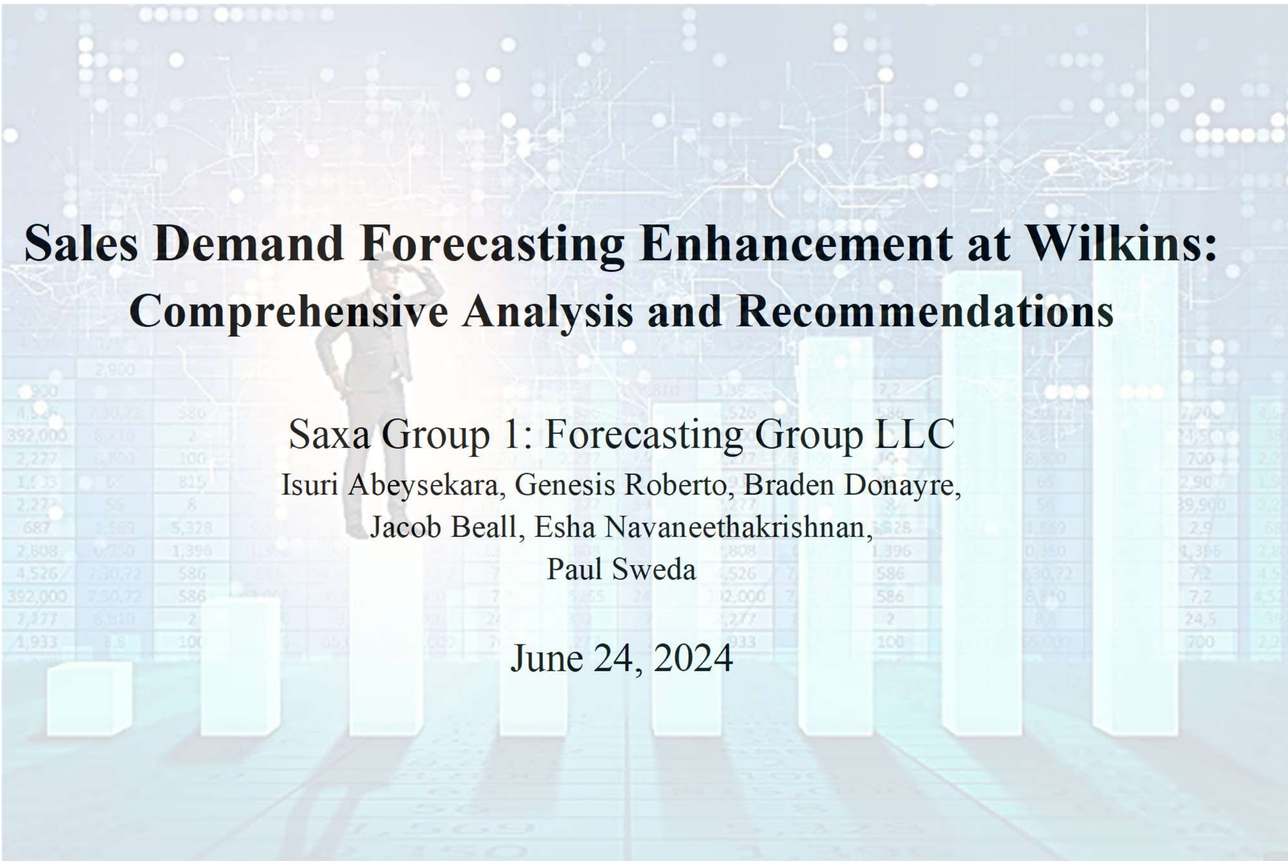


Sales Demand Forecasting Enhancement at Wilkins:
Comprehensive Analysis and Recommendations
Saxa Group 1: Forecasting Group LLC
Isuri Abeysekara, Genesis Roberto, Braden Donayre,
Jacob Beall, Esha Navaneethakrishnan,
Paul Sweda
June 24, 2024

