Welcome
Using R and Tableau at Worthington Industries:  
*Price optimization for high-mix, low-volume environments*

Steve Bartos  
Manager, Predictive Analytics  
Worthington Industries

Wil Davis  
Analyst, Predictive Analytics  
Worthington Industries
Agenda

Why pricing? A Brief History of Analytics at Worthington Industries

A Machine Learning Approach Misfire.

Power to the People! Let them choose with Tableau and R

The Complete Stack. Rserve, CI/CD and Version Control
Analytics at Worthington Industries
Worthington Industries

✓ Founded in 1955 and headquartered in Columbus, OH
✓ Publicly traded on the NYSE under the ticker WOR
✓ 10,000 employees & 5,000 customers; 80 facilities in 11 countries
✓ Employee, customer, supplier and investor-centered philosophy
✓ Leader in safety management and injury prevention – company wide goal of zero accidents and injuries
✓ Named one of “America’s Safest Companies” by Occupational Hazards magazine, 2008
✓ Named to Fortune’s “100 Best Companies to Work For” list five times

W&I Processed 7 MILLION TONS OF STEEL and is the largest purchaser of steel in the U.S. behind auto makers.

W&I is the largest alternative fuel cylinder and system supplier in the world, manufacturing 400,000 UNITS in FY2016.

WAVE produced 1 BILLION LINEAR FEET of ceiling grid in FY2016. That’s equal to 8 times around the world.

Worthington manufactured over 26 MILLION CYLINDERS joining alloys and accessories for industrial markets in more than 70 COUNTRIES.

WI Produced over 45 MILLION Balloon Time, Coleman, BernzOmatic and Worthington branded consumer products for jobsite, home and outdoor activities.
What if we had more accurate forecasts for these FGs?
Opportunities for another crown jewel of analytics?
Business Problem

While we are tracking the won, lost, and WIP quotes at a macro level, we do not currently have a system to identify potential opportunities to demand a premium to the market based on market, region, or product (item).
### Key Questions Addressed by Analytics

<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What happened?</td>
<td>What is happening?</td>
<td>What will happen?</td>
</tr>
<tr>
<td>Reporting</td>
<td>Alerts</td>
<td>Extrapolation</td>
</tr>
<tr>
<td><strong>Insight</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How and why did it happen?</td>
<td>What’s the best action?</td>
<td>What’s the best/worst that can happen?</td>
</tr>
<tr>
<td>Modeling</td>
<td>Optimization</td>
<td>Prediction</td>
</tr>
</tbody>
</table>

Source: Davenport et al. *Analytics at Work: Smarter Decisions, Better Results*
SCOPE tools: from historical pricing data to our won/lost quote data

“We have to find a way of making the important measurable, instead of making the measurable important.”
Robert McNamara, U.S. Sec. of Defense
### Key Questions Addressed by Analytics

<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Past</strong></td>
<td><strong>Present</strong></td>
<td><strong>Future</strong></td>
</tr>
<tr>
<td>What happened?</td>
<td>What is happening?</td>
<td>What will happen?</td>
</tr>
<tr>
<td>Reporting</td>
<td>Alerts</td>
<td>Extrapolation</td>
</tr>
<tr>
<td><strong>How and why did it happen?</strong></td>
<td><strong>What’s the best action?</strong></td>
<td><strong>What’s the best/worst that can happen?</strong></td>
</tr>
<tr>
<td>Modeling</td>
<td>Optimization</td>
<td>Prediction</td>
</tr>
</tbody>
</table>

Source: Davenport et al. *Analytics at Work: Smarter Decisions, Better Results*
A Machine Learning Approach
Price Elasticity and Optimization

A traditional Price Elasticity Curve...

Price Elasticity at Worthington Steel...
Model Design

- ~8,000 observations
- 150 original variables (50 usable)
- >500 engineered variables
Overfitting

Cold Rolled? yes
CR Strip? yes

Price < $50
Win

Price < $60

Lose
Win

Confounding variables!

Probability of Winning

Normalized Price
How should features be chosen to maximize accuracy and minimize overfitting?
Let the User Choose with Tableau (and R)!
Ability to analyze variances in absolute and percentage terms

User input for real-time market pricing

Support for sensitivity and what-if analysis of different price points
Users select which training observations to include/exclude. This improves business relevance and reduces the likelihood of overfitting.
Historical wins and losses by price are provided for context.

Tooltip provides individual quote data.

Chart type and layout chosen to imply logistic regression context.
Model identifies price that maximizes expected value

Expected value as a function of price

Model predicts win probability given simulated price

Leakage from sub-optimal pricing

Expected value defined as:

\[
EV = Price \times P(Winning)
\]
Model Status:

- Model is valid!

Traffic light indicates model quality

Tooltip provides additional context and statistical measures
What makes a successful journey?

• Strong **engagement** from the business

• **Alignment** across functional areas (sales, analytics, IT) – are we solving a real business problem?

• Introduce complexity **in pieces** to reduce learning curve

• Allow users to **see their data** (in addition to the model)
Tableau and R Technology Stack
1. Develop and push model as an R package
2. On merge to master, build to an internal CRAN mirror
3. Install/update package on Rserve server from CRAN
4. Call package and function from R script in Tableau
Call package and function from R script in Tableau

```
pe_lm <- function(var, win_status) {
  w1 <- factor(win_status, levels = c("lost", "won"))
  df <- data.frame(var, w1)
  wts <- round((1-table(w1)/length(w1)) * 100)
  df$wts <- as.numeric(wts[w1])
  fit <- glm(w1 ~ var, family = "binomial", data = df, weights = wts)
  price_var_vec <- seq(min(var), max(var), by = (max(var) - min(var)) / (length(var) - 1))
  price_var_vec <- price_var_vec[rank(fit$data$var, ties.method = "first")]
  ndf$price_var_vec <- data.frame(var = price_var_vec)
  ndf$win_prob <- predict.glm(fit, newdata = ndf, type = "response")
  out <- list(ndf = ndf, fit = summary(fit))
  out$status <- pe_lm.status(out)
  return(out)
}
```

Results are computed along Quote Key.

```
SCRIPT_REAL(''
library(pe_lm)
x <- pe_lm(
  var = .arg1,
  win_status = .arg2
)
x$ndf$win_prob
',
'AVG([Variance.ParamCtrl]),ATTR([Status]])
')
```
Why build and maintain this infrastructure?

- **Efficiency** of continuous integration and continuous deployment
- **Stability and reproducibility** of version control
- Automated **testing** in R (testthat package)
- **Performance** gains from distributed and concurrent processing – R and Tableau are running on separate servers
- **Security** of a CRAN repository that is behind the firewall
What’s next?

1. Temporally-weighted training data to improve responsiveness to market changes

2. Incorporate CRM data to improve customer and market context

3. Clustering and classification trees to recommend market segmentations (filter selections)
Thank you!

Contact or CTA info goes here

Steve Bartos
@OldDirtyBarGraph
Steve.bartos@worthingtonindustries.com

Wil Davis
github.com/wkdavis
William.davis@worthingtonindustries.com