NCCC-134

APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

# Herding in the USDA International Baseline Projections

by

Rabail Chandio and Ani Katchova

Suggested citation format:
Chandio, R. and A. Katchova. 2022. “Herding in the USDA International Baseline Projections.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. [http://www.farmdoc.illinois.edu/nccc134].

# Herding in the USDA International Baseline Projections

**Rabail Chandio, Ohio State University (chandio.1@osu.edu)**
**Ani L. Katchova, Ohio State University (katchova.1@osu.edu)**

Paper presented at the NCCC-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management, 2022.

**Abstract**

USDA's annual Agricultural Baseline Projections contribute significantly to agricultural policy in the United States, and hence their accuracy is vital. The baseline projections present a neutral policy scenario assuming a specific macroeconomic situation and allow the analyses of alternative policies and their micro and macroeconomic impacts in the United States. We investigate the trends and heterogeneity in the incidence of bias in the USDA International Baseline Projection reports from 2002 to 2021. The evaluation of bias as it varies geographically, temporally, and by crop-variable allows us to make inferential judgments about the sources of this bias. First, we use the dynamic time warping algorithm to examine whether experts tend to group together the projections for certain crops across different countries, producing similar projection trends. We find that projection series for all countries in the sample are correlated with the United States in their trends. Second, we compute the bias in projections and decompose it by projection horizon. Third, we assess whether the bias is higher across crops or across countries with more substantial evidence for grouping behavior and find that for soybeans imports, soybeans ending stocks, and wheat area harvested, similarity in projection trends with the United States lowers the bias while for most other crop-variables it increases it. This suggests that the projections for our sample countries are unnecessarily made to follow similar trends to the United States projections which proves to be a bias inducing choice in most cases.

# 1 Introduction

Due to its significant contribution to the United States agricultural policy, minimizing the bias in the USDA's annual Agricultural Baseline Projections is vital. The projections present a neutral policy scenario assuming a specific macroeconomic situation and allow the analyses of alternative policies and their micro and macroeconomic impacts in the United States. The baseline projections help evaluate local and foreign policy change scenarios and their subsequent implications for United States farmers (Skorbiansky, Childs, and Hansen, 2018; Langholtz et al., 2012). USDA also utilizes the baseline projections to inform revisions to Farm Bills and aid in annual Presidential budgeting. Therefore, any policy evaluations utilizing the baselines projections will be as useful and informative as the projections are accurate. Recently, academic research has started evaluating the accuracy of these projections, and has made initial discoveries about the incidence of bias and limited informativeness of the projections in certain cases (Bora, Katchova, and Kuethe, 2021b; Regmi et al., 2021; Isengildina-Massa et al., 2020; Kuethe, Hubbs, and Sanders, 2018). However, from a policy perspective, understanding the source of bias is essential for minimizing it and improving the projections, which has not received much attention in the literature.

To identify and decompose any present bias, we first need to understand how the baseline projections are prepared. Released each year by the USDA Interagency Agricultural Projects Committee, baseline projections combine model-based values and judgment-based adjustments to these values (USDA Agricultural Projections to 2030). Experts from various committees in USDA, including the Economic Research Service, World Agricultural Outlook Board, and the Office of Chief Economist, evaluate the model-based results and adjust them until a point of consensus is reached. We, however, only observe the finalized projection values, so we utilize a novel approach to identify the sources of bias.

In this study, we investigate the heterogeneity in the incidence of bias in the USDA International Baseline Projection reports from 2002 to 2021 where bias is the difference between the projected values and the realized values. The evaluation of bias as it varies geographically, temporally, and by crop-variable allows us to make inferential judgments about the sources of bias. We answer three main questions. First, we examine whether experts tend to group together the projections for certain crops across different countries (i.e. herding behavior), producing similar projection trends. Second, we compute the bias in projections and decompose it by projection horizon. Third, we assess whether the bias is higher across crops or across countries with more substantial evidence for grouping behavior. In-depth familiarity with the baseline procedures that comes from studying the documentation and literature allows us to define case-based hypothesis about any potential bias in the projections. While previous studies find the incidence of bias and hence lack of informativeness in the baseline projections (Bora, Katchova, and Kuethe, 2021a), we are not aware of any studies that considered the heterogeneity in bias which may help identify its source.

Herding is a behavioral phenomenon often observed in financial markets, when investors and experts with private information align their choices and decisions with others as a risk management strategy. It can be rational if the individuals make the choice to align their decisions based on superior private information, or it can be irrational if individuals ignore their private information in order to adopt similarity with others (Devenow and Welch, 1996). Behavioral finance research suggests that propensity to "herd" is a response to a private cost minimization strategy by individuals. Huang et al. (2017) find that the majority of people choose to follow the group consensus regardless of their individual prior beliefs as long as there is no significant cost of agreeing with the group opinion. Moreover, when the institution (such as USDA) or the forecasting team is considered a single entity, research shows that the forecast behavior of experts within the institution and/or a team are affected by and aligned with the overall beliefs of the institution/team and hence they herd together (Benchimol et al., 2020; Van Campenhout and Verhestraeten, 2010).

Whether it is rational or irrational, herding increases volatility in stock and commodity markets and is suboptimal for the market (Blasco, Corredor, and Ferreruela, 2012; Wang and Wang, 2018). Yet, instances of herding behavior are common in stock market investments, land valuations by experts, and financial forecasts by analysts, among other things (Trueman, 1994; Herzenstein, Dholakia, and Andrews, 2011).

Similarly, herding together the agricultural baseline projections among various countries and commodities can potentially be a suboptimal and bias inducing choice for the international baselines projections. Therefore, our study identifies a potential source of bias in the baselines projections. Furthermore, by specifically highlighting the crops, variables, and countries where this bias persists, we propose a cost-effective way to reduce these errors, and improve the accuracy and informativeness of the baselines projections.

Since simple correlation measures and auto-regressive lag models do not satisfactorily measure herding behavior in the projections of various countries, we employ time-series techniques novel to our field. Utilizing a rich time-series data where each time period nests another projection series (called path forecasts), we use the Dynamic Time Warping algorithm to assess the degree and rationality of herding in the baselines projections. The algorithm has been adapted to various fields in recent applications (Berndt and Clifford, 1994; Müller, 2007; Jeong, Jeong, and Omitaomu, 2011; Varatharajan et al., 2018). We identify the cases where herding is an irrational behavioral choice undertaken to minimize the risk of inaccuracy and estimate its contribution to the overall bias in the projections. We find that projections for all countries in our data are statistically significantly aligned with the United States in their trends for all of the crops and variables we look at. Moreover, for most crops and variables including corn, soybeans, and wheat total consumption, this correlation is associated with significantly higher errors in projections. Moreover, in the cases where the correlation is rational i.e. the realized values also depict a similarity in trends, it is often associated with lower errors in projections. Our findings have implications for the USDA baseline experts as well as government agencies and users of the baselines reports.

We make three main contributions to the literature. First, our study identifies that the projections for all countries included in the USDA International Baseline Projection reports are made correlated with the United States beyond what the realized values are for each country, which may inform USDA on another criterion that needs to be checked prior to releasing their projection reports. Second, we provide conclusive evidence that for other countries' projections of most crop specific variables, correlation with the United States is associated with higher bias/lower accuracy of these countries' projections. This informs the USDA baseline experts to assess the accuracy of the baseline models. If the excessively correlated projections are a result of the model, the model input or relations necessitate updating but if the model presented dissimilar projections that were smoothed to look similar later, this suggests that the changes may not be needed. Third, we recognize the heterogeneity in the relationship between projections' bias and projections' correlation.

By highlighting the variables for which correlation reduces the error in projections, we allow for higher accuracy in the projections. Overall, these insights can be incorporated by the team preparing the USDA baselines projections to minimize excessive similarity in the projections, which is decreasing their accuracy.

The remainder of the paper is organized as follows. Section 2 describes the USDA International Baseline Projections and the variables included in our study. Section 3 details the empirical strategy, which is followed by presentation and discussion of the results in section 4. Section 5 contains the concluding remarks.

## 2 Data

We use the official USDA International Baseline Projections data from 2002 to 2021 which includes 10-year domestic (United States) and international (other countries and regions) projections for several crops each year. We limit our analysis to corn, soybeans, and wheat for the variables area harvested, yield, imports, exports, ending stocks, and total consumption for each year since 2002. Balance sheet equation dictating the relationship of the variables we study is as follows

$$\begin{aligned} &BeginningStocks + Production + Imports \\ &= Exports + TotalConsumption + EndingStocks \end{aligned} \tag{1}$$

where the $BeginningStocks_t = EndingStocks_{t-1}$, making it a redundant variable, and $Production = AreaHarvested \times Yield$. We focus only on the variables that are identified independently.

The available baselines data also include the realized values for up to three years before the release date of the reports. We utilize these limited historical data in each year's report to construct an annualized panel data set for "actual" or "realized" values that are used for bias calculations and accuracy evaluations of the projections.

The baseline projections have a structure which is statistically referred to as nested time-series (or path-forecasts) data, where each year nests the series of ten incremental horizon projections. A representative projection $\hat{Y}_{rcvt}$ is the projection series for country $r$ (belonging to an unbalanced panel of 34 countries observed annually over the study period), for crop $c \in$ {*corn, soybeans, wheat*}, variable $v \in$ {*yield, area harvested,* *imports, exports, total consumption, ending stocks*}, and report year $t \in \{2002, ..., 2022\}$. $\hat{Y}_{rcvt}$ is a series that has a length of $H = 10$, where $h$ represents the different projection horizons such that $\hat{Y}_{rcvt} = (\hat{Y}_{h_0}, \hat{Y}_{h_1}, ..., \hat{Y}_{h_9})$.

# 3 Methods

There are three main components of our empirical analysis. First, we estimate the degree of similarity among various countries' baseline projections using the dynamic time warping (DTW) algorithm. Second, we compute the errors in historical projections and assess the size of bias for each crop and country in the USDA projections where bias is defined as the difference between the projected values and the actual realized values. Finally, we use regression analysis to study the relationship between the degree of herding and the size of bias in the projections.

## 3.1 Evaluating the Degree of Similarity

We begin our analysis by evaluating the differences in projections of specific countries for each crop, variable, report year, and projection horizon to estimate the degree of herding. We use a dynamic time warping algorithm to compute the distance between each set of projection series for each crop-variable-year-horizon combination (for instance, "corn yield for report year 2010, and the entire projection horizon of 10 periods" OR "soybean area harvested for report year 2015") and evaluate whether the projections exhibit similar trends. The algorithm finds the minimum distance needed to make two time-series as similar as possible. We use this method to compute the distances between all available country pairs for each crop-variable-year-horizon to determine the closest projection "neighbors" of the top producers for each crop with the closest distance among all countries.

We suppress the indices *cvt* since they will remain the same for each pair of countries whose projections are being compared. To determine the distance between the projections for any two countries $\hat{Y}_{r_i}$ and $\hat{Y}_{r_j}$, we define the two time-dependent series $\hat{Y}_{r_i}$ and $\hat{Y}_{r_j}$ and compute an expansive local cost matrix ($LCM$) between them. The $LCM$ is populated by pairwise comparison of each horizon's projections for $\hat{Y}_{r_i}$ with each horizon's projections of $\hat{Y}_{r_j}$, resulting in a square matrix of dimensions $10 \times 10$ since $length(\hat{Y}_{r_i}) = length(\hat{Y}_{r_j}) =$

10 for the 10 horizons. The $LCM$ matrix is defined as:

$$LCM(\hat{Y}_{r_i}, \hat{Y}_{r_j}) = \begin{pmatrix} d_{\hat{Y}_{r_i h_0}, \hat{Y}_{r_j h_0}} & d_{\hat{Y}_{r_i h_1}, \hat{Y}_{r_j h_1}} & \cdots & d_{\hat{Y}_{r_i h_0}, \hat{Y}_{r_j h_9}} \\ d_{\hat{Y}_{r_i h_1}, \hat{Y}_{r_j h_0}} & d_{\hat{Y}_{r_i h_1}, \hat{Y}_{r_j h_1}} & \cdots & d_{\hat{Y}_{r_i h_1}, \hat{Y}_{r_j h_9}} \\ \vdots & \vdots & \ddots & \vdots \\ d_{\hat{Y}_{r_i h_9}, \hat{Y}_{r_j h_0}} & d_{\hat{Y}_{r_i h_9}, \hat{Y}_{r_j h_1}} & \cdots & d_{\hat{Y}_{r_i h_9}, \hat{Y}_{r_j h_9}} \end{pmatrix}$$

where each matrix element $d_{\hat{Y}_{r_i h_a}, \hat{Y}_{r_j h_b}} = \sqrt{|\hat{Y}_{r_i h_a} - \hat{Y}_{r_j h_b}|^2}$ denotes the Eucleadean distance between $a^{th}$ and $b^{th}$ horizon projections of series $\hat{Y}_{r_i}$ and, $\hat{Y}_{r_j}$ respectively and $a, b \in h$.