1. Executive Summary

This paper delves into the challenges and strategies of demand forecasting for Wilkins Regulator Company (Wilkins), a Zurn Company. The study analyzes the existing forecasting methods, evaluates the accuracy of the Q1 2005 forecast, and investigates the impact of any forecast inaccuracies. We take it a step further and deploy three other forecasting techniques: ARIMA, Naive and Moving average, to enable comparisons with the current forecast methods, based on statistical analysis and historical quarterly data. We considered the impact of trends and seasonal patterns, correlation between product demand and various economic indicators, such as the number of housing starts. Vital time series statistical tools such as Autocorrelation (ACF) and Partial Autocorrelation Function (PACF) allowed us to dissect the structure of the data, laying the groundwork for a comprehensive comparison of the Naive, Moving Average, and ARIMA and current forecast model. Through this in-depth analysis, we aim to equip Bernie Barge to leverage the ARIMA model to forecast the next 3 quarters of this year.

2. Introduction

Bernie Barge, the newly promoted inventory manager, has been tasked to forecast sales demand for an upcoming meeting in January 2025. He wonders if there is a more reliable and efficient method to predict the demand of the plant's products, specifically the Pressure Vacuum breakers (PVB) and Fire valve (FV), within the irrigation and fire production product family, respectively. As a new manager, Bernie wants to bring fresh ideas to the current forecasting team, comprised of Chris Connor, General Manager and Rick Fields, Sales Manager. We've been brought in to assist Bernie in evaluating current forecasting practices and determine if a better forecasting model could be deployed to better predict demand, and ultimately, maximize sales.

3. Current State Assessment

We observed that Wilkins General Manager and Sales/ Marketing manager meet on quarterly basis to forecast demand for each product family, leveraging an approached dubbed as the "Forecasting Master", which relies on knowledge of industry trends, competition strategies and product sales history. This approach primarily relies on average unit sold in the last four quarters x projected daily sales for the next 12 months. Then, a disaggregated view of individual units are computed as a % of total product family, in order to apply somewhat precise level of forecast to individual products. The "Forecasting Master" system, while seemingly straightforward, may lack the necessary sophistication to ensure accurate demand predictions for various product lines. While this approach offers a degree of simplicity, it introduces the potential for inaccuracies stemming from inherent biases, lack of comprehensive data analysis, or even planning for seasonal spikes as the current forecast only looks at the past four quarters. This can lead to a cascade of negative consequences, impacting inventory management, production efficiency, and ultimately, missed sales and customer satisfaction. The limitations of the current system were demonstrably evident in Q1 2005. The PVB product family forecast projected unit sales of 53,560, exceeding actual units sales of 48,519, a miss by 5,041 units or 9.4%. This overestimation signifies a potential for significant financial repercussions, which could lead to several consequences, such as:

- **Inventory Carrying Costs**: Over-forecasting translates to excess inventory, incurring storage, handling, and potential obsolescence costs.
- **Stock-Outs:** Under-forecasting can lead to stock-outs, resulting in lost sales opportunities and customer dissatisfaction.
- **Missed Sales Opportunities:** Inaccurate forecasts can hinder Wilkins' ability to capitalize on market demand, particularly for their newly launched "Fixed Pressure Fire Valve."
- **Production Inefficiencies:** Inaccurate forecasts create production planning challenges, leading to inefficiencies such as idle time, overtime, and scheduling issues.

To mitigate these risks and optimize resource allocation, Wilkins should consider exploring more data-driven forecasting models that incorporate historical sales data, seasonality trends, and external economic factors such as unemployment rate, bank prime loans and housing starts.

4. Relationship Analysis

To assess the relationship between the demand for the PVB product family (total) and the Fire Valve (FV) product family with external economic factors, we conducted a simple linear regression analysis. Our findings, as seen on Table 1, indicate that while the unemployment rate and bank prime loan rate show non-significant p-values, suggesting

Table 1: Relationship with Exogenous variables

Linear regression p-value results	Unemployment Rate	Bank Prime Loan	Total Housing Starts
Total PVB	0.97840	0.99470	0.00409 **
Total FV	0.865	0.952	0.410

*** significant p-value*

weak evidence of a direct relationship with PVB demand, the Total number of Housing Starts exhibits a significant p-value of 0.00409. This significant finding suggests a strong positive relationship between the number of housing starts and demand for PVB products. Additionally, Table 1 revealed significant limitations in establishing a strong relationship between the Fire Valve product family and these economic variables, as indicated by low statistical significance values. This implies that neither the Unemployment Rate, total housing starts, nor the Bank Prime Loan Rate significantly influence demand for Fire Valve products. These results emphasize the critical importance of understanding and monitoring the Total Number of Housing Starts as a pivotal indicator of construction activity and its direct impact on demand for PVB plumbing products. This insight is crucial for accurately forecasting and managing inventory levels to meet market demand effectively.

