Crop options:
modeling farmer intra-season decisions to abandon as an American put option

Asa Watten, Department of Agriculture, Food, and Resource Economics. Michigan State University. Email: wattenas@msu.edu

David Hennessy, Department of Agriculture, Food, and Resource Economics. Michigan State University. Email: hennes64@anr.msu.edu


Copyright 2018 by Asa Watten and David Hennessy. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Abstract

We provide structure to the problem of estimating crop yield distributions and incorporate the farmer’s economic decision to continue to husband a crop or abandon throughout a growing season. We employ the mathematical technology of a Pólya-Eggenberger Urn process, which gives a binomial tree structure to crop progress as well as other useful properties for modeling crops such as: (1) it produces a beta distribution in the limit for crop yield conditional on husbanding the crop to the end of the season; (2) it allows for path dependence such that good intra-seasonal crop growth makes further good outcomes more likely, which allows early-season outcomes to have large impacts on expected yield; and (3) it provides a structure such that expected values can be computed recursively and farmers can exercise the option to abandon. The model has a variety of uses including determining the impact of insurance on abandonment, scenario modeling under climate change, and valuing technological improvements. In a numerical illustration of winter wheat in Kansas, our example farmer would be 7.5 percentage points more likely to abandon if their insurance coverage level increased from 75% to 80%; a 20% increase in the variance of the yield distribution leads to a 2 percentage point increase in abandonment; and halving mid-season costs reduces abandonment by 6.5 percentage points.
1 Introduction

While it is well understood that crop production is stochastic process—weather, pests, and disease arrive randomly and affect crop outcomes—the endogenous component of this stochastic process, intra-seasonal farm management decisions, has been studied far less. As their crop matures, farmers update their expectations about end-of-season crop yields and may adjust their production plans during the growing season accordingly. An important economic choice that farmers may take is to abandon a crop altogether. Following a cold patch, period of dry weather, or lower than expected prices farmers may choose to neglect or plow under a stressed crop if expected yields are insufficient to cover harvest costs—or if abandoning the crop would lead to a more lucrative insurance payment. In this paper we make progress on understanding endogenous intra-season management decisions by modeling the option to abandon the crop and set out a plan for estimating the models parameters from farm-level data.

Crop abandonment is not a trivial issue. Many regions on the boundary of a crop’s climate zone experience weather stresses which may lead poor crop growth or total loss. Some examples are cotton in West Texas and Oklahoma, corn in the Dakotas, and maize, ground-nuts, and cotton in the semi-arid regions of Africa. To illustrate our model we focus on rain-fed winter wheat in Kansas. Figure 1 shows a time series of box-plots of the proportion of acres of winter wheat that were planted in a given county that were not harvested and for each Ag District (a collection of counties) using USDA-NASS Quick Stats data for the years 1970 to 2007. It is apparent from the figure that high levels of abandonment are not uncommon with
8% of observations showing greater than 50% abandonment and 46% of the data above 10% abandonment.

The presence of abandonment in the data can lead to poor predictions when researchers attempt to estimate the distribution of yield outcomes for a crop. If the researcher chooses to only include yield data on acres that were harvested, they ignore what may have been lower yields had the crop been husbanded through to harvest. This will tend to skew estimated distributions to the right. If instead, the researcher includes observation with zero yield, they will find that commonly used continuous distributions do a poor job at capturing probability mass at zero. One might be tempted instead to treat yields as exogenous and estimate the distribution using non-parametric methods or allow for a discontinuous mass at zero. This
strategy, however, ignores the economic decision to abandon a crop, which interacts with the opportunity cost alternative uses of land and time and the price the farmer expects to command at harvest time. We propose that these economic factors shift the threshold at which a farmer is indifferent between continuing to care for or to abandon their crop. Further, as these factors are likely to change from season to season, an estimated distribution from time-series yield data only will not account for shifts at the point where a farmer will abandon from season to season.

To address this methodological shortcoming, we develop a model of a farmer’s yield expectations over the course of a growing season, allowing the farmer to change their production plan in response to it. The mathematical technology that we deploy is the Pólya-Eggenberger urn model, which we augment to allow for the option to abandon the crop mid-season. Pólya-Eggenberger urn models involve an algorithm in which a theoretical urn containing colored balls, say green representing good weather events and brown for bad, is sampled at any given time period. If, for example, brown is drawn, then some number of additional brown balls are added to the urn, increasing the probability of drawing a brown ball at future draws. Pólya-Eggenberger urn have a number of desirable properties for modeling yield expectations as described by Hennessy (2011): (1) they have a finite distribution—as do yields; (2) in the limit, as the number of periods becomes large, the distribution of outcomes is in the form of a Beta distribution, which is quite flexible; and (3) the process allows the researcher to explicitly model path dependence where, for example, good crop growth early on in a season increase the probability of seeing good crop growth later on in the season.