We find the distance between the two projection series by defining $\phi(k)$ to be the path from $d_{\hat{Y}_{r_i h_0}, \hat{Y}_{r_j h_0}}$ to $d_{\hat{Y}_{r_i h_9}, \hat{Y}_{r_j h_9}}$ where $k = (1, 1), ..., (h, h)$. For a given path $\phi$, we compute the Euclidean distance/dissimilarity between the projections for two countries $Y_{r_i}$ and $Y_{r_j}$ as

$d_\phi(\hat{Y}_{r_i}, \hat{Y}_{r_j}) = \sum_k [(LCM(k)) \times m_\phi(k)]$

where $m_\phi(k)$ is the per-step weighting coefficient, allowing to only add the cost/distance that falls on the path $\phi(k)$. Next, we find a path $\phi$ within the $lcm$ matrix that gives the minimum total distance between the projections in two countries $d_\phi(\hat{Y}_{r_i}, \hat{Y}_{r_j})$. Hence, we use the the DTW algorithm to find the minimum distance between the two country's projections for the same horizon by solving the following optimization problem:

$$DTW(\hat{Y}_{r_i}, \hat{Y}_{r_j}) = min_\phi(d_\phi(\hat{Y}_{r_i}, \hat{Y}_{r_j})) \tag{2}$$

We impose two constraints on the minimization problem to avoid meaningless loops or inefficient paths. First, we require monotonicity which restricts the direction of the path taken within the $lcm$ to only increasing projection horizon for at least one of the projection series:

$$\phi : \phi\,(k+1) \geq \phi(k) \tag{3}$$

Second, we impose the Sakoe-Chiba window constraint to restrict the path that can be taken to a window $h^{window}$ around the principal diagonal of the $lcm$. We use a window $h^{window}$ as the only valid points of the $lcm$ where a path $\phi$ can be made in the range

$$|\left(a, b - h^{window}\right), \left(a, b + h^{window}\right)| \tag{4}$$

for all $(a, b)$ along the principal $LCM$ diagonal. This constraint allows only reasonable lead and lag relationships to be considered while two projection series are being compared. We set $h^{window} = 2$ restricting $\hat{Y}_{r_i}$'s projection for, say, horizon 5 to only be compared to $\hat{Y}_{r_j}$'s projections from horizon 3 to 7 allowing the algorithm to detect at max a 2-year lead or a 2-year lag relationship between the two series. To ensure comparability across different countries'projections, we scale the baseline projections data using z-score normalization for each series being compared.

As mentioned before, the distance calculation detailed above is repeated for each crop $c$, variable $v$, report year $t$ and the projection series of each country being compared have the length of horizon $h$. To compute the overall distance in projections of two countries, we average the DTW distances between projections for two countries $\hat{Y}_{r_i}$ and $\hat{Y}_{r_j}$ over all the report years. For instance, the distance between corn-yield projections for countries $r_i = US$ and $r_j = China$ considering the whole projection horizon $H = 10$ is computed as follows:

$$distance(\hat{Y}_{r_icv}, \hat{Y}_{r_jcv}) = \frac{1}{length(T)} \sum_{t=2002}^{2021} (DTW(\hat{Y}_{r_icvt}, \hat{Y}_{r_icvt})) \tag{5}$$

where $distance(\hat{Y}_{r_icvh}, \hat{Y}_{r_jcvh})$ gives average difference in projections of two countries for each crop $c$ and variable $v$ and $DTW(\hat{Y}_{r_icvt}, \hat{Y}_{r_jcvt})$ refers to the minimization problem described in equations 2, 3, and 4. We compute the standard errors for the computed distance using the bootstrapping method with 250 replications.

If a country's distance from a benchmark country is statistically indistinguishable from zero, we infer that its projections are correlated with the benchmark country, in other words, the two countries are similar in trends. Repeating this process for all crop-variable combinations and all benchmark countries provides enough information to evaluate if and where grouping behavior of countries' projections occurs.

## 3.2 Computing Bias in Projections

The error in the baseline projections is defined as the difference between the projected value and the actual value. We employ two measures for assessing projections accuracy which are common in the literature: the mean absolute percentage error — MAPE

$$MAPE_{rcvh} = \left( \frac{1}{T} \sum_t |100(\hat{Y}_{rcvth} - Y_{rcvth})/Y_{rcvth}| \right) \tag{6}$$

and the root mean squared percentage error — RMSPE

$$RMSPE_{rcvh} = \left( \frac{1}{T} \sum_t (100(\hat{Y}_{rcvth} - Y_{rcvth})/Y_{rcvth})^2 \right) \tag{7}$$

where $Y_{rcvth}$ is the actual value realized for the projection $\hat{Y}_{rcvth}$. $error_{rcvh}$ ($MAPE_{rcvh}$ or $RMSPE_{rcvh}$) is the average error calculated over the report years in the projections for country $r$, crop $c$, variable $v$, and horizon $h$. The average error can also be computed for each report year ($error_{rcvt}$) by averaging over $h$ instead of $t$.

## 3.3 Relationship between Bias and Herding

We estimate the relationship between similarity in projection trends and the bias in the baseline projections. This allows us to answer whether grouping the projections for various countries is the optimal strategy in the case of limited information. Moreover, we can evaluate the heterogeneity in the impact of herding on bias across crops, variables, horizons, and report years. Using different countries (United States, Brazil, or China) as benchmark countries for herding in the projections, we estimate the following equation:

$$\begin{aligned} log(error)_{rh} = \beta_0 + \beta_1 log(DistanceFromBase)_{rh} + \beta_2 CorrelatedWithBase_r \\ + \beta_3 DistanceIsRational_r + \beta_4 (CorrelatedWithBase_r \times DistanceIsRational_r) \\ + \epsilon_{rh} \end{aligned} \tag{8}$$

for each crop and variable separately, using the error calculated above. $log(DistanceFromBase_{rh})$ is the log of computed DTW distance of country $r$'s projections from the Base country for each of the projection horizons. $CorrelatedWithBase_r$ is an indicator variable taking the value 1 if, on average, country $r$'s projections for the given crop-variable are correlated with the Base country, and 0 otherwise. $DistanceIsRational_r$ is an indicator variable that takes the value 1 if the distance in the realized series lies within the confidence interval of the average distance in the projections of country $r$ from the Base country. The final term is the interaction of the two indicator variables, reflecting the effect of rational correlation in projections on the error in the projections for country $r$.

We estimated two versions of this equation for our analysis. When the United States is set as the benchmark country, we only include the variables $log(DistanceFromBase_{rh})$ and $DistanceIsRational_r$. That is because all the other countries' projections are correlated with the United States and hence both the remaining variables are redundant in the regression.

# 4 Results

Figures 1 through 6 graphically display the results from the estimations in sections 3.1 and 3.2. We use the dynamic time warping algorithm to measure the distance in projections of various countries compared to the projections of the United States for corn, soybeans, and wheat yields, areas harvested, and total consumption. We also depict the errors in the projections for each country decomposing it by projection horizon. Our results show strong correlation in the baseline projections trends with the United States. Top panels of figures 1 through 6 show that the projections for each of the 12 plotted countries do not significantly differ from the United States in their USDA baseline projections for corn and soybeans yield, area harvested, and total consumption (all confidence intervals contain 0 regardless of the value of the point estimate for distance in projections of each country). On the other hand, projection error trends in the bottom panels of figures 1 through 6 show that while the projection errors for the United States remain the lowest, herding does not ensure similar low errors for projections of other countries though their errors are statistically indistinguishable from those of the United States. When either China or Brazil are set as the base country, the projections are significantly different for several countries from those of the base country, showing reasonable variation in similarity across the countries by both crop and variable. However, what that means for the projection error requires further investigation.

Moreover, the distances in realized values of different countries from the United States (in the top panels of figures 1 to 6) are visually higher (and outside the 95% confidence interval) than the distances in the projections for corn and soybeans yields as well as total consumption. This suggests that yields' and total consumption's projections may be irrationally grouped with the United States' for a number of countries.

**Figure 1:** Corn Yield- Correlation Estimates and Error Calculations

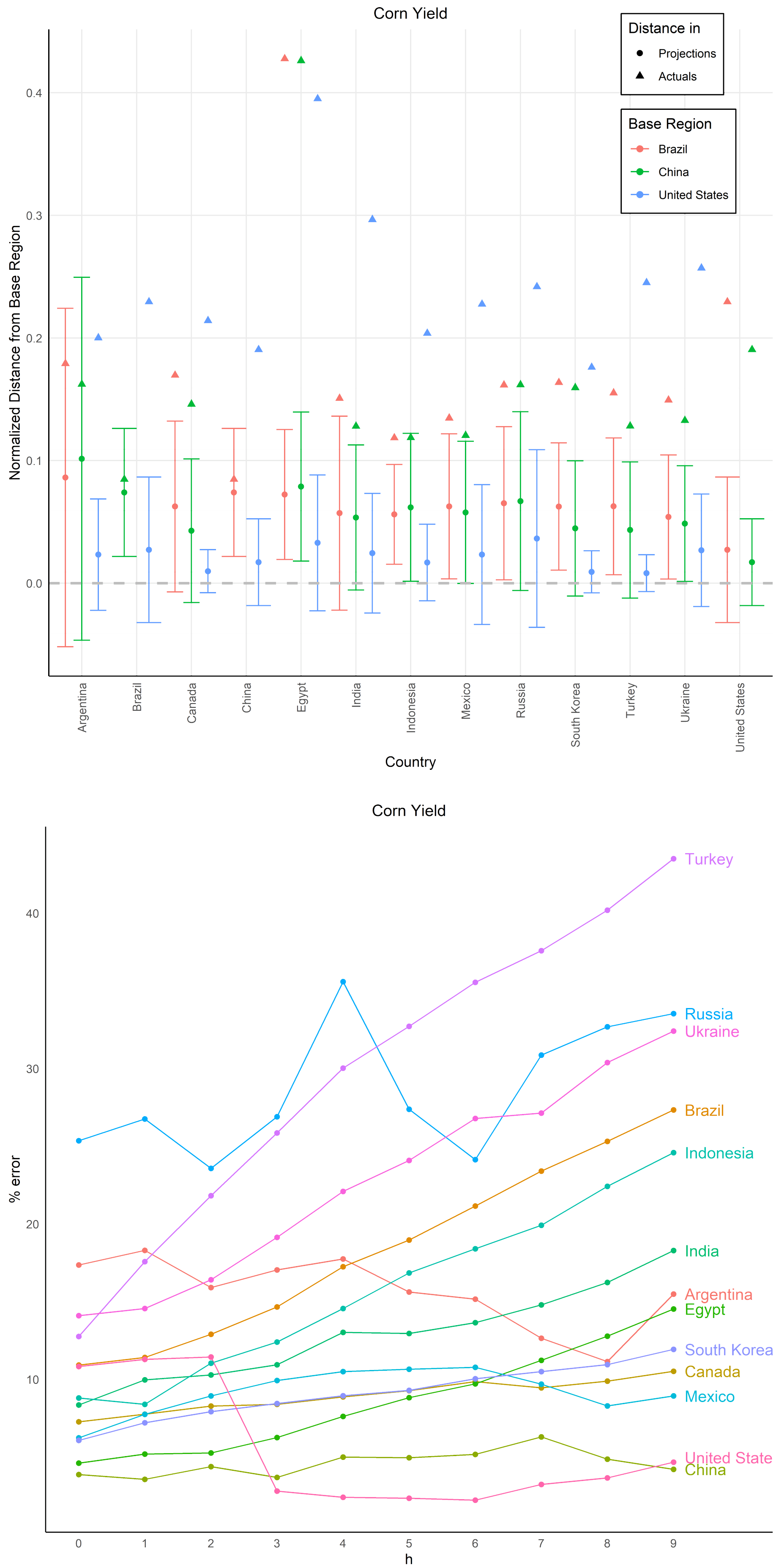
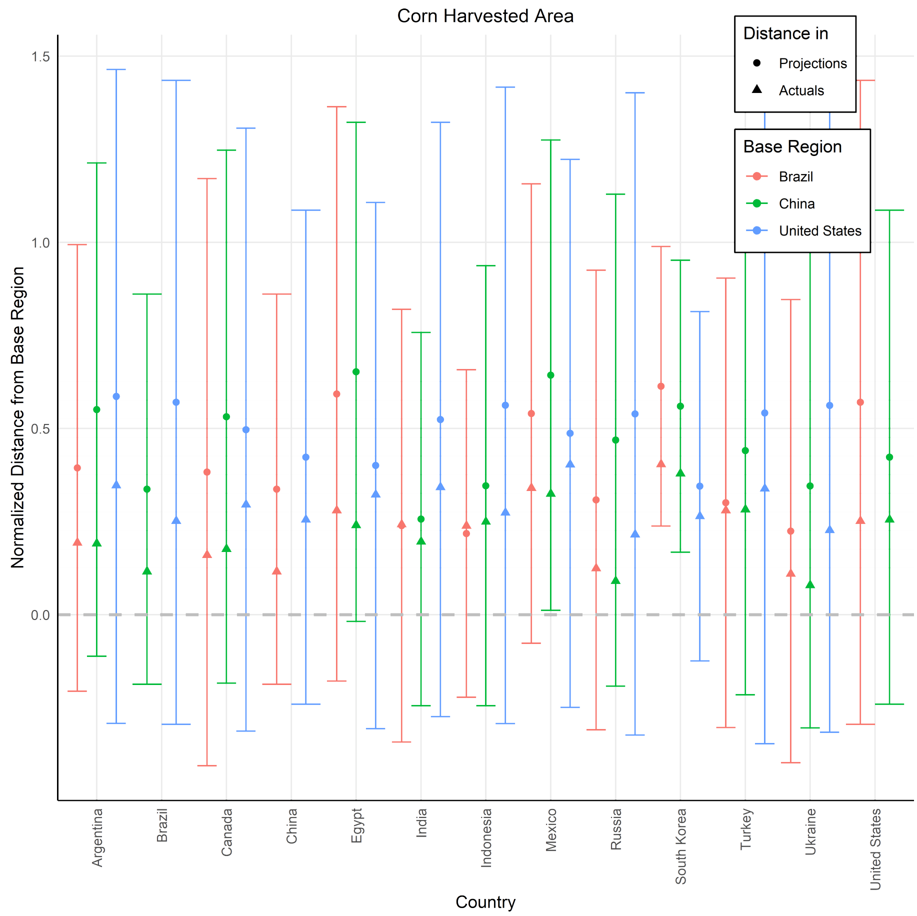


