hat{Y}_{T+1} + \tilde{e}$$
or in Simetar we can use
$$\tilde{Y}_{T+1} = NORM(\hat{Y}_{T+1}, \sigma_e)$$
or for more complete information
$$\tilde{Y}_{T+1} = \hat{Y}_{T+1} + \sigma_e * NORM(0, 1)$$
5. Simulate 500 possible future values for $\tilde{Y}_{T+1}$, and report the range and probability about the forecasted mean.

# Major Activities in Simulation Modeling

- Estimating parameters for a forecasting model

$$\tilde{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{\beta}_2 Z_t + \tilde{e}_t$$

- Risk can be simulated with different distributions, e.g.
  - $\tilde{e} = NORMAL(Mean, St.Deviation)$
  - $\tilde{e} = BETAINV(Alpha, Beta, Min, Max)$
  - $\tilde{e} = Empirical$(Sorted Values and their probabilities) and so on.
- We estimate the parameters $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$, recover the residuals $\hat{e}$ and specify the distribution for $\tilde{e}$.
- Simulate random values from the distribution $\tilde{e}$ and validate that simulated value comes from the parent distribution.
- Apply the model to analyze risky alternatives and decisions.
  - Statistics and probabilities.
  - Charts and graphs (PDF, CDF, Stoplight).
  - Rank risky alternatives.

# Role of a Forecaster

- Analyze historical data series to quantify the patterns for the variable.
- Extrapolate the pattern into the future for a forecast using quantitative models.
- In the process, become an expert in the industry so you can identify structural changes before they are observed in the data – incorporate new information into forecasts. Think carefully and look for unexpected.

# Types of Forecasts

Three types of forecasts

1. Point or deterministic forecast

$$\hat{Y} = 10.00$$

2. Range forecast

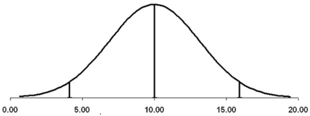
$$\hat{Y} = [8.0, 12.0]$$

3. Probabilistic forecast

$$\hat{Y} = [8.0, 12.00]$$

$$P(8 <= \hat{Y} <= 12) = 50\%$$

$$P(4 <= \hat{Y} <= 16) = 95\%$$




Forecasts are never perfect so simulation is a way to present the most possible outcomes. In other words, it is harder to prove your forecast is wrong when you use a simulation approach.

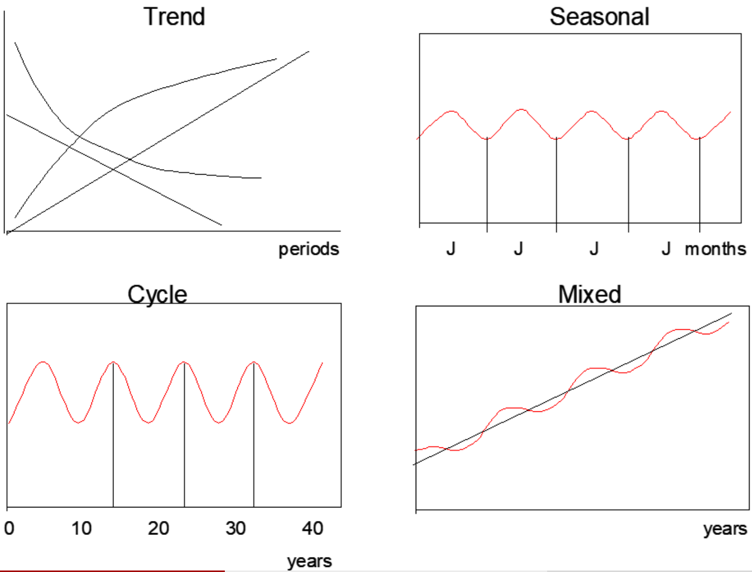
# Forecasting Tools in AGEC-622

- Trend regression, including linear and non-linear.
- Multiple regression.
- Seasonal analysis.
- Cyclical analysis.
- Autoregressive process.
- Finance-style stochastic processes.
- Multivariate forecasting.