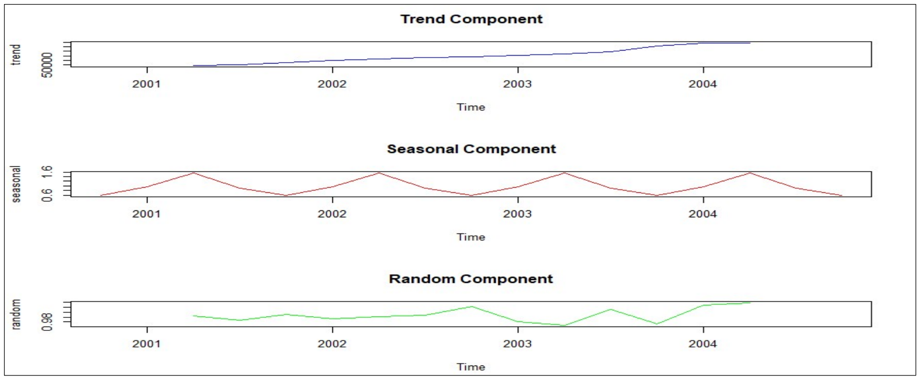
5. Data Analysis

Creating an initial model. To create a demand forecast for the PVB product family for the next three quarters of 2005, we fitted both an ARIMA (Autoregressive Integrated Moving Average) and an ETS (Error, Trend, and Seasonality) model to the historical sales data. The ARIMA model, selected using the 'auto.arima' function, and the ETS model, fitted using the 'ets' function, both provide forecasts indicating the expected total PVB sales

As demonstrated above, the ETS forecast plot shows predicted sales values with confidence intervals for the next three quarters of 2005, capturing the seasonal and trend components of the data. The ARIMA forecast plot, generated with similar steps, also shows predicted sales for PVB with confidence intervals, indicating an overall decline in sales as the quarters progress.

Analyzing trend, seasonal and random patterns. The analysis of demand for the PVB product family from 2001 to 2004 reveals several key patterns as seen on ***Figure 1***. The ***trend*** component of the demand exhibits a consistent upward trajectory throughout the observed period, with a slight leveling off noted towards the end of 2004, suggesting a potential stabilization in demand growth.

Figure 1: Trend, Seasonal and Random Patterns



Seasonal variations occur regularly quarterly, with demand peaking in the first quarter and reaching troughs in the third quarter each year. This recurring pattern indicates a strong seasonal impact on demand, likely influenced by factors such as construction seasons or other industry-specific cycles.

The ***random*** component, which represents unexplained fluctuations in demand, appears minimal. This suggests that the observed variations in demand are predominantly accounted for by the identified trend and seasonal patterns. Overall, the analysis indicates increasing demand trends over time, punctuated by predictable peaks and dips throughout the year.