**Figure 2:** Corn Area Harvested - Correlation and Error Calculations



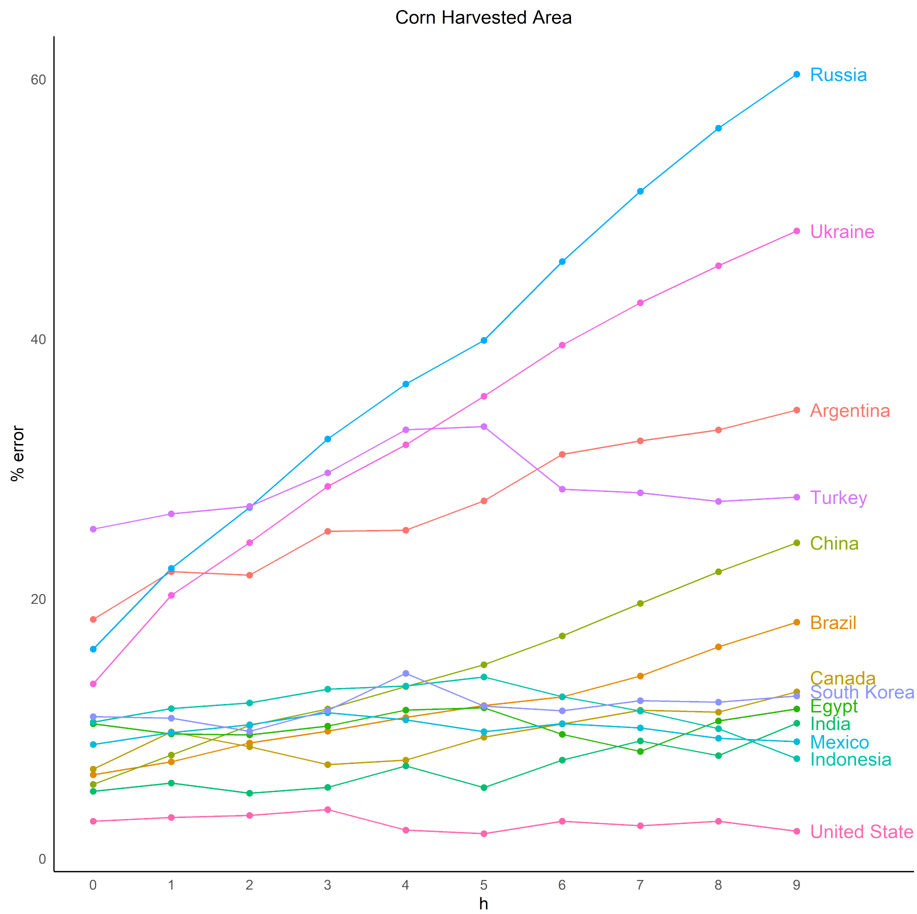
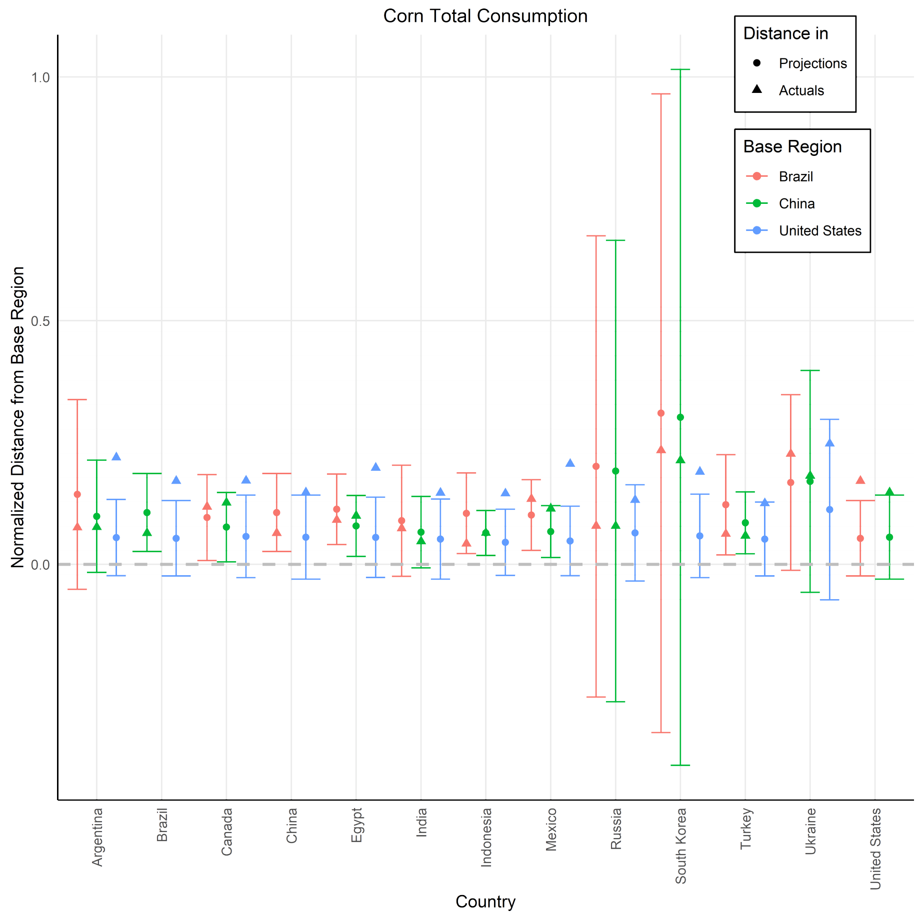



**Figure 3:** Corn Total Consumption - Correlation and Error Calculations



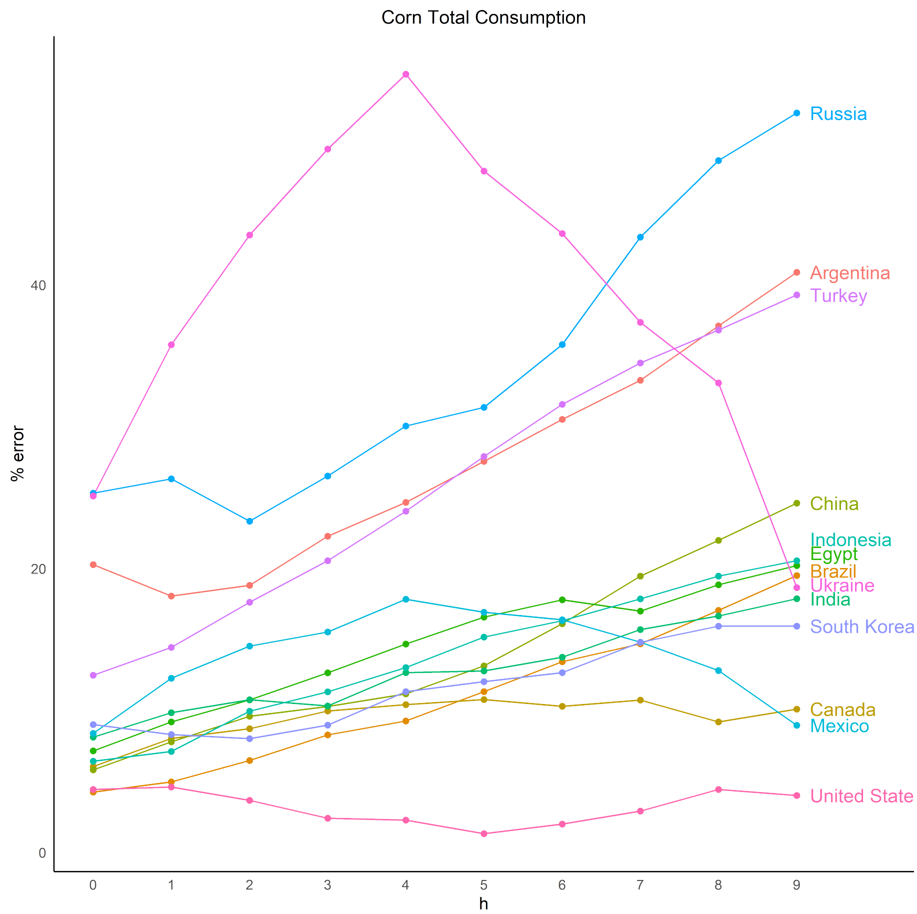
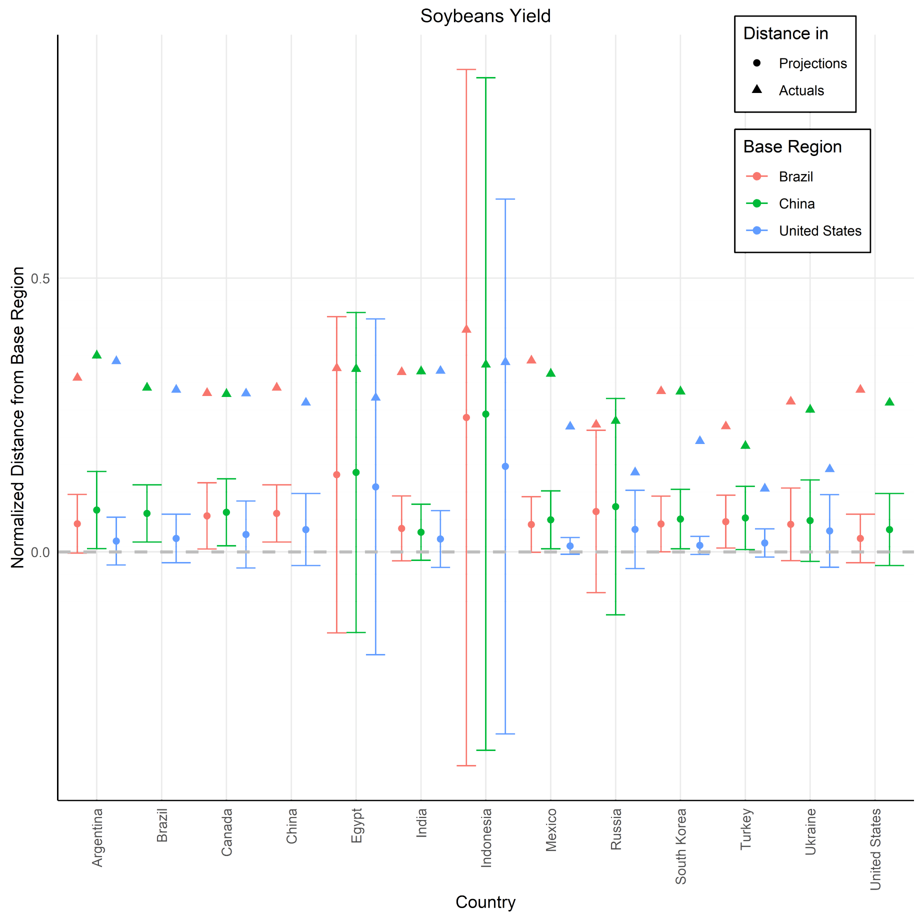