This work has the potential to be used in many applications. In the numerical ex-
ample in Section 4, we explore the effects of insurance, mid-season costs and higher yield variance on the probability that an example farm will abandon the crop. However, with farm-level data, one may use our model estimate its parameters and test a variety of counterfactual scenarios on yield and abandonment patterns such as: (1) mid-season production forecasting; (2) valuing technological advancements such as hardier seeds; (3) modeling the effect of changing weather patterns.

2 Related Literature

Researchers have made use of the beta distribution to model crop yield as it has the favorable attribute for modeling yields, which have a lower bound of zero and some finite maximum, of having a support from zero to one, which can be transformed to span any finite support and produces a fairly flexible shape with only two parameters that need to be estimated (three if the maximum is also simultaneously estimated, and four if both the maximum and minimum yields are as well). The flexibility of beta distribution have made it popular in the literature for modeling yield distributions. Additional uses of the beta distribution for modeling yield distributions are Borges and Thurman (1994), Babcock and Hennessy (1996), Goodwin (2009), and Classen and Just (2011) for examples. An early example of the use of the beta distribution for crop yields is Nelson and Preckel (1989). Their paper investigates the effect of input choices on yield distributions using the model of a beta distribution, what the authors refer to as a conditional beta distribution, that has parameters that are determined by a function of a vector of inputs. Nelson and Preckel’s interest is distinct from
our own as they hope to probe the effect of a input and production plan on yield. They do this by first estimating the parameters of the beta distribution on each farm and then estimating the relationship of the parameters as a function of a vector of inputs. In contrast, this work (as well as Hennessy 2011) is concerned with a farmer’s updating of information as they observe the crop’s growth and then changes to the farmers production plan, which we assume to optimally chosen. Hennessy (2011), from which this paper draws from significantly, makes use of a Pólya urn process, which gives a beta distribution in the limit as the time steps approach infinity, to model the information updating of the farmer through the growing season. This paper extends Hennessy (2011) by explicitly including the option to abandon in a similar fashion as an American put option.

The beta distribution is not the only distribution to be used by researchers to for crop yields. Other researchers have used a log-normal specification to model yield distributions (Goodwin, Roberts, and Coble, 2000), while others work (Goodwin and Ker, 2002; Ker and Coble, 2003) argue for non-parametric methods. Chen and Miranda (2004) compare the fit of estimates of cotton yield from county-aggregated data for counties in Texas at high risk of crop failure from popular distributions. They note that all of these continuous distributions are disadvantaged, including the beta distribution, if there is a probability mass on crop failure or no yield.

There is a smaller literature on factors that contribute to crop abandonment, which has primarily focused on regression analysis of adverse intra-seasonal events on yield and abandonment outcomes. For example Mulungu and Tembo 2015 finds evidence that lack of moisture early in the season was associated with crop abandonment
in Zambia. Our approach in contrast provides significantly more structure, which allows for a greater flexibility in modeling scenarios of interest.

3 Conceptual model

Our model begins with a risk-neutral\(^1\) farmer who has rational expectations of the distribution of harvest time yields. After sewing the crop, the farmer observes the progress of the crop and soil conditions and updates their expectations of the final yield distribution. The model includes two types of intra-season costs: 1) the cost of continuing to husband the crop eg application of herbicide and; 2) the opportunity cost of continuing to harvest, which we will think of as the cost of forgoing an insurance payout. We further assume that the optimal production plan is at the outset of the season and that the only choice which the farmer is making is actively making decisions on is the whether or not to abandon the crop or continue the production plan. At this time we fix crop prices to focus on the dynamics of crop progress. Future work will include dynamic prices using, which incorporates a second binomial tree.

For the structure of the model we formulate an augmented Pólya-Eggenberger urn process, which we will refer to simply as the “urn process” for brevity. The urn process is augmented by including the option of the farmer to abandon, and therefore forgo further evolution of the process. We will see that this augmentation changes the distribution of yields at harvest time relative to an urn process without abandonment.

\(^1\)We hope to extend the model to include risk aversion in subsequent work.
However, before we consider the augmentation, we first consider the foundations of the standard Pólya-Eggenberger urn.

