Sales Forecasting At “James & Sara Dairy Farm”

Posted on October 10, 2019

Introduction

Sales forecasting helps a farm translate historical records into purchasing, staffing, cash-flow, production, and distribution decisions. The original project analyzes thirty-six months of sales from James and Sara Dairy Farm and observes a clear seasonal pattern: sales are generally stronger from October through March and weaker during late spring and summer. It then creates monthly dummy variables, estimates a seasonal regression, adds a time trend, compares mean squared error, and forecasts months thirty-seven through forty-eight. That analytical sequence is appropriate, but the explanation needs greater statistical precision. A model that fits historical data well does not necessarily forecast well, and the complete thirty-six observations are not reproduced in the article, so the reported coefficients cannot be independently recalculated. This expanded analysis preserves the original equations and family-business context while explaining how the models work, how to validate them, how to convert forecasts into operating decisions, and how uncertainty should be communicated.

The Business Question

The farm does not need a forecast merely to know whether next December will be busier than next June. It needs estimates that support decisions about feed, veterinary supplies, processing capacity, packaging, transport, labor, refrigeration, and working capital. The forecasting horizon is twelve months because the family wants a plan for the fourth year after observing three years of monthly sales. Monthly data are suitable for identifying annual seasonality, although thirty-six observations remain a small sample. The business should therefore use a transparent model that can be updated each month and should avoid treating a single point forecast as certainty.

Understanding the Historical Pattern

The first step is visual inspection of the time series in chronological order (Hyndman & Athanasopoulos, 2021). The original project reports higher sales from October through March and lower sales during the remaining months. That pattern may reflect weather, household consumption, school calendars, holiday demand, milk yield, pricing, customer routines, or local distribution conditions. A chart should also be examined for trend, unusual spikes, sudden level changes, and changes in seasonal amplitude. If high-season sales become progressively larger while low-season sales remain stable, a multiplicative model may fit better than an additive model. If all months rise by approximately the same number of units each year, an additive seasonal regression is reasonable.

Data Quality Before Modeling

Forecast quality cannot exceed record quality. The farm should confirm that all thirty-six figures use the same definition of sales: units, revenue, liters, or another measure. Revenue data can rise because of price changes even when physical demand falls, so volume and value should ideally be modeled separately. Missing invoices, returns, spoiled products, bulk contracts, and one-time sales should be identified. Dates must reflect when a sale occurred rather than when payment was received. The accounts department should retain a clean monthly series and a note for exceptional events. These notes become essential when a future model appears to “miss” a month that was actually affected by a festival, disease event, equipment failure, or supply interruption.

Encoding Monthly Seasonality

A regression model cannot use month names directly without converting them into indicator variables. The original analysis creates eleven monthly dummies and leaves December as the reference category. January equals one for January observations and zero otherwise, February behaves similarly, and the remaining indicators follow the same rule. December is represented when all eleven indicators equal zero. One month must be omitted to avoid perfect multicollinearity, commonly called the dummy-variable trap. The intercept therefore represents estimated December sales, while each monthly coefficient represents the average difference between that month and December.

Interpreting the First Seasonal Model

The original seasonal equation is reported as:

$$\text{Sales} = 279.666 + 3.333\ \text{Jan} - 15.667\ \text{Feb} - 40.667\ \text{Mar} - 65\ \text{Apr} - 97\ \text{May} - 131.333\ \text{Jun} - 136.667\ \text{Jul} - 109.333\ \text{Aug} - 77.667\ \text{Sep} - 43.667\ \text{Oct} - 22.333\ \text{Nov}.$$

With December as the reference, the intercept predicts approximately 279.7 sales units in December. January is estimated about 3.3 units above December, while July is about 136.7 units below December. The equation confirms the visual seasonal pattern. Its coefficients should not be read as causal effects of the calendar month. They summarize recurring differences associated with month after averaging across the three observed years.

Reconstructing Monthly Predictions

To forecast a January under the seasonal-only model, the farm adds the January coefficient to the intercept, producing approximately 283.0 units. A June estimate is approximately 148.3 units, and a July estimate is approximately 143.0 units. The model gives the same prediction for every January because it contains no trend. That feature is useful if demand is stable, but it becomes unrealistic when sales are gradually increasing or decreasing. The residual for each observation is actual sales minus predicted sales. These residuals reveal whether the model systematically underestimates recent years and overestimates early years, which would indicate an omitted trend.