**Figure 4:** Soybeans Yield - Correlation and Error Calculations



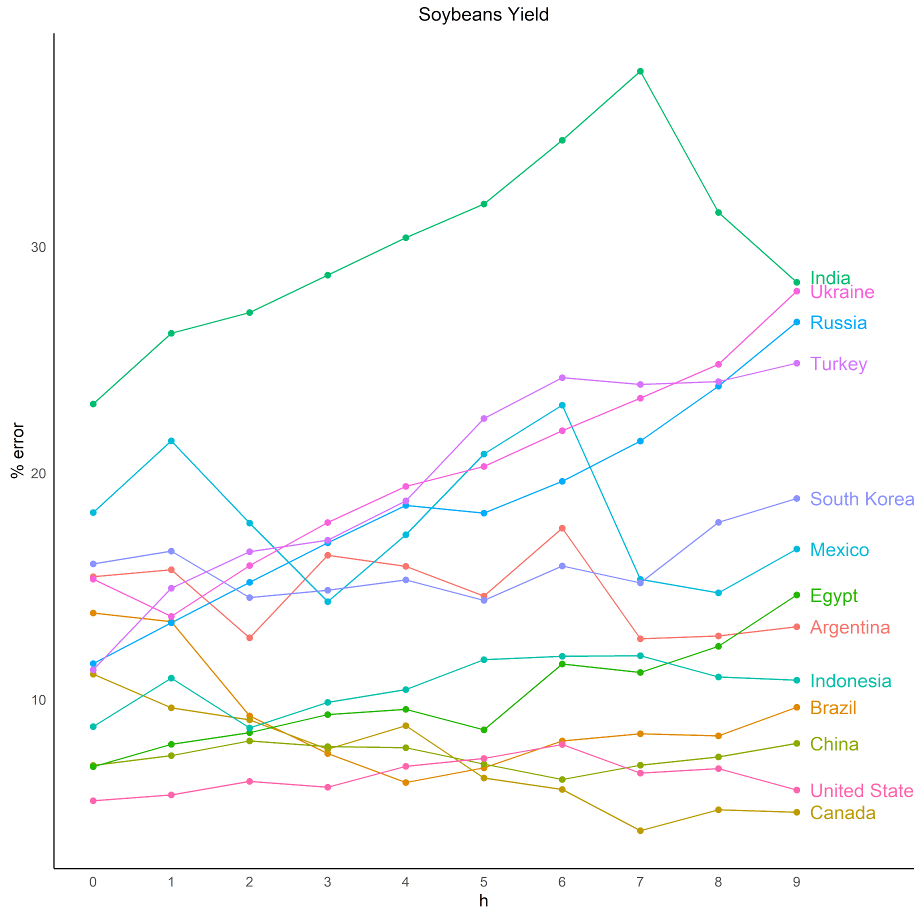
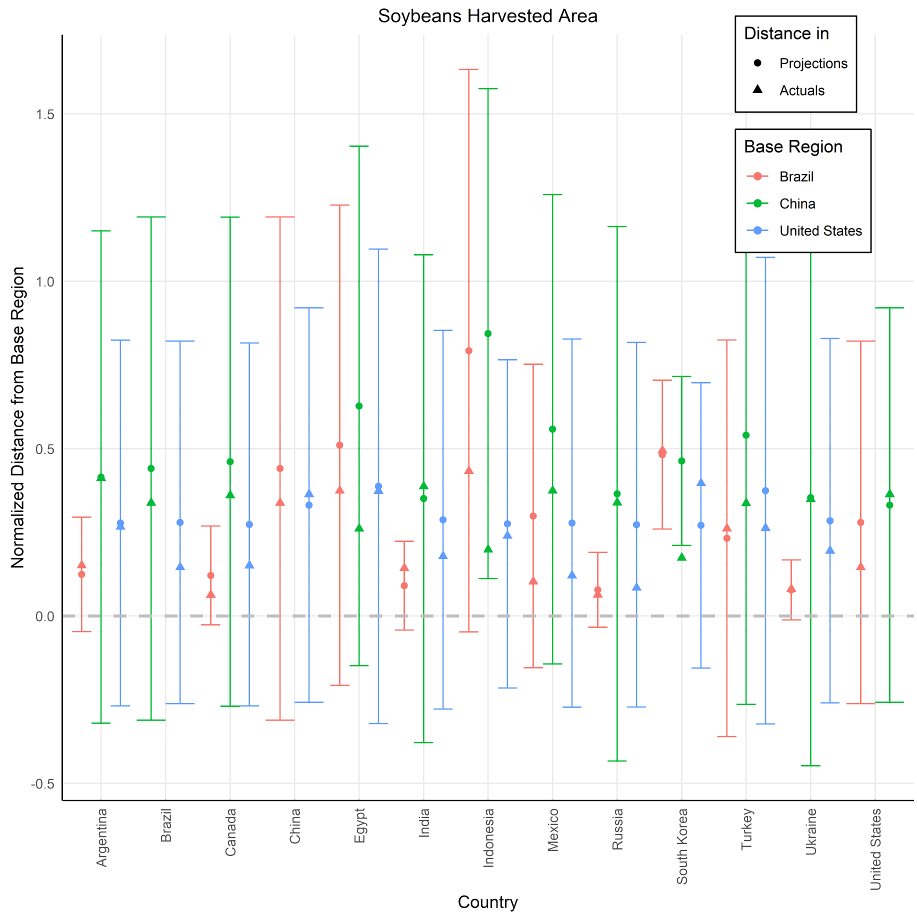



**Figure 5:** Soybeans Area Harvested - Correlation and Error Calculations



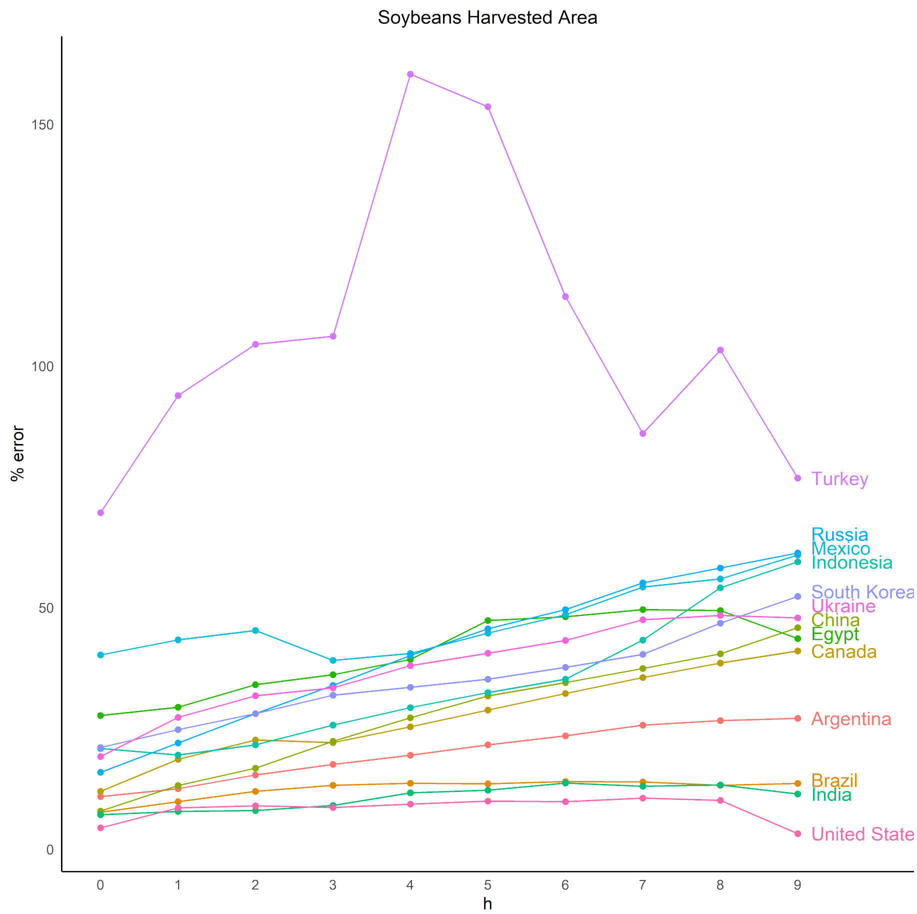
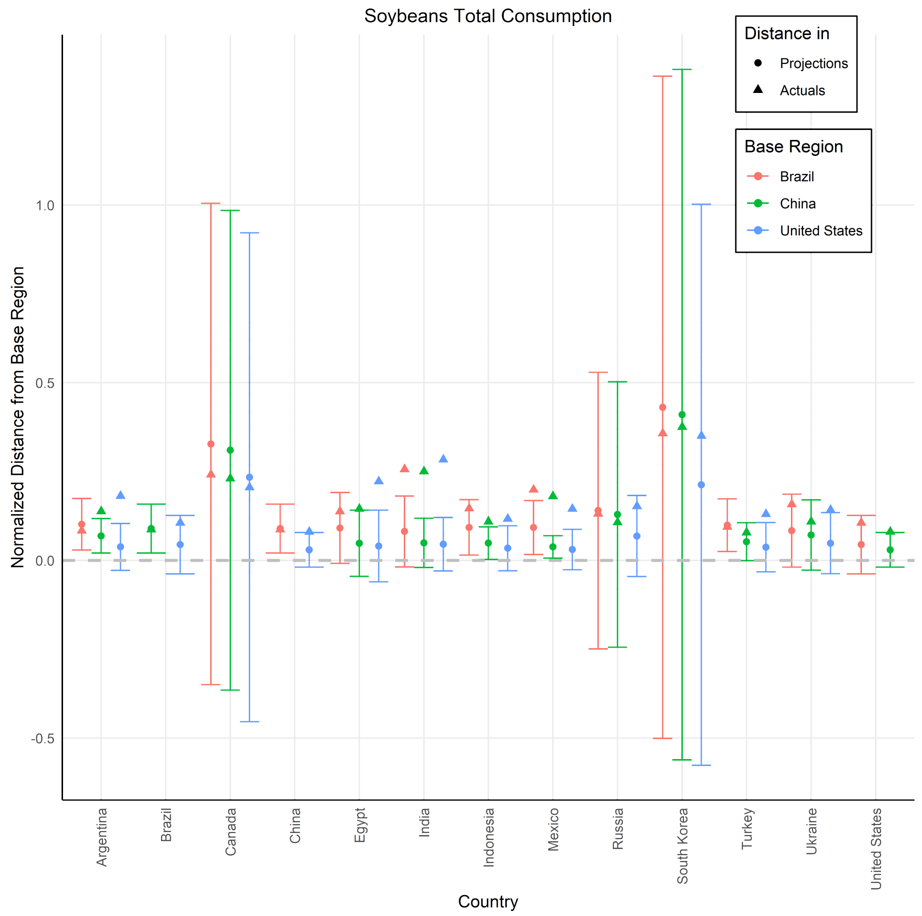
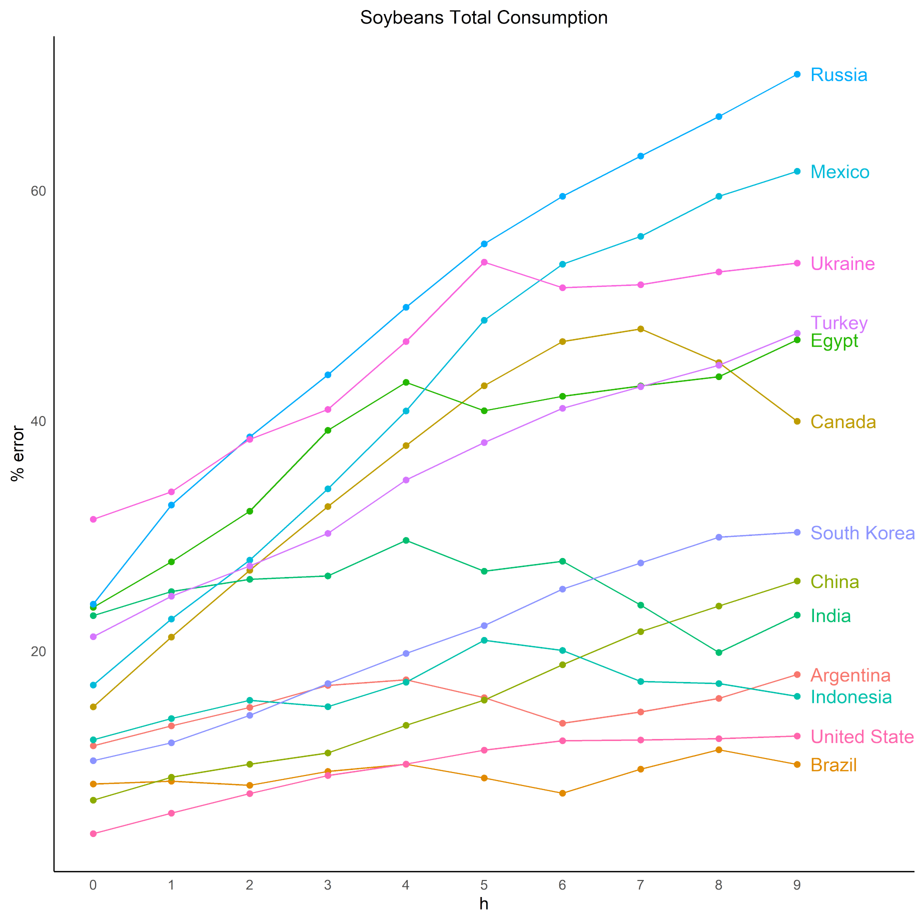



**Figure 6:** Soybeans Total Consumption - Correlation and Error Calculations



## 4.1 Heterogeneity in Bias

To better understand the relationship between accuracy of the projections and the correlation among various countries, we compute the DTW distances between all countries and the benchmark country for each projection horizon. This approach allows the projections to be correlated with the benchmark country differently across the projection horizon. Figures 7 through 14 show the distances among projections of top producing countries for corn, soybeans, and wheat, and the United States for each of the projection horizons. The solid lines show the average distance from the United States for each projection horizon while the dashed line shows the realized value's distance between the United States and a given country. Countries are distinguishable by colors.