6. Modeling and Analysis

Model significance (R^2, Beta coefficients). To quantify the trend observed in the PVB demand data, a linear regression model was applied using the 'TimeIndex' as the predictor variable for quarterly demand. As seen on *Figure 2*, the regression analysis yielded an R-squared (R^2) value of 0.08218, indicating that approximately 8.2% of the variance in demand can be explained by the linear trend captured by 'TimeIndex'. However, the adjusted R-squared (adjusted R^2) of 0.021 suggests that the model's explanatory power is limited.

```
Call:
lm(formula = `Total PVB` ~ TimeIndex, data = combined_data)

Residuals:
   Min     1Q Median     3Q    Max
-25715 -20533  -4770   8701  56350

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -11324008    9821938  -1.153    0.267
TimeIndex        5684       4904   1.159    0.265

Residual standard error: 24860 on 15 degrees of freedom
Multiple R-squared:  0.08218,	Adjusted R-squared:  0.021
F-statistic: 1.343 on 1 and 15 DF,  p-value: 0.2646

R-squared: 0.08218452
Beta (coefficient) estimate: 5683.895
Model summary:

Call:
lm(formula = `Total PVB` ~ TimeIndex, data = combined_data)

Residuals:
   Min     1Q Median     3Q    Max
-25715 -20533  -4770   8701  56350

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -11324008    9821938  -1.153    0.267
TimeIndex        5684       4904   1.159    0.265

Residual standard error: 24860 on 15 degrees of freedom
Multiple R-squared:  0.08218,	Adjusted R-squared:  0.021
F-statistic: 1.343 on 1 and 15 DF,  p-value: 0.2646
```

Figure 2: Model significance results

Further assessment using the F-statistic ($F = 1.343$, $p = 0.265$) indicates that the model is not statistically significant at the conventional significance level (α = 0.05). These findings suggest that the linear regression model with 'TimeIndex' alone may not fully capture the complex dynamics of demand variation for the PVB product family. Additional preprocessing or transformation steps may be necessary to enhance the model's predictive capability.

7. Time Series Analysis

The Autocorrelation Function (ACF) plot for PVB demand on *Figure 3* shows a gradual decay as the lag increases, with significant spikes observed at lag 2 and lag 4, while other spikes remain within the range of +/- 0.5. This pattern suggests strong autocorrelations at these lags, indicating potential seasonality or recurring patterns in demand data.

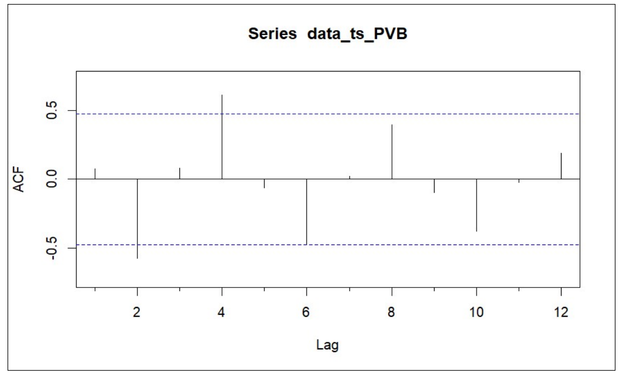


Figure 3: ACF Plot for PVB products

The Partial Autocorrelation Function (PACF) plot on *Figure 4* highlights a significant negative spike at lag 2, with minimal significant correlations at smaller lags. This suggests that the direct influence of earlier observations on current demand is primarily captured by lag 2, indicating a possible AR(2) process in the demand data.

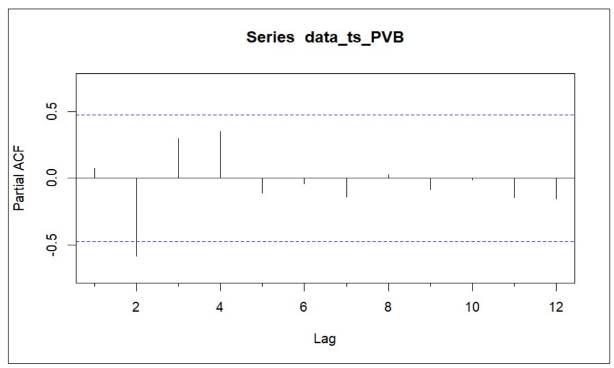


Figure 4: PACF Plot for PVB products

These ACF and PACF analyses provide insights into the temporal dependencies within the PVB demand data, suggesting the presence of seasonal effects and potential ARIMA model specifications that could be explored further to improve forecasting accuracy.