Adding a Time Trend

The second model adds a sequential time variable representing months one through thirty-six. The reported equation appears to contain a typographical omission after the coefficient 1.05444. Interpreted as a trend model, it is:

Sales = 286.33 + 1.05444(Time) + 24.333 Jan + 22.333 Feb + 24 Mar + 5 Apr − 30.333 May − 64.667 Jun − 43.333 Jul − 68.333 Aug − 51.333 Sep − 33.667 Oct − 16.333 Nov.

The time coefficient indicates an estimated increase of about 1.05 sales units per month, or about 12.65 units per year, holding seasonal position constant. The monthly coefficients now describe differences from December after allowing for the upward trend. Because the trend and seasonal effects are estimated together, these coefficients should not be compared mechanically with the first equation without examining model diagnostics.

Forecasting Months Thirty-Seven Through Forty-Eight

If observation one is January of the first year, observation thirty-seven is January of the fourth year. The January forecast becomes 286.33 + 1.05444(37) + 24.333, or approximately 349.7 units. February uses time 38 and the February dummy, producing roughly 348.7 units. March uses time 39 and its coefficient, producing about 351.5 units. December uses time 48 with all monthly dummies equal to zero, yielding about 336.9 units. These calculations illustrate the method, but the precise calendar alignment must be confirmed. If the first observation was not January, the dummy assignment for periods thirty-seven through forty-eight must follow the actual sequence.

Why the Trend Model Appears Better

The original project states that the seasonal-and-trend model has a lower mean squared error than the seasonal-only model. Mean squared error averages squared residuals, giving larger mistakes more weight (Newbold et al., 2013). A lower in-sample MSE indicates closer fit to the thirty-six observations. Adding a trend almost always improves or preserves in-sample fit because it gives the model another explanatory variable. The improvement becomes meaningful only if it is large enough and survives evaluation on data not used to estimate the coefficients. Adjusted R-squared, standard error, residual patterns, and coefficient uncertainty should also be considered.

Creating a Holdout Test

A more credible comparison would estimate each model on the first twenty-four months and test forecasts on months twenty-five through thirty-six. The farm could compare mean absolute error, root mean squared error, and mean absolute percentage error across the twelve holdout months (Makridakis et al., 2022). This procedure imitates the actual task of forecasting an unseen year. If the trend model fits the first three years beautifully but performs poorly on the holdout, it may be overfitting or responding to a temporary rise. With only thirty-six observations, rolling-origin evaluation can provide several tests: estimate through month twenty-four and forecast month twenty-five, then extend the sample one month and repeat.

Residual Diagnostics

Residuals should appear approximately random around zero (Montgomery et al., 2015). A sequence of positive residuals followed by negative residuals suggests missing trend or level change. A repeated seasonal wave suggests inadequate monthly effects. Large isolated residuals require investigation rather than automatic deletion. Autocorrelation matters because consecutive sales months may be related after accounting for seasonality. If residuals remain correlated, standard regression estimates of uncertainty may be unreliable, and a seasonal time-series model such as exponential smoothing or seasonal ARIMA may improve forecasting.

Alternative Forecasting Models

The dummy-variable regression is attractive because every coefficient is understandable. The farm should nevertheless compare it with simple benchmarks. A seasonal naïve forecast sets each month equal to the same month in the previous year. A moving average smooths recent observations but can lag trend. Holt-Winters exponential smoothing estimates level, trend, and seasonality while giving greater weight to recent data. Seasonal ARIMA models can capture autocorrelation but require more observations and careful diagnosis. The best model is not the most complex; it is the one that performs reliably on unseen data and can be maintained by the business.

Forecast Intervals

A point estimate such as 350 units creates false precision if presented alone. Forecast intervals communicate a plausible range based on historical error and model uncertainty. A high-season month might be forecast at 350 units with an interval wide enough to include materially lower or higher demand. Inventory and staffing plans can then use service-level decisions rather than the exact midpoint. The family might plan ordinary production around the central estimate, secure flexible labor for the upper range, and maintain supplier arrangements that protect against shortage without holding excessive perishable inventory.

Separating Demand From Capacity

Recorded sales can be lower than demand when the farm lacks product, transport, refrigeration, or staff. A forecasting model built only on completed sales will reproduce the constraint and underestimate potential demand. The accounts data should therefore be combined with stockout records, unfilled orders, returns, and lost customers. Likewise, unusually high sales caused by discounting may not represent normal full-price demand. Forecasting should distinguish what customers wanted from what the farm was able or willing to supply.