# Define Data Patterns

- A time series is a chronological sequence of observations for a particular variable, typically over fixed intervals of time.
  - Daily
  - Weekly
  - Monthly
  - Quarterly
  - Annual
- Six patterns for time series data (data we work with is time series data because use they are generated over time).
  - Trend
  - Cycle
  - Seasonal variability
  - Structural variability
  - Irregular variability
  - Black Swans

# Patterns of Time Series Data



# Simplest Forecast Method

- The Mean is the simplest forecast method.
- Deterministic forecast of a fixed mean.

$$\hat{Y} = \bar{Y} = \frac{\sum Y_i}{N}$$

- Forecast error for a mean forecast is a residual.

$$\hat{e}_i = Y_i - \bar{Y}$$

  or

$$\hat{e}_i = Y_i - \hat{Y}$$

  for a conditional mean
- Standard deviation of the residuals is the measure of the error (or risk) for the forecast.

$$\sigma_e = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{N-1}}$$

- Range forecast of: $\hat{Y} \pm \sigma_e$
- Probabilistic forecast, where $\tilde{e}$ represents the risk. $\tilde{Y} = \hat{Y} + \tilde{e}$ so the forecast can be represented as $P(8 <= \hat{Y} <= 12) = 50\%$.

# Trend Forecast

- A trend is a general up or down movement in the values for a variable over time.
- Economic data often contains at least one trend
- A trend often represents some form of long-term growth or decay.
- Trends can caused by many different underlying forces, such as:
  - Technological changes, e.g., crop yields
  - Change in tastes and preferences
  - Change in income and population
  - Market competition
  - Inflation and deflation
  - Policy changes

# Linear Trend Forecast Models

- Linear trend model (no risk)

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} T_t$$

where $T_t$ is time trend and is represented as: $T = 1, 2, 3, ...$ or $T = 1985, 1986, .., 2019$.

- Estimate parameters for forecast model using OLS.
- Probabilistic forecast of the variable with a linear trend becomes

$$\tilde{Y}_t = \hat{Y}_t + \tilde{e}$$

which is rewritten using a *Normal Distribution* for $\tilde{e}$

$$\tilde{Y}_{t+i} = NORM(\hat{Y}_{t+i}, \sigma_e)$$

where $t$ reflects the last historical observation and $\sigma_e$ is the standard deviation of the residuals.

# Non-Linear Trend Forecast Models

- Polynomial trend model

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 T_t + \hat{\beta}_2 T_t^2 + \hat{\beta}_3 T_t^3$$

where $T_t$ is the time variable and is represented as:

$$T = 1, 2, 3, \ldots$$

$$T^2 = 1, 4, 9, \ldots$$

$$T^3 = 1, 8, 27, \ldots$$

Estimate parameters (beta-hats) using OLS

- Probabilistic forecast from trend becomes again

$$\tilde{Y}_{t+i} = NORM(\hat{Y}_{t+i}, \sigma_e)$$

# Steps to Develop a Trend Forecast

- Plot the data.
  - Identify linear or non-linear trend.
  - Generate $T$, $T^2$, $T^3$ if necessary.

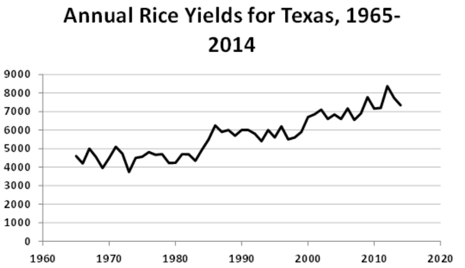
- Estimate regression model using OLS.
  - Observing a low $R^2$ is common.
  - $F$ ratio and $t$-test will be significant if a trend is statistically present.

- Simulate model using:

$$\tilde{Y}_{t+i} = \hat{Y}_{t+i} + NORM(0, \sigma_e)$$

- Report probabilistic forecast.

# Linear Trend Model Forecast



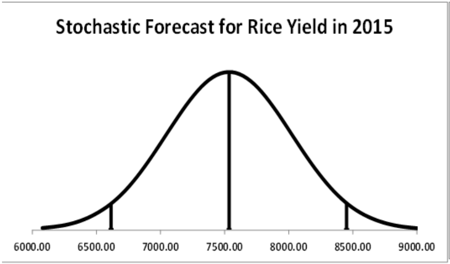

| Texas Rice Yield | |
|---|---|
| Mean | 5698.82 |
| StDev | 1148.803 |
| 95 % LCI | 5323.377 |
| 95 % UCI | 6074.263 |
| Min | 3740 |
| Median | 5650 |
| Max | 8370 |
| Skewness | 0.305479 |
| Kurtosis | -0.83617 |