8. Comparison of Forecasting Approaches

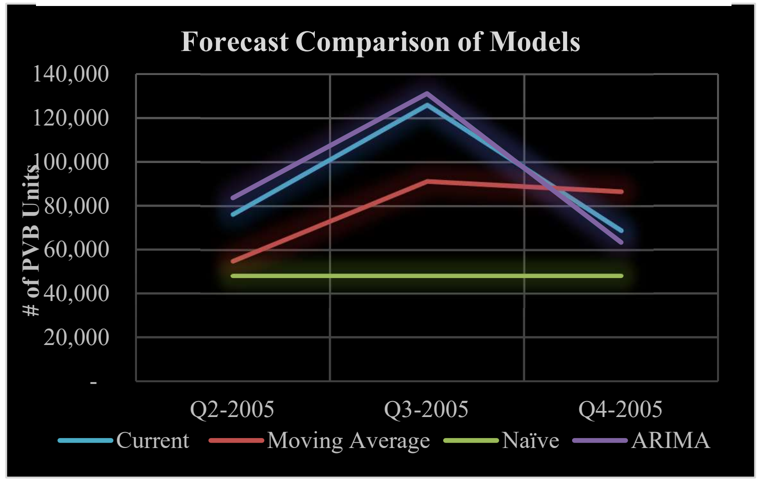
Comparison of time-series models. A naive model assumes that the next period's demand will be exactly the same as the last period. As such, naive forecast values are simply equal to the most recently observed value.

Figure 5: Quarterly forecast of Units by Model Q2-Q4 2005

Current <dbl>	Moving Average <dbl>	Naive <dbl>	ARIMA <dbl>
76272.27	54929.77	48159	83457.92
126037.51	91256.73	48159	131106.92
68843.42	86634.94	48159	63234.92

For this data set, this means the naive model forecasts future quarter sales as equal to the last reported quarters' sales. As seen in the table above, the naive model forecasted the next 3 quarters sales of PVB to be equal to 48,159 units. A Moving Average model uses the average sales over the past 'x' periods to forecast the future sales. This method of averaging the demand over a specific period of time allows the model to smooth out the impact of short-term fluctuations and focus on the longer-term trends from the historical data. In our moving average model, we used the previous six months (2 quarters) average to forecast the future quarter sales. In the table above, we can see that the moving average model forecasted lower sales for all three of the next quarters. Lastly, Autoregressive Integrated Moving Average (ARIMA) models are special type of regression models in which the dependent variable has been stationarized and the independent variables are all lags of the dependent variable (AR-component) and/or lags of the errors (the MA-component). ARIMA models are a great tool for time series analysis and forecasting, but ARIMA models do not account for exogenous variables, like the unemployment rate and housing starts.

Figure 6: Quarterly forecast of Units by Model Q2-Q4 2005



As seen on ***Figure 5*** above, we forecasted and obtained the # of units for the next 3 quarters using ARIMA, Naive and 2-MA against the current forecast method. We plotted forecast on a graph as seen on ***Figure 6***. It is obvious from the graph that ARIMA(purple line) proved to be the best forecasting model, as it captured the 3 quarters closest to the current forecast predictions (blue line).

When we compare the ARIMA model to the current forecast (see ***Figure 6***), we see that it follows the trending spike occurring in Q3-2005, and then dipping back to Q2 level at the end of 2005. As we dig deeper into the numbers (see ***Table 2***), we see that ARIMA

Table 2: Comparing Forecasted units: Current and ARIMA

Period:	Current model	ARIMA model	Differences: ARIMA vs Current *Unit change*	*% change*
Q2-2005	76,272	83,457	7,185	9%
Q3-2005	126,037	131,106	5,069	4%
Q4-2005	68,842	63,234	(5,608)	-8%
Total	271,151	277,797	6,646	2%

forecasted 9% higher in Q2, 4% higher in Q3, and 8% lower in Q4 when compared to current, with an overall 2% higher forecast than current forecast, as seen on ***Table 2.*** This tells us that ARIMA could be our best model to address potential stock-out issues while also ensuring that Wilkin's does not over forecast.

We also compared the AIC for each of the models, as on ***Figure 7***. ARIMA had the lowest AIC of 273.69, Moving Average (322.68), Current forecast (AIC of 340.03) and lastly Naive at 431.07. Because we could not obtain the AIC value for Naive model, we obtained the SSE (sum of squared errors) as the next best comparative statistic to use for comparative purposes. Given the below values, ARIMA remains the superior model.