As noted in Hennessy 2011, the Pólya-Eggenberger urn is suitable for modeling crop production and preferable to the more frequently binomial tree process, geometric brownian motion, for the following reasons: (1) unlike geometric brownian motion which gives rise to a log normally distributed stochastic outcome with a support from zero to positive infinity, Pólya-Eggenberger urn produces a distribution of outcomes with a bounded support, which is appropriate for considering yield outcomes; (2) the model allows for path dependence in that good growing outcomes early in the season (drawing a green ball) increase the probability of future good growing outcomes (increase the proportion of green balls in the urn); and, as mentioned above, (3) the limiting distribution of the process, the beta distribution, is quite flexible allowing for a bell shape, a U shape, as well as skewness.

3.1 Pólya-Eggenberger urn process

A Pólya-Eggenberger urn process is a model in which a theoretical urn contains colored balls—say green and brown. In the model, at every time period $t$, a single ball is drawn, observed, and then replaced with $r$ number of balls of the same color. Consider a Pólya-Eggenberger urn process that stops at harvest time $T$. Drawing a green ball will represent for us an incremental improvement in the growth of the crop given the crops current development—a step up in a binomial tree discussed below. A brown ball then represents an incremental lack of improvement or decline in the crop—a step down in the binomial tree. The yield outcome for the farm in a given
season is then a linear transformation of the proportion of green balls in the urn at harvest time $T$.\footnote{Again, conditional on no abandonment.}

Assume that the initial parameterization of green and brown balls, $g^*$ and $b^*$, is a function of the optimal ex-ante production plan, which we assume is fixed in this paper other than the option to abandon. It is convenient for our notation to begin the initial time period at $t = 1$ so that $t \in \{1, T\}$. For our model we will fix the number of balls to be added to the urn of the same color, $r$, to 1.\footnote{In the Pólya literature this parameterization of the model would be written as a model with a replacement matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.} Fixing $r = 1$ we lose no flexibility of the distribution of outcomes. This is because with the standard Pólya-Eggenberger urn process (without an option to abandon), the distribution of outcomes (proportion of green balls in the urn) in the limit as $t \to \infty$ is given by the beta distribution $\beta(g^*, b^*)$ (Mahmoud 2009 p. 51). Therefore, with $r = 1$, any Beta distribution can be achieved by choice of $g^*$ and $b^*$.

This process can be illustrated as a binomial tree with nodes $n_{j,t}$ (see Figure 1). We denote the index of any node with the ordered pair $(j, t)$, where $j \in \{1, t\}$ is ordered from the largest proportion of green balls to the smallest for each $t$. Therefore, given a current location at node $n_{j,t}$, the probability an upward (downward) step to $n_{j,t+1}$ ($n_{j-1,t+1}$) is equal to (one less) the proportion of green balls at $n_{j,t}$

$$\pi_{j,t} \equiv \frac{g^* + t - j}{g^* + b^* + t - 1}$$  \hspace{1cm} (1)
3.2 Augmenting the Pólya-Eggenberger urn process to accommodate abandonment

To incorporate abandonment, we execute a three step algorithm: 1) compute the transition probabilities of arriving at all nodes from each node’s antecedent nodes conditional on husbanding the crop to harvest and given the parameters $g^*$ and $b^*$. This is done in the section above. 2) Recursively compute the expected value at each node given the opportunity cost of insurance and future husbanding costs beginning with the terminal nodes $n_{i,T}$. 3) Recompute the probability of arriving at each node, beginning with the first node $n_{1,1}$, and exiting the binomial tree (abandoning the crop) whenever the expected value at a node equals or falls below the insurance payment (which is 0 if no insurance).

For step two of the algorithm, define the value at $T$, the end of the growing period, as
\[ v_{j,T} \equiv \max\{\mu_{j,T} - k_T, I\} \] where \( I \in [0, 1] \) is the insurance payment and \( k_T \) is the cost of cultivation (harvest) at time \( T \). The nominal value is simply a linear transformation of \( v_{j,T} \). The value at node \( n_{j,t} \) is then defined recursively as

\[
v_{j,t} = \max \left( I, \mu_{j,t} v_{j,t+1} + \left(1 - \mu_{j,t}\right) v_{j+1,t+1} - k_t \right)
\] (2)

starting with harvest time nodes, \( n_{T} \). Note that all previous periods costs are sunk—any \( k_t \) for all \( t' < t \) are sunk at \( t \). Insurance plays a role in the following way, if arriving at node \( n_{j,t} \) and \( v_{j,t} \leq I \), then the farmer abandons their crop, which is irreversible, and their final payout is \( I \). In other words, the farmer exits the binomial tree.