**Figure 7:** Corn Yield - Distance in projections from the United States by projection horizon

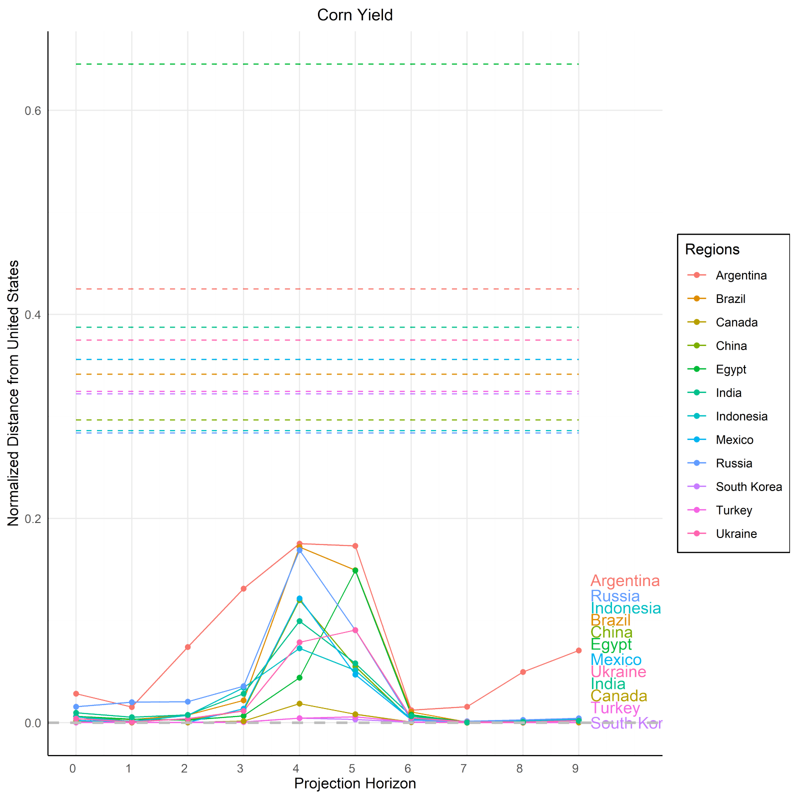


**Figure 8:** Corn Area Harvested - Distance in projections from the United States by projection horizon

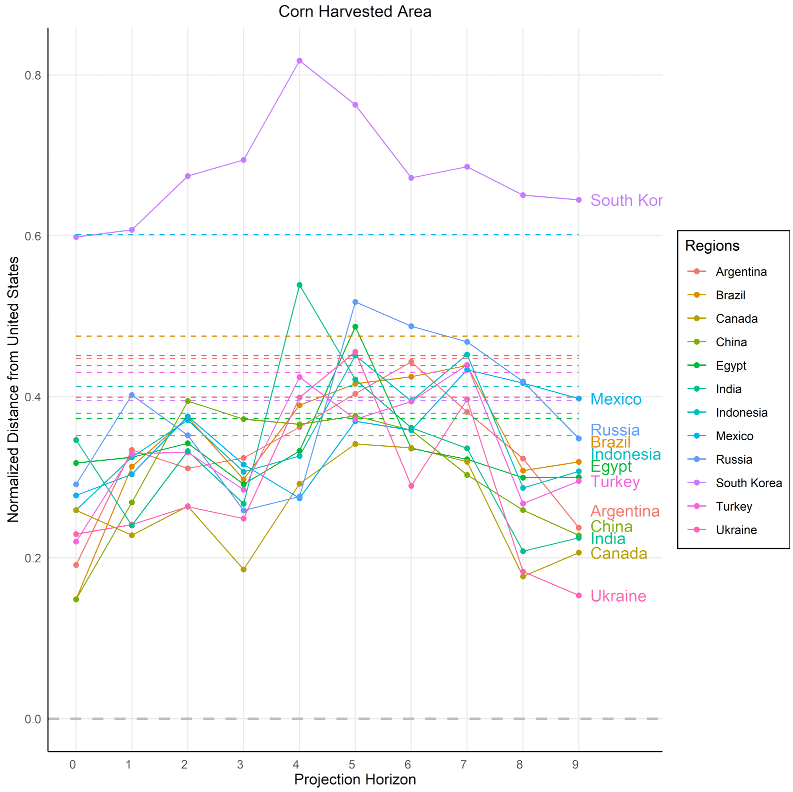



**Figure 9:** Corn Total Consumption - Distance in projections from the United States by projection horizon

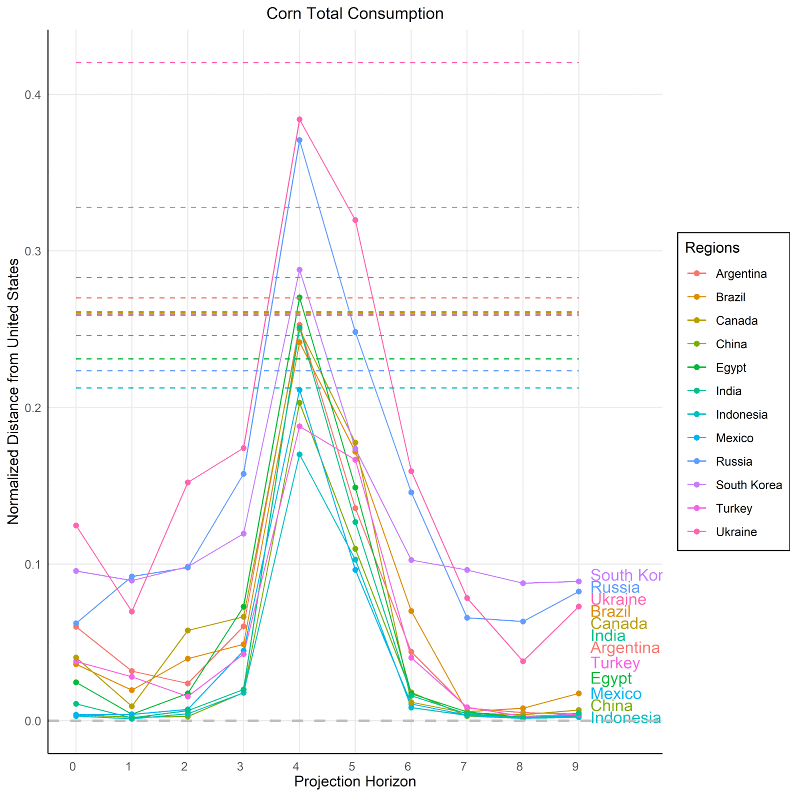


**Figure 10:** Soybeans Yield - Distance in projections from the United States by projection horizon

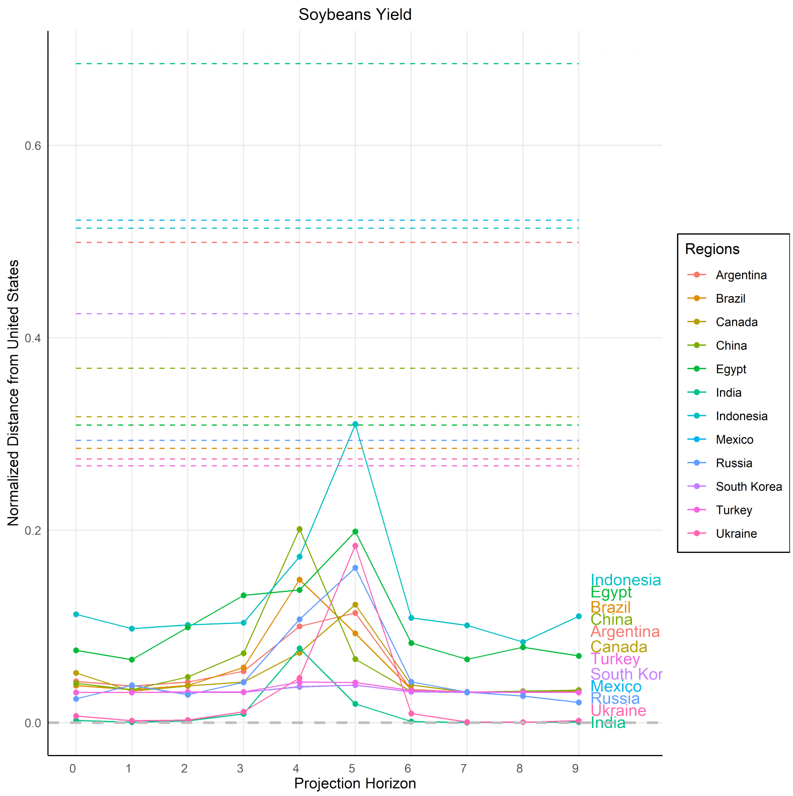



**Figure 11:** Soybeans Area Harvested - Distance in projections from the United States by projection horizon

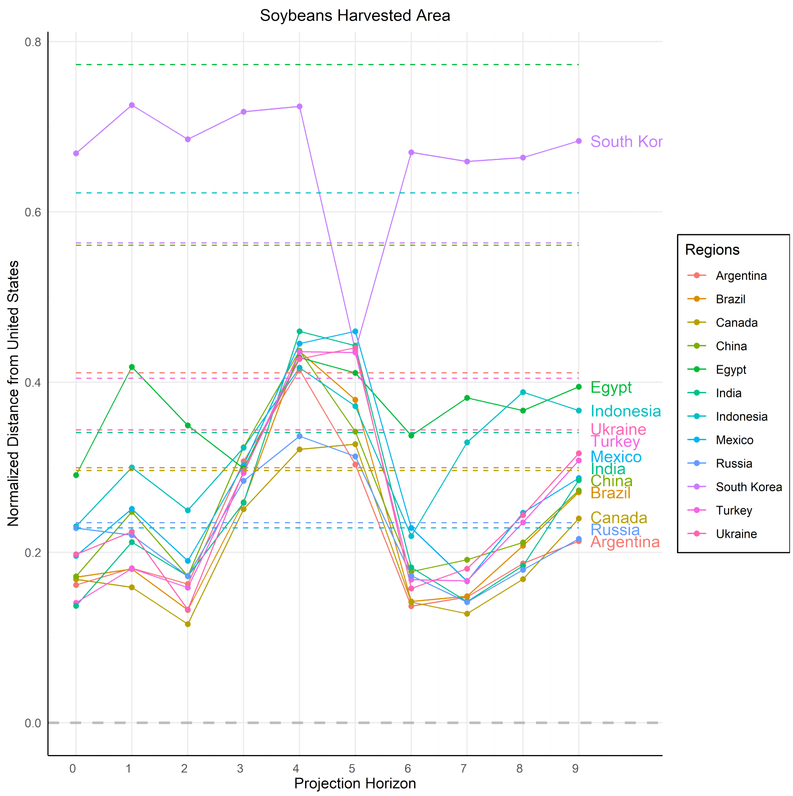


**Figure 12:** Soybeans Total Consumption - Distance in projections from the United States by projection horizon

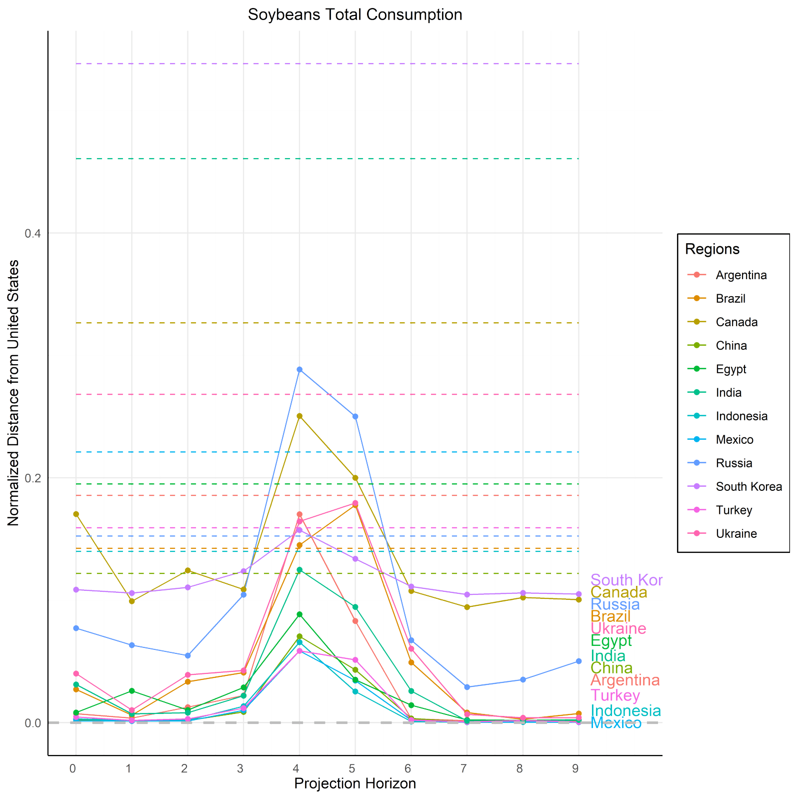



**Figure 13:** Wheat Yield - Distance in projections from the United States by projection horizon

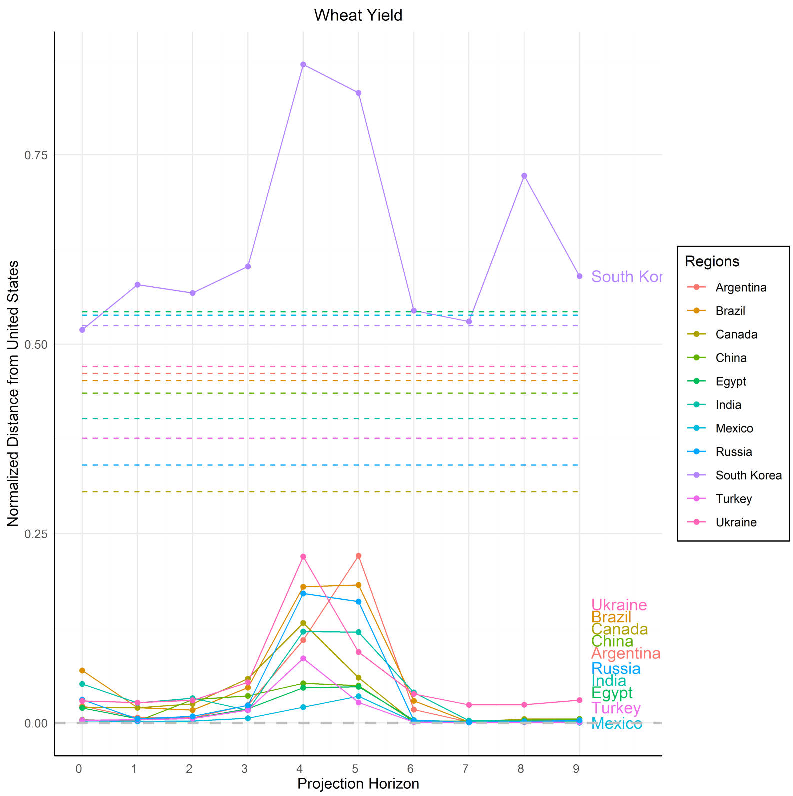


**Figure 14:** Wheat Area Harvested - Distance in projections from the United States by projection horizon

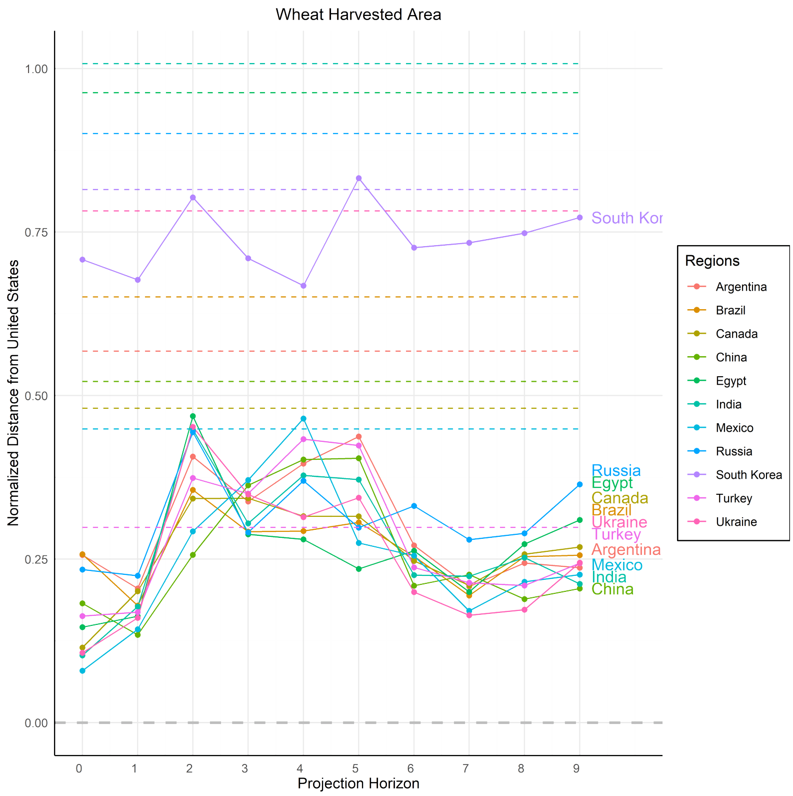
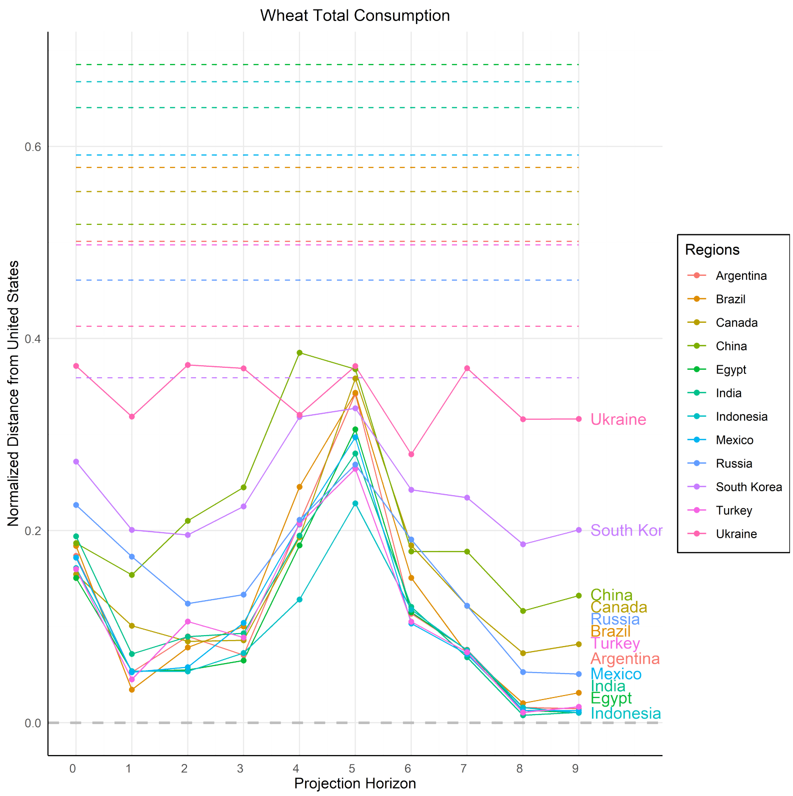



**Figure 15:** Wheat Total Consumption - Distance in projections from the United States by projection horizon



It is interesting to note that for most variables, the projections are much more correlated with the United States in the early projection horizons and the farther projection horizons. In the early horizons, the available historical data strongly guides the projection process, but, we observe that instead of the data for the specific countries, the first few horizons are made to follow the trends of the United States. This suggests that the overall correlation observed in the previous set of figures is driven more by the farthest points of the projection horizons. Therefore, the overall correlation with the United States, the correlation for each projection-horizon, and the rationality of overall correlation all play a role in the errors presented in figures 1 through 6. To investigate that relationship, we estimate equation 8.

## 4.2 Relationship between Herding and Bias

Results for estimation equation 8 for each of the three base countries (United States, China, and Brazil) are presented separately in tables 1-3.

**Table 1:** Distance and Accuracy - Base country is US

| Variable | Yield | Area Harvested | Imports | Exports | Total Consumption | Ending Stocks |
|---|---|---|---|---|---|---|
| **Corn** | | | | | | |
| Distance from Base$_{rh}$ (logged) | -0.0027 | 0.001 | -0.0862 | -0.0578 | -0.0127*** | -0.0994 |
| | (0.0027) | (0.0251) | (0.2208) | (0.0782) | (0.0038) | (0.0605) |
| Distance is Rational$_r$ | -0.1127*** | | 0.0589 | -0.2395 | -0.0207 | -0.6549 |
| | (0.0169) | | (0.1468) | (0.2402) | (0.0132) | (0.4139) |
| **Soybeans** | | | | | | |
| Distance from Base$_{rh}$ (logged) | -0.0039 | 0.0152 | 0.4055*** | -0.0712** | -0.0137** | 0.3549* |
| | (0.0032) | (0.0243) | (0.0622) | (0.0308) | (0.0059) | (0.1809) |
| Distance is Rational$_r$ | -0.0174 | -0.2559** | -0.0027 | -0.1608 | -0.0543 | 1.2547*** |
| | (0.0115) | (0.1186) | (0.1137) | (0.2210) | (0.0402) | (0.2640) |
| **Wheat** | | | | | | |
| Distance from Base$_{rh}$ (logged) | 0.0014 | 0.047** | -0.0138 | -0.2791*** | -0.0201*** | -0.0331 |
| | (0.0025) | (0.0183) | (0.0347) | (0.0761) | (0.0040) | (0.0573) |
| Distance is Rational$_r$ | -0.0321*** | 0.1229*** | -0.0244 | 0.0523 | 0.0174 | |
| | (0.0113) | (0.0267) | (0.0395) | (0.1513) | (0.0136) | |

This table shows the estimation results for equation $log(error)_{rh} = \beta_0 + \beta_1 log(Distance\_From\_Base)_{rh} + \beta_3 Distance\_is\_Rational_r + \epsilon_{rh}$ for each crop and variable. Since all countries are significantly correlated with the United States in their projections, the other terms are irrelevant for estimation with United States as the base country. Each column and panel shows the results for a separate regression for the crop-variable labeled in the table. Parentheses contain robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

Each column and crop panel in table 1 shows estimation results corresponding to equation 9 with United States as the base country. The coefficient on *Distance from Base*$_{rh}$

*(logged)* measures the percentage change in the error if the projections are 1% farther from the base country (i.e. United States). Row 1 shows that for all variables except area harvested, projections being more distant from the United States are associated with a decrease in the projection errors of those countries. That is, decreasing the correlation with the United States in corn projections for all countries, which on average are highly correlated with the United States, is associated with decreases in their error/bias.