Model <chr>	Metric <chr>	Value <dbl>
ARIMA	AIC	273.6994
Naive	SSE	431.0719
Moving Average	AIC	322.6827
Current	Forecast Value	340.0354

Figure 7: Comparison of AIC values between models

Forecast at Product family level or individual products. The choice between forecasting demand with an ARIMA model at the individual product level versus the product family level depends on several factors, as summarized in ***Table 3***. Forecasting demand at an individual product level provides insights into the demand patterns, seasonality, and trends specific to each product. This specificity is valuable for products that have unique customer segments and varying trends and seasonality. This method facilitates precise demand planning & inventory management, minimizing stockouts and excess inventory situations by addressing the specific needs and fluctuations of each product. While there are clear benefits to forecasting at the individual product level, there are limitations for small companies in regard to the resources they have available for demand forecasting and the data they have access to. Individual product demand forecasting requires sufficient historical data for each individual product, which may be challenging for new products like the new Fire Valve product. Forecasting at the product family level can prove beneficial as it provides demand patterns across multiple products, smoothing out

Table 3: Pros & Cons of Forecasting at Product Family Vs. Individual Item

	Product Family	Individual item
Benefits	• Takes less time, resources • Enables strategic planning • Smoother trend • Reduced computation cost	• Allows for precise forecasting • Enables targeted strategies • Improved resource allocation
Challenges or Limitations	• Loss of detail and less targeted planning • Potential for unit stockouts	• Takes time and effort • Reliant on data availability • Increased model complexity and computation demand

noise and variability at the individual product level. This stability provides a clearer picture of overall market trends and broader demand patterns for management to make important decisions about inventory management. By avoiding the complexity of individual product forecasting, management can improve their ability to make dynamic decisions about inventory each quarter.

9. Conclusion and Recommendations

We recommend Bernie adopt advanced statistical models like ARIMA and ETS for demand forecasting due to their ability to leverage historical data, capture trends, and account for seasonality, resulting in more reliable forecasts compared to current methods. Integrating economic indicators such as housing starts will further enhance accuracy, given our findings of a significant correlation with PVB demand.

These recommendations are based on the improved accuracy and comprehensiveness of these models, which effectively capture trends and seasonal patterns. Incorporating economic indicators ensures forecasts align with broader market conditions, enhancing resilience and reliability. Regular validation and adjustments will keep forecasts responsive to market dynamics, minimizing risks of over- or under-forecasting. Developing individual forecasts for each product variant will provide detailed insights, optimizing inventory management and reducing costs.

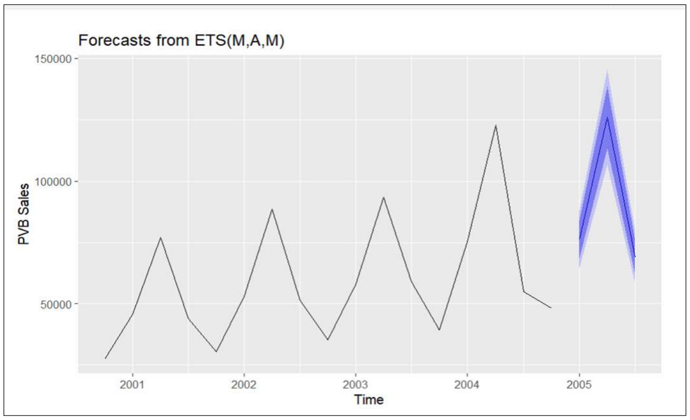
To convince management, we propose presenting data-driven comparisons of forecast accuracy between current methods and advanced models. Emphasizing the correlation between housing starts and PVB demand will further drive the need for integrating economic indicators. Quantifying financial benefits, such as cost savings and revenue gains from improved inventory management, will further support the case.

We also recommend a pilot program to implement these models over a few quarters, demonstrating tangible benefits firsthand. Addressing implementation risks, we propose a structured integration plan that includes team training and data analyst support.