For the final step of the algorithm, computing the distribution of harvest time yields or values (including the probability of abandonment), we begin at the first node and recursively compute probabilities of arriving at each subsequent node accounting for dead branches of the tree. If the expected value \( v_{j,t} \) of continuing to husband the crop at any node \( n_{j,t} \) is greater than the insurance payout, the probability of arriving node \( n_{j,t} \) is the sum of the probabilities of having previously arrived at the antecedent nodes (which are 2 at most) and then moving up or down the tree to node \( n_{j,t} \). If the expected value \( v_{j,t} \) of continuing to husband the crop at the node \( n_{j,t} \) is less than or equal to the insurance payment \( I \), then the farmer abandons the crop and the probability of arriving at the node is 0.\(^4\)

\(^4\)This makes an assumption that there is a small cost \( \varepsilon \) of continuing to husband the crop at every time period.
The probability of arriving at any node may also be defined non-recursively, but painfully. Define the function \( h(\cdot) \), which takes two sets of ordered brown pairs \((j, t)\) and \((j', t')\) such that \( t' > t \) and \( j' \geq j \),

\[
h((j, t), (j', t')) \equiv \Pr(n_{j', t'}|n_{j, t})
\]

where \( \Pr(n_{j', t'}|n_{j, t}) \) is the probability of arriving at node \( n_{j', t'} \) given a current position at node \( n_{j, t} \) in the standard Pólya and Eggenberger urn model without the option to abandon.

The number of green balls and brown balls at time node \( n_{j, t} \) is \( g_{j, t} = g^* + t - j \) and \( b_{j, t} = b^* + j - 1 \). Then, the following defines a change in \( t \) and \( g \): \( \Delta t \equiv t' - t \), \( \Delta g \equiv \Delta t - (j' - j) \). Adapted from Mahmoud (2009 p. 51) in our notation, we can write:

\[
\Pr(n_{j', t'}|n_{j, t}) = \frac{\binom{\Delta t}{\Delta g} B(g_{j, t} + \Delta g, b_{j, t} + \Delta t - \Delta g)}{B(g_{j, t}, b_{j, t})}
\]

Where \( B(a, b) \) is the beta function. Let

\[
\hat{t}_j = \begin{cases} 
0 & \text{if } v_{j, t} \neq I \forall t \\
\text{t' s.t. v}_{j, t} = I & \text{otherwise.}
\end{cases}
\]

Let \( j \) be the smallest number row \( j \) where \( v_{j, t} = I \) for any \( t \). Construct a \((T + 1 - j) \times 1\) vector \( l \) where the \( i^{th} \) entry, \( l_i \), is the ordered pair \((i + j - 1, \hat{t}_j)\) for all \( \hat{t}_j > 0 \). Construct a second vector \( p \) of the same dimension where the \( p_1 = h((1, 1), l_1) \) and \( p_i = h((1, 1), l_i)) - p_{j-1} \cdot h(l_{j-1}, l_i) \) for all \( i > 1 \). The entry \( p_i \) is then the probability arriving at the node with the address \( l_i \) without passing through any previous node in the list. Then, the probability of terminal value, \( v_{j, T} \), at the start of the season for all \( v_{j, T} > I \).

\[
\Pr(v_{j, T}|n_{1, 1}) = h((1, 1), (j, T)) - \sum_{i=1}^{j+1-j} p_i \cdot h(l_i, (j, T))
\]

The probability of exercising the option is then 1 less the sum of the above probabilities in (6).
\[
pr(a_T) = 1 - \sum_j \text{pr}(n_{j,T}).
\] (7)

Armed with a full set of probabilities, given a set of model parameters, we can then estimate the parameters using Maximum Likelihood Estimation and simulate the effect of marginal changes on the parameters to the rate of abandonment.

## 4 A numeric example

For illustrative purposes, we investigate the properties and sensitivity of a specific parameterization of the model using parameter values that are reasonable for non-irrigated winter wheat in Kansas. The values we will use are the following:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g^* = b^*)</td>
<td>8.4</td>
</tr>
<tr>
<td>min yield</td>
<td>0</td>
</tr>
<tr>
<td>max yield</td>
<td>80</td>
</tr>
<tr>
<td>(I)</td>
<td>.75</td>
</tr>
</tbody>
</table>