Moreover, the coefficient on *Distance is Rational$_r$* shows the percent change in error if the true distance in the realized values are within the confidence interval of the distance in the projections. That is, the distance in the projections is "rational" and reflects the realized values' distance in the trends of the historical data. Since all countries' projections are correlated with the United States projections, this variable can be interpreted as the change in error by switching from an irrational correlation to a rational correlation. Then, switching from an irrational to rational correlation in the projections of corn yield is associated with a $(100 * (e^{\beta_3 - 1})\%)$ 10.7% decrease in the projection error. Except for corn imports, which mostly have no trend since they are often assumed to remain constant across the projection horizon (see figure 16 in the appendix), all other variables display a reduction in projection error when the correlation with the United States in projections is rational.

For soybeans, projection errors in yield, exports, and total consumption decrease the more the projections for each country differs from the United States. Soybeans imports, on the other hand, have little to no variation in their trends (see figure 19) so this coefficient is identified solely from the variation in levels of distances across countries. Moreover, all variables for soybeans have a lower error associated with lower correlation with United States in projections, except for ending stocks. *Distance is Rational$_r$* seems to be related with a higher error in soybeans ending stocks. That is, among the countries where the realized values are as correlated as the projections, higher error is observed (recall that *Distance is Rational$_r$* is an indicator variable taking the value 1 if the distance in projections matches the distance in realized data). This is because our data has low variation in corn ending stocks errors (errors are high for almost all regions) while there is less variation in the variable *Distance is Rational$_r$*. Therefore, only a few outlying countries, for which the distance is rational, are identifying this coefficient and they happen to have overall higher errors. When it comes to rationality, we observe that if the distance in projections is rational (i.e. distance in projections is similar to the distance in realized values) it is associated with significantly lower errors for area harvested while all the other variables have negative but insignificant coefficients.

For wheat, we observe that distance from base country in area harvested is significantly and positively associated with the projection error. That is, increasing the distance to the United States is associated with higher error which suggests that, for wheat area harvested, aligning the projection trends with the United States is a good strategy to reduce bias. For exports and total consumption, the situation is opposite: increasing the distance is significantly related to lower errors for both of these variables. The *Distance is Rational*$_r$ variable, on the other hand, has a significant negative coefficient for wheat yields but a positive estimate for wheat area harvested. This implies that if the distance in projections is close to the distance in realized values for wheat yields, it is related to lower error in the projections. However, even if the distance in projections is rational, it will cause higher error in wheat area harvested projections which is a puzzling result.

Overall, there are two main takeaways from these results. First, on average, projections for countries that are significantly different from the United States are associated with lower projection errors for these countries for most variables across all three crops, even if marginally so. This suggests that the projections of other countries that are following similar trends to the United States projections which proves to be bias inducing in most cases. The exceptions are wheat area harvested, soybeans imports and ending stocks where increasing the distance from the United States is significantly associated with increases in the projection error for these countries and variables. Second, in the case that the realized values for a country follow a similar trend to that of the United States, correlating the projections is, on average, likely to reduce the error in the projections. This is reflected in the variable "Distance is Rational$_r$".

We also use Brazil and China as the base regions because they are major producers and trade partners of the United States. Therefore, we also assess the correlation in projections by setting China and Brazil as benchmark countries instead of the United States. Tables 2 and 3 show the estimation results from equation 8 by setting China and Brazil as base countries, respectively.

Since the projections for all countries are statistically correlated with the United States in their trends, using other top producers as the benchmark country allows us to

**Table 2:** Distance and Accuracy - Base country is China

| Variable | Yield | Area Harvested | Imports | Exports | Total Consumption | Ending Stocks |
|---|---|---|---|---|---|---|
| **Corn** | | | | | | |
| Distance from Base$_{rh}$ (logged) | -0.0038 | -0.0033 | -0.0488 | 0.1319 | -0.0108*** | -0.0035 |
| | (0.0023) | (0.0113) | (0.0540) | (0.1779) | (0.0028) | (0.2214) |
| Correlated with Base$_r$ | 0.007 | 0.015 | 0.4829*** | -0.0664 | 0.0288 | 0.0775 |
| | (0.0109) | (0.0310) | (0.1064) | (0.1765) | (0.0257) | (0.0739) |
| Distance is Rational$_r$ | -0.0589 | 0.0183 | 0.0513 | | 0.0198 | 0.311*** |
| | (0.0398) | (0.0558) | (0.1405) | | (0.0210) | (0.0798) |
| Distance is Rational × Correlated with Base$_{rh}$ | | | | | -0.0503 | |
| | | | | | (0.0320) | |
| **Soybeans** | | | | | | |
| Distance from Base$_{rh}$ (logged) | -0.0038*** | 0.0085 | -0.0471 | -0.043 | -0.0061 | 0.2632* |
| | (0.0011) | (0.0242) | (0.0360) | (0.0507) | (0.0053) | (0.1346) |
| Correlated with Base$_r$ | 0.0183 | 0.0215 | 0.0149 | | -0.0234 | -1.2679*** |
| | (0.0142) | (0.0303) | (0.0543) | | (0.0334) | (0.1888) |
| Distance is Rational$_r$ | -0.0267*** | 0.0137 | 0.0083 | 0.618*** | -0.0311 | -0.0043 |
| | (0.0093) | (0.0297) | (0.1843) | (0.0688) | (0.0359) | (0.0563) |
| Distance is Rational × Correlated with Base$_{rh}$ | | | -0.2442** | | -0.0466 | 1.1856*** |
| | | | (0.1198) | | (0.0494) | (0.2606) |
| **Wheat** | | | | | | |
| Distance from Base$_{rh}$ (logged) | 0.0013 | 0.0205 | -0.0144 | -0.2105*** | -0.0129** | -0.0399 |
| | (0.0024) | (0.0216) | (0.0361) | (0.0753) | (0.0052) | (0.0254) |
| Correlated with Base$_r$ | 0.0112 | -0.0988* | -0.1726* | 0.1713 | 0.0072 | 0.2434*** |
| | (0.0128) | (0.0538) | (0.0972) | (0.1587) | (0.0081) | (0.0858) |
| Distance is Rational$_r$ | -0.0285* | 0.0188 | -0.0982*** | 0.053 | -0.0507*** | 0.3855*** |
| | (0.0162) | (0.0158) | (0.0362) | (0.1626) | (0.0138) | (0.0252) |
| Distance is Rational × Correlated with Base$_{rh}$ | -0.0011 | | | | | |
| | (0.0182) | | | | | |

This table shows the estimation results for equation $log(error)_{rh} = \beta_0 + \beta_1 log(Distance\_From\_Base)_{rh} + \beta_2 Correlated\_with\_Base_r + \beta_3 Distance\_is\_Rational_r + \beta_4(Correlated\_with\_Base_r \times Distance\_is\_Rational_r) + \epsilon_{rh}$ for each crop and variable. Each column and panel shows the results for a separate regression for the crop-variable labeled in the table. Parentheses contain robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

evaluate whether the other major producer countries offer an alternate error minimizing approach to grouping. While limited information may result in projections for other countries following major producers, it does not necessarily mean that the United States trends should be applied globally. It is also reasonable to expect that for some crops and variables, following the trends of China or Brazil in projections may reduce errors because of the massive global contribution of these countries while for others it may not. Our results depict that heterogeneity.

For all crops and variables in table 2, increasing the distance from China reduces the error in projections on average, with the exception of soybeans ending stocks which

**Table 3:** Distance and Accuracy - Base country is Brazil

| Variable | Yield | Area Harvested | Imports | Exports | Total Consumption | Ending Stocks |
|---|---|---|---|---|---|---|
| **Corn** | | | | | | |
| Distance from $\text{Base}_{rh}$ (logged) | -0.0056*** | -3e-04 | -0.3767 | -0.1494 | -0.0188*** | -0.028 |
| | (0.0018) | (0.0098) | (0.4008) | (0.1114) | (0.0034) | (0.0788) |
| Correlated with $\text{Base}_r$ | 0.0255 | 0.0107 | -0.1017 | -0.0305 | 0.0413 | 0.3006*** |
| | (0.0163) | (0.0416) | (0.1710) | (0.1918) | (0.0404) | (0.0617) |
| Distance is $\text{Rational}_r$ | -0.0379 | -0.0213 | 0.0201 | -0.477*** | 0.0177 | -0.2641 |
| | (0.0346) | (0.0345) | (0.2836) | (0.1749) | (0.0175) | (0.3844) |
| Distance is Rational × Correlated with $\text{Base}_{rh}$ | | | | | -0.0383 | |
| | | | | | (0.0464) | |
| **Soybeans** | | | | | | |
| Distance from $\text{Base}_{rh}$ (logged) | -0.0036* | 0.0105 | -0.3541*** | -0.0035 | -0.0155*** | -0.2813** |
| | (0.0019) | (0.0082) | (0.0841) | (0.0324) | (0.0046) | (0.1100) |
| Correlated with $\text{Base}_r$ | 0.0285** | 0.1124* | 0.0573 | | -0.0101 | 0.3261*** |
| | (0.0134) | (0.0642) | (0.0590) | | (0.0408) | (0.0462) |
| Distance is $\text{Rational}_r$ | -0.025** | -0.1314* | -0.0053 | -0.1479 | -0.041 | -0.5401* |
| | (0.0099) | (0.0721) | (0.2627) | (0.1499) | (0.0513) | (0.3055) |
| Distance is Rational × Correlated with $\text{Base}_{rh}$ | | | | | -0.0235 | |
| | | | | | (0.0752) | |
| **Wheat** | | | | | | |
| Distance from $\text{Base}_{rh}$ (logged) | -0.0021 | -0.0351** | 0.0015 | 0.1641 | -0.0093** | -0.0307 |
| | (0.0019) | (0.0157) | (0.0188) | (0.1386) | (0.0037) | (0.0245) |
| Correlated with $\text{Base}_r$ | 0.0044 | 0.0044 | -0.0092 | | 0.0071 | 0.0978** |
| | (0.0082) | (0.0345) | (0.0321) | | (0.0243) | (0.0377) |
| Distance is $\text{Rational}_r$ | -0.0392*** | 0.0063 | 0.0942 | 0.3496 | 0.011 | 0.2823*** |
| | (0.0113) | (0.0466) | (0.0598) | (0.3560) | (0.0224) | (0.0834) |
| Distance is Rational × Correlated with $\text{Base}_{rh}$ | | | -0.1341 | | -0.0116 | |
| | | | (0.0912) | | (0.0290) | |

This table shows the estimation results for equation $log(error)_{rh} = \beta_0 + \beta_1 log(Distance\_From\_Base)_{rh} + \beta_2 Correlated\_with\_Base_r + \beta_3 Distance\_is\_Rational_r + \beta_4(Correlated\_with\_Base_r \times Distance\_is\_Rational_r) + \epsilon_{rh}$ for each crop and variable. Each column and panel shows the results for a separate regression for the crop-variable labeled in the table. Parentheses contain robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

has a positive significant coefficient. However, a correlation in projection trends with China is associated with significantly lower errors in wheat area harvested and imports, along with soybeans ending stocks. This suggests that, on average, wheat area harvested and imports, and soybeans ending stocks projections trends are following China and this alignment reduces error. On the other hand, the most interesting result to note in table 3 is that none of the coefficients for the variable "Correlated with $\text{Base}_r$" are negative and significant for any crops or variables. Contrarily, corn ending stocks, soybeans yield, area harvested and ending stocks, along with wheat ending stocks all depict a significant positive relationship between other countries' correlation with Brazil and their higher errors. Moreover, even rational correlation between other countries and Brazil (coefficient of the variable "Distance is Rational$_r$") is not significantly related to reductions in error.

# 5 Conclusion

While the USDA International Baseline Projections are prepared through a combination of model-based values and expert analysts' judgements, the baseline projections for all other countries are found to have an underlying correlation with the United States. Our results show that this is bias reducing for a few crop specific variables, but contributes to higher error in other cases, compromising the overall accuracy of the projections. Given the importance of the baseline projections in domestic agricultural policy, it is imperative to identify where the projection correlation among countries that is beyond the correlation in the realized values is increasing bias and reducing accuracy, so that it can be addressed.

We employ various methods to identify the correlation in the projections of different countries, assess their degree of accuracy/bias, and map the relationship between projections' similarity in terms of correlation and projections' error. Our results show that only select variables that are grouped together in their projection trends are associated with reduced errors while most of the others are not. Soybeans imports and ending stocks, and wheat area harvested are the only three crop-variable combinations where similarity/correlation in projection trends with the United States is associated with more accurate projections for the other countries. Among other crop-variable combinations, our results show that correlation in projection trends is significantly decreasing the accuracy of the projections in corn total consumption, soybeans exports, soybeans total consumption, wheat exports, and wheat total consumption. These findings can be used by the team preparing the USDA baselines projections by checking that the projections' correlations does not exceed realized values' correlations as this may decrease their accuracy.

# References

Benchimol, J., M. El-Shagi, Y. Saadon, et al. 2020. *Do Expert Experience and Characteristics Affect Inflation Forecasts?*. Bank of Israel-Research Department.

Berndt, D.J., and J. Clifford. 1994. "Using dynamic time warping to find patterns in time series." In *KDD workshop*. Seattle, WA, USA:, vol. 10, pp. 359–370.

Blasco, N., P. Corredor, and S. Ferreruela. 2012. "Does herding affect volatility? Implications for the Spanish stock market." *Quantitative Finance* 12:311–327.

Bora, S.S., A.L. Katchova, and T.H. Kuethe. 2021a. "The Accuracy and Informativeness of Agricultural Baselines.", pp. .

—. 2021b. "The Rationality of USDA Forecasts under Multivariate Asymmetric Loss." *American Journal of Agricultural Economics* 103:1006–1033.

Devenow, A., and I. Welch. 1996. "Rational herding in financial economics." *European economic review* 40:603–615.

Herzenstein, M., U.M. Dholakia, and R.L. Andrews. 2011. "Strategic herding behavior in peer-to-peer loan auctions." *Journal of Interactive Marketing* 25:27–36.

Huang, R., M. Krishnan, J. Shon, and P. Zhou. 2017. "Who herds? Who doesn't? Estimates of analysts' herding propensity in forecasting earnings." *Contemporary Accounting Research* 34:374–399.

Isengildina-Massa, O., B. Karali, T.H. Kuethe, and A.L. Katchova. 2020. "Joint Evaluation of the System of USDA's Farm Income Forecasts." *Applied Economic Perspectives and Policy*, pp. .

Jeong, Y.S., M.K. Jeong, and O.A. Omitaomu. 2011. "Weighted dynamic time warping for time series classification." *Pattern recognition* 44:2231–2240.

Kuethe, T.H., T. Hubbs, and D.R. Sanders. 2018. "Evaluating the USDA's Net Farm Income Forecast." *Journal of Agricultural and Resource Economics* 43:457–468.

Langholtz, M., R. Graham, L. Eaton, R. Perlack, C. Hellwinkel, and G. Daniel. 2012. "Price projections of feedstocks for biofuels and biopower in the US." *Energy Policy* 41:484–493.

Müller, M. 2007. "Dynamic time warping." *Information retrieval for music and motion*, pp. 69–84.

Regmi, M., A.M. Featherstone, B.C. Briggeman, and D. Subedi. 2021. "Assessing and Improving USDA's Farm Income Baseline Projections.", pp. .

Skorbiansky, S.R., N. Childs, and J. Hansen. 2018. "Rice in Asia's feed markets." Working paper.

Trueman, B. 1994. "Analyst forecasts and herding behavior." *The review of financial studies* 7:97–124.

Van Campenhout, G., and J.F. Verhestraeten. 2010. "Herding behavior among financial analysts: A literature review.", pp. .

Varatharajan, R., G. Manogaran, M.K. Priyan, and R. Sundarasekar. 2018. "Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm." *Cluster Computing* 21:681–690.

Wang, G., and Y. Wang. 2018. "Herding, social network and volatility." *Economic Modelling* 68:74–81.