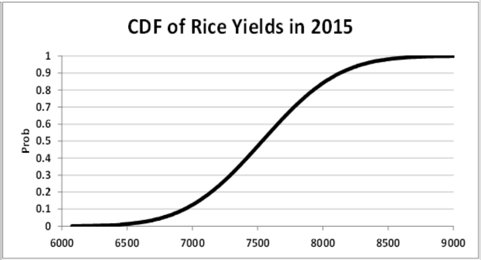
| Simple Trend Regression | |
|---|---|
| | Texas Rice Yield |
| Intercept | -137566 |
| Slope | 72.01061 |
| R-Square | 0.83495 |
| F-Ratio | 242.8215 |
| Prob(F) | 2.09E-20 |
| S.E. | 4.62118 |
| T-Test | 15.58273 |
| Prob(T) | 1.28E-20 |
| S.D. Residuals | 466.7163 |

| Simple Deterministic Forecast | | | | Trend Forecast of the Means | | | | |
|---|---|---|---|---|---|---|---|---|
| Mean | 5698.82 | =B69 | | | Y-Hats | Stochastic | | Stoch Y-Hats |
| | | | | 2015 | 7535.091 | -522.1413 | 2015 | 7012.949 |
| Range Forecast using mean and std deviati | | | | 2016 | 7607.101 | 304.9594 | 2016 | 7912.061 |
| Upper Range | | 4550.017 | =B69-B70 | 2017 | 7679.112 | 775.4362 | 2017 | 8454.548 |
| Mean | | 5698.82 | =B69 | 2018 | 7751.122 | 132.7596 | 2018 | 7883.882 |
| Lower Range | | 6847.623 | =E73+B70 | 2019 | 7823.133 | -893.1347 | 2019 | 6929.998 |
| | | | | 2020 | 7895.144 | -205.5286 | 2020 | 7689.615 |
| | | | | | =$B$81+$B$82*G75 | =G75 | | =H75+I75 |
| | | | | | | =NORM(0,$B$89) | | |

# Linear Trend Model Forecast

| Simetar Simulation Results for 500 Iterations. | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Linear!K70 | Linear!K71 | Linear!K72 | Linear!K73 | Linear!K74 | Linear!K75 |
| Mean | 7535.018733 | 7606.887246 | 7678.886251 | 7751.251965 | 7822.910474 | 7895.242295 |
| StDev | 467.4026346 | 466.5737346 | 466.1773899 | 465.3475314 | 467.5138607 | 466.6523975 |
| CV | 6.203071965 | 6.133569745 | 6.070898495 | 6.003514445 | 5.976213869 | 5.910551951 |
| Min | 6077.51955 | 6104.282335 | 6285.908085 | 6407.642608 | 6199.439017 | 6501.259321 |
| Max | 9002.216388 | 8970.753819 | 9074.013636 | 9114.930194 | 9284.876319 | 9314.504986 |
| Iteration | 2015 Stoch Y-Hats | 2016 Stoch Y-Hats | 2017 Stoch Y-Hats | 2018 Stoch Y-Hats | 2019 Stoch Y-Hats | 2020 Stoch Y-Hats |



# Non-Linear Trend Regression

Estimated equation of

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 T + \hat{\beta}_2 T^2 + \hat{\beta}_3 T^3$$

to capture the non-linear trend in the data.

[demo]

# Is a Trend Forecast Enough?

- A seasonal pattern may mask a trend, so the final model will need both trend and seasonal terms.
- Cyclical or structural variability may mask the trend so need a more complex model.
- Bottom line is that trend may be where we start, but we can generally benefit from a more complex model.

# Types of Forecast Models

- Two types of models:
  1. Causal or structural models.
  2. Univariate (time series) models.
- **Structural models** identify the variables ($Xs$) that explain the variable ($Y$) we want to forecast, the residuals are the unexplained fluctuations we simulate.

  $$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 P_y + \hat{\beta}_2 P_x + \hat{\beta}_3 Z + \hat{\beta}_4 T + \tilde{e}$$

  Ex: Demand models have own and cross prices, income, population, and other variables for tastes and preferences (often includes trend).
- **Univariate models** forecast using past observations of the same variable.
  - Advantage is you do not have to forecast the structural variables.
  - Disadvantage is no structural equation to test alternative assumptions about policy, management, and structural changes.

  $$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 Y_{t-1} + \hat{\beta}_2 Y_{t-2} + \tilde{e}$$

# Meaning of the CI and PI in Simetar

- Confidence intervals and prediction intervals are not the same thing and the terms are often confused.
- CI: Confidence interval calculations take sample data and produce a range of values that likely contains the population *parameter* (e.g., the conditional mean) that you are interested in. Because the data are random, the interval is random. A 95% confidence interval will contain the true parameter with probability 0.95. That is, with a large number of repeated samples, 95% of the intervals would contain the true parameter
- PI: Prediction intervals contain the value of the *dependent variable*, given specific settings of the independent variables, with some particular probability. Thus, a prediction interval is an interval associated with a random variable yet to be observed, with a specified probability of the random variable lying within the interval. For example, we may predict a 95% interval for the forecast of corn price in 2019/20. The actual corn price in 2019/20 should lie within the interval with probability 0.95.