Price, Volume, and Product Mix

If “sales” means revenue, the upward trend may partly reflect price inflation rather than increasing volume. The farm should model liters or units by product and then apply expected prices to create a revenue budget. Milk, yogurt, cream, butter, and other products may have different seasonal patterns and shelf lives. An aggregate forecast can be accurate while concealing a shortage in one category and excess in another. Product-level forecasts should be reconciled with the total so that management understands both the broad cash-flow outlook and operational detail.

Operational Use: Feed and Herd Planning

Demand forecasts should inform, but not mechanically dictate, biological production. Milk supply depends on herd size, lactation cycles, feed quality, animal health, weather, and veterinary management. The farm can compare expected demand with a separate supply forecast. When demand exceeds expected production, options include adjusting contracts, purchasing from qualified partners, shifting product mix, or communicating availability. Increasing herd size for one high forecast would be risky because livestock decisions have long lead times and continuing costs. Scenario planning is more appropriate than a single deterministic response.

Operational Use: Processing and Cold Storage

Higher winter sales may require additional packaging, refrigeration, delivery capacity, and maintenance preparation. Equipment failure during a peak month has a larger opportunity cost than failure in a low month. Preventive maintenance can therefore be scheduled before the forecasted high season. During lower-demand months, the farm can consider products with longer shelf lives, promotional activity, or planned downtime, provided these decisions remain economically justified. The forecast should connect each month with capacity utilization rather than remain only an Excel output.

Operational Use: Labor and Cash Flow

Seasonal forecasts help schedule temporary labor, driver hours, and administrative work. They also support a cash budget. The farm can estimate receipts, feed purchases, wages, packaging, utilities, loan payments, taxes, and emergency reserves. A growing sales trend may require more working capital because production and packaging costs are paid before customers settle invoices. The forecast should be integrated with payment timing and credit risk. Revenue growth does not automatically mean improved cash availability.

Communicating the Forecast to the Family

The family presentation should begin with the observed seasonal pattern, then explain the two models in ordinary language. December is the reference month, monthly coefficients show typical differences, and the trend estimates gradual growth. A chart should display historical sales, fitted values, and the twelve forecasts with intervals. Management should be told which assumptions can fail: the seasonal pattern may change, price increases may alter demand, a major customer may leave, or production may be constrained. The purpose is to support decisions and review, not to certify the future.

A Monthly Updating Routine

After each month, the farm should enter actual sales, calculate forecast error, record exceptional events, and refresh the model. A dashboard can track cumulative error and bias. Persistent underforecasting means the model or assumptions need revision; persistent overforecasting can produce waste and cash strain. The farm should formally reselect the model at least annually or when a structural change occurs. Retaining prior forecasts prevents hindsight bias and allows the family to learn whether decision quality improves.

Limitations of the Current Project

The article provides only the first twelve observations, not the complete thirty-six-month series, regression output, standard errors, MSE values, or forecast table. The reported coefficients therefore cannot be independently verified from the published information. The trend equation also appears to omit the word “Time.” These limitations do not invalidate the approach, but they require transparent qualification. A complete submission should attach the original spreadsheet, define the sales unit, show all observations, preserve formulas, and document the exact Excel regression settings.

Conclusion

The James and Sara Dairy Farm data show recurring monthly seasonality and an apparent upward trend. Monthly dummy variables provide a clear way to compare each month with December, while a sequential time variable allows sales to increase across years. The reported trend-and-seasonal model fits the historical data better according to MSE and can generate forecasts for periods thirty-seven through forty-eight. However, in-sample fit is not enough. The farm should validate the model on unseen months, inspect residuals, compare it with seasonal naïve and exponential-smoothing alternatives, and report forecast intervals. The forecast becomes valuable only when connected with feed, herd capacity, processing, cold storage, labor, transport, product mix, and cash flow. Updated monthly and interpreted with uncertainty, the model can become a practical family-management tool rather than a one-time spreadsheet exercise.

References

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting, 38*(4), 1346–1364.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting* (2nd ed.). Wiley.

Newbold, P., Carlson, W. L., & Thorne, B. M. (2013). *Statistics for business and economics* (8th ed.). Pearson.