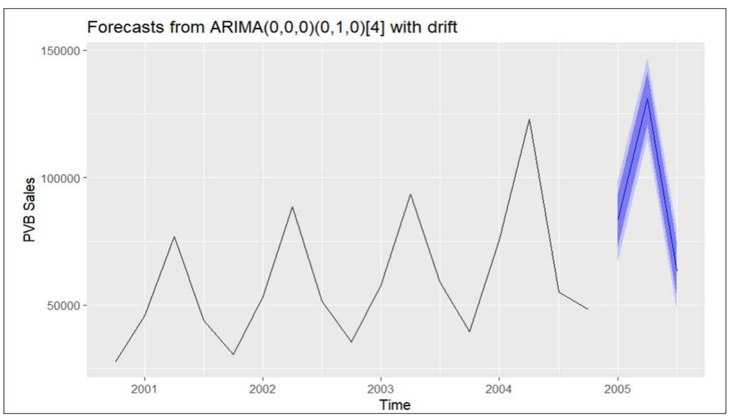
By presenting compelling data, financial benefits, and a structured implementation strategy, we believe management will recognize the value in adopting advanced forecasting methods. This will lead to more accurate demand forecasts, optimized inventory management, and improved operational efficiency for Wilkins.

Appendix

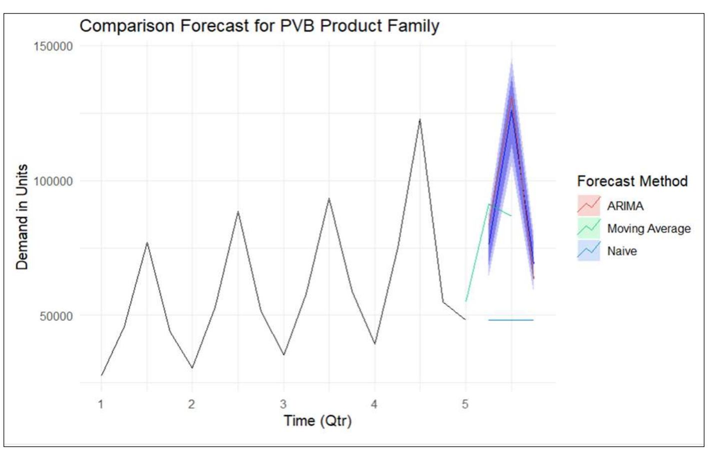
Appendix I: PVB Sales forecast - Moving Average model



Appendix II : PVB Sales Forecast - Arima Model



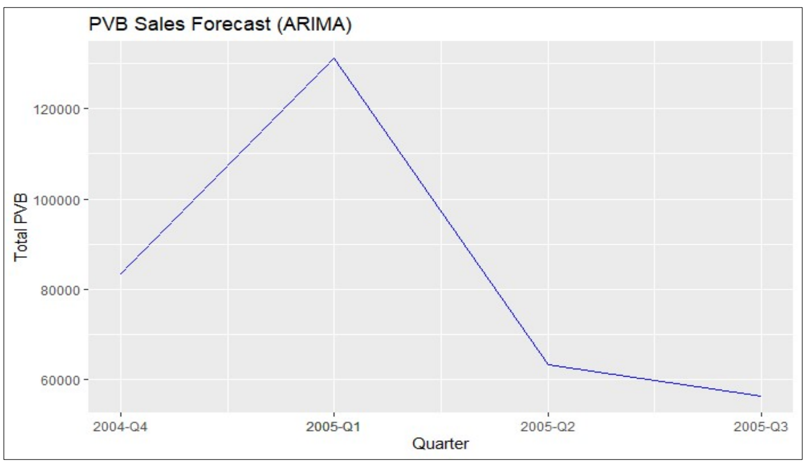
Appendix III : PVB Product Family Forecast - Comparison



Appendix IV : AIC - Comparison

```
AIC for ARIMA model: 273.6994
AIC for Naive model: 431.0719
AIC for MA model: 322.6827
AIC for Current model: 340.0354
```

Appendix V : PVB Sales Forecast (ARIMA)



Appendix VI : PVB Sales Forecast Current Model