**HUYNH DO**

**Module 8, Market 2**

Choose at least two series from the broadband to perform Time Series and complete the following prompts:

1. **Create a sequence chart to visualize the series that you’ve selected and describe any trends or seasonality you observe.**



**Observations:**

1. **Trend**: The data shows a general upward trend from 1999 to 2001, indicating growth in market value over this period.
2. Plateau and Variability: The market faces increasing volatility, including periods of standstill and slight falls, as the expansion slows around the middle of 2001.
3. **Recovery**: From late 2002 to 2003, the market resumes an upward trend, ending at a higher number of subscribers.

**Seasonality:**

There is no clear seasonal pattern in this dataset, as the fluctuations seem irregular and not tied to specific intervals or time of the year.

1. **Create a second sequence chart