For the vector \(k\) we will set the cost of the middle time period—\(k_T/2\) provided that \(T\) is even—to 9% of the \(t = 0\) expected nominal value of the crop at harvest. The remaining entries of the vector \(k\) are then set to zero. This is comparable to a single mid-season production cost used for an application herbicide and pesticide of $16 per acre which is based on 2018 Farm Management Guides for Non-Irrigated
Crops from Kansas State University’s AgManager website\(^6\). The percentage is found by multiplying the averages of bushels per acre and commodity price per bushel of wheat for the years 2000 to 2017. Mean, variance, minimum and maximum yield were determined by the Kansas county with the largest average yield from the years 1970 to 2017 (USDA-NASS, 2018). While 7 bushels per acre was the smallest average yield observed at the county level in the data, zero is assumed to a reasonable lower bound for an individual farm. Similarly, the number of green and brown balls (from the mean and variance assuming a symmetric yield distribution conditional on harvest) and maximum yield the pooled county level observations are distinct from actual farm-level variance and maxima, we assume that these values are reasonable approximations for some farms. For now we ignore the distribution of parameter values within pooled communities of farms (counties or NASS reporting districts). Figure 2 shows the probability mass function with the number of therms is set to 50.

The number of time periods was chosen so that the computations for simulation could be executed in a reasonable time. Increasing the number of time periods has the effect of making the probability mass of the yield distribution smoother, however it is unlikely that a farmer in actuality is updating their information and making decisions to abandon continuously, but rather at discreet intervals.
In Figure 3 we plot the probability of abandoning the crop by harvest time $T$ at insurance levels from none to 85% and the by mid-season cost from none to 20% of beginning of season expected nominal value of the crop—just over a doubling of our baseline. In Figure 4 we plot the probability of abandoning the crop by harvest time for the same levels of insurance and against different standard deviations of the non-augmented beta distribution (the distribution if the decision to abandon is removed). In both plots, baseline values of the example parameterization are shown by the dashed lines and the number of time periods in the model is $T = 150$ to give
sufficient smoothness when plotting as computational rounding error may produce noise when drawing the contours.

Figure 5: Contour plot of probability (%) of abandoning crop at varying levels of insurance and yield variance

Note: the contours were plotted using a resolution of 100 x 100. Together with rounding in the computation of each PMF may lead to some noise in the plot.

Figure 3 shows that, an increase 75% to 80% insurance coverage increases the probability that the farmer will abandon by 7.5 percentage points holding the baseline level of mid-season costs fixed, whereas decreasing coverage to 70% decreases the probability that the farm will abandon by 6.8 percentage points. Increasing (decreasing) mid-season costs by five percentage points, increases (decreases) the probability that the farmer will abandon by 6.9 (4.7) percentage points. The effect of changes in
mid-season costs generally increases as the level of insurance also increases. Figure 4 shows that increasing the variance from the baseline parameter values by 20%, increases the probability of abandonment by 2 percentage points and the effect on the probability of abandonment from increasing the variance stays relatively constant for the range of insurance levels plotted.

5 Limitations and future work

Our model makes a number of simplifying assumptions or omissions regarding insurance. It assumes that a farmer may abandon their crop and collect insurance for any reason. In practice, it may be that a farmer must show that they acted in good faith to receive insurance. Further, as written the model ignores the potential effect of insurance claims on future season insurance costs. This is likely not the case to the extent that insurance pricing is computed from the history of yields on a given plot. However, the discounted future cost of this consideration can easily be added to the model if a formula for determining future insurance costs were specified.

The model in its current state is useful for understanding how farmer behavior reacts to parameter changes conditional on a both the decision to plant the crop and the optimal production plan for the crop at the farm at the beginning of the season. It does not include a model for choosing which crop and weather to plant the crop. Similarly, our model does not address the endogenous choice insurance take-up. Instead, we have assumed risk-neutral farmers, who may be indifferent to purchasing actuarially fair insurance policies. Risk aversion strikes us as important dimension
to expand future work.

When taking the model to data it is important make two additions to the model. First, to the extent that farm practices incrementally improve over a time series, and therefore expected yields conditional on no abandonment have increased, the researcher must adjust re-scale yields over the series as some function of time. Fortunately, there are existing methodologies for trend adjusting yields (Edwards, 2012). Second, as stated, the model holds fixed the price that the crop commands throughout a given season. This can be addressed easily enough by including a geometric brownian motion model of the crop prices. The parameter $\sigma$, the probability of an upward step in price of size $\delta$ (fixed) in the next time period can be estimated from historical data or from futures options, as is commonly done in the real options literature. This addition would result in $T \times T$ possible harvest time revenue outcome nodes from the perspective of the initial time period.

References