# Appendix

**Figure 16:** Corn Imports - Distance in projections from the United States by projection horizon

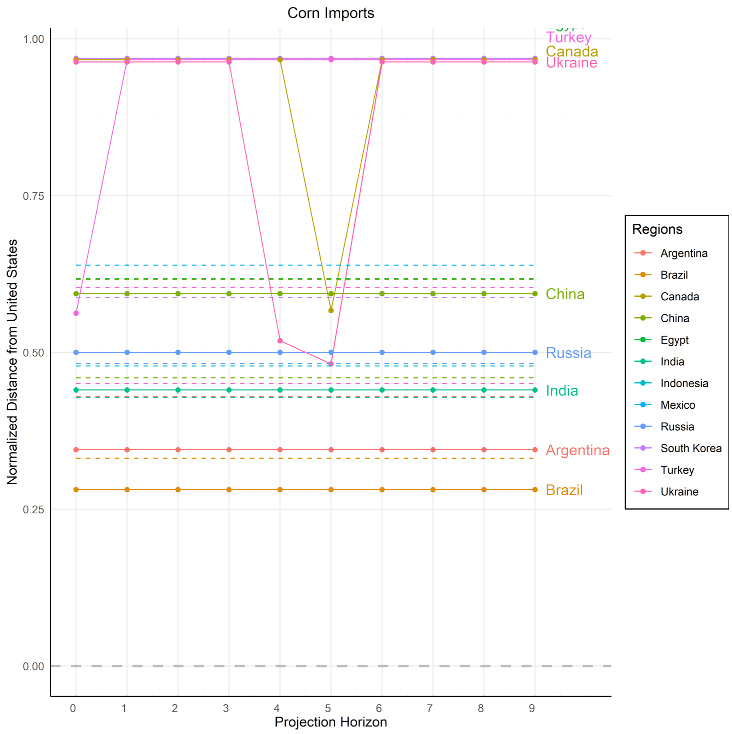



**Figure 17:** Corn Exports - Distance in projections from the United States by projection horizon

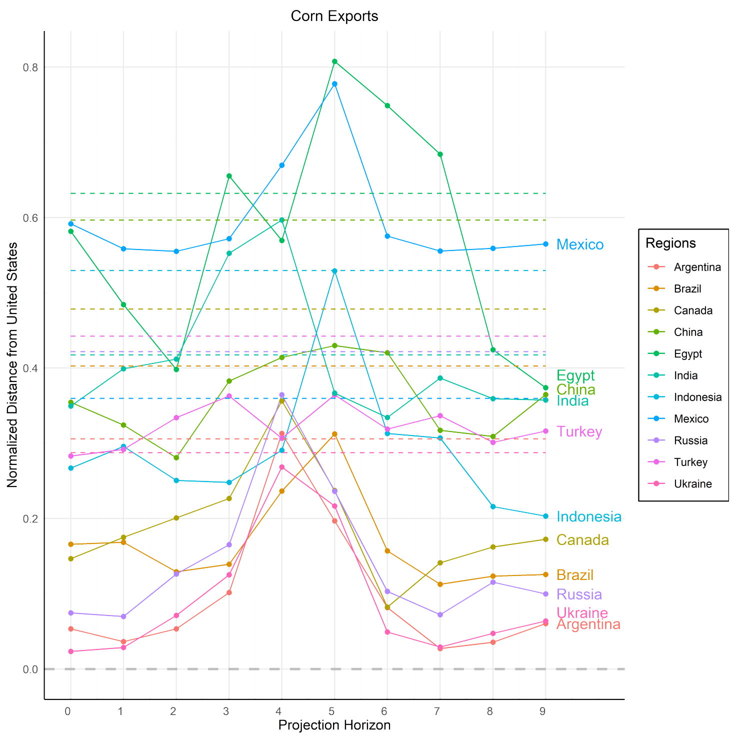


**Figure 18:** Corn Ending Stocks - Distance in projections from the United States by projection horizon

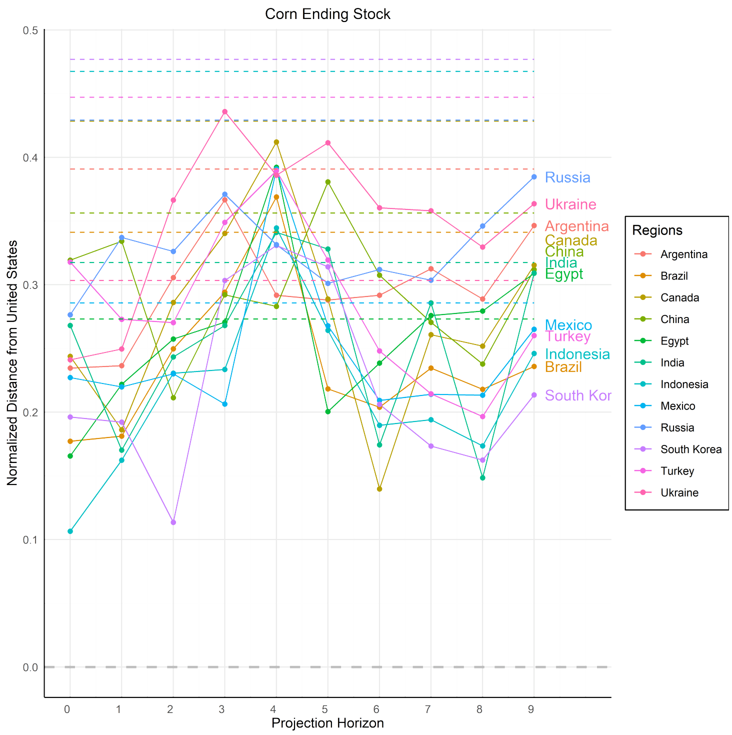



**Figure 19:** Soybeans Imports - Distance in projections from the United States by projection horizon

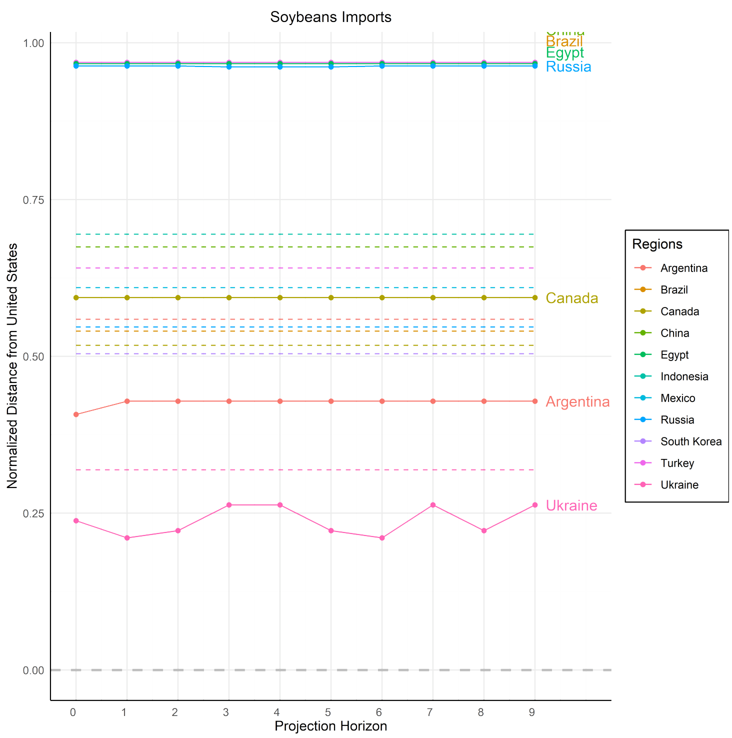


**Figure 20:** Soybeans Exports - Distance in projections from the United States by projection horizon

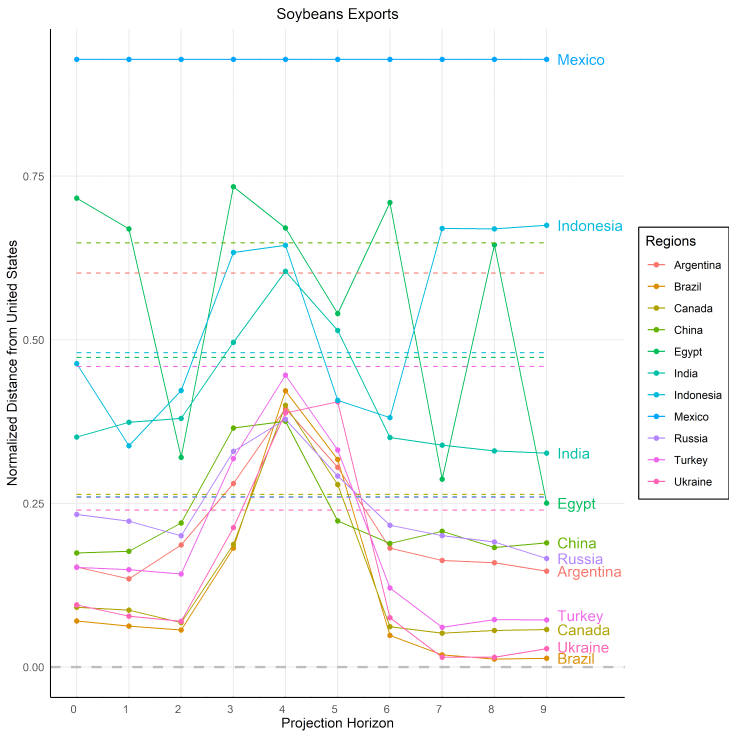



**Figure 21:** Soybeans Ending Stocks - Distance in projections from the United States by projection horizon

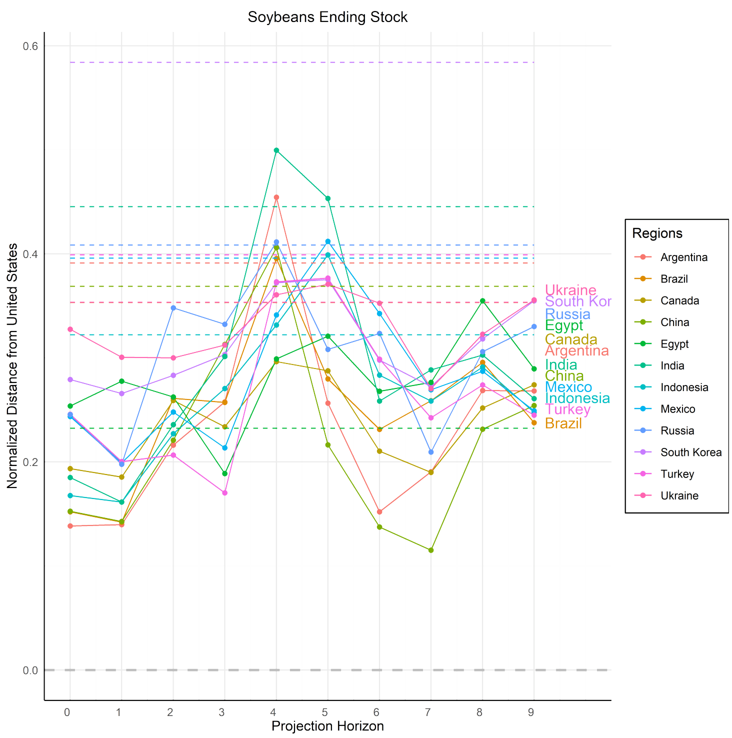


**Figure 22:** Wheat Imports - Distance in projections from the United States by projection horizon

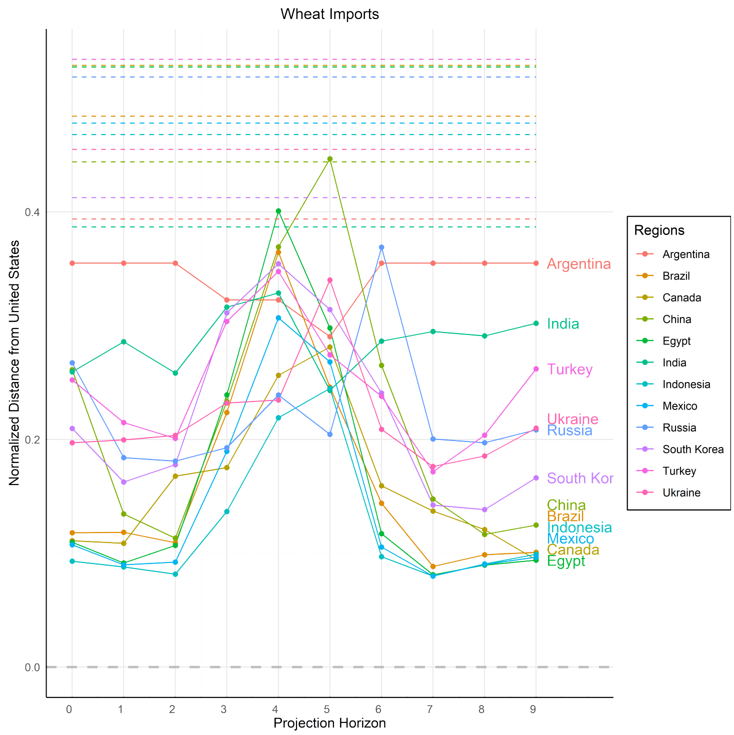



**Figure 23:** Wheat Exports - Distance in projections from the United States by projection horizon

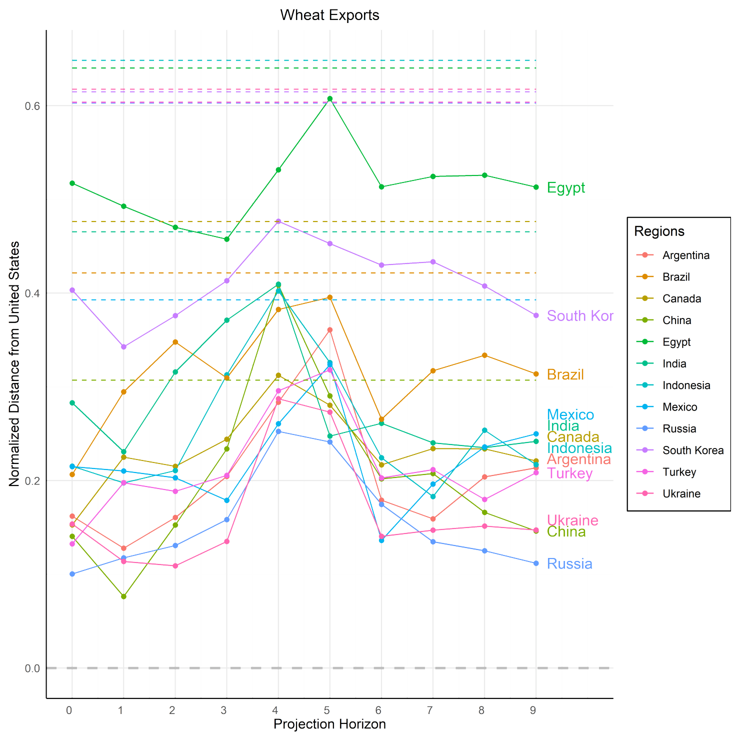


**Figure 24:** Wheat Ending Stocks - Distance in projections from the United States by projection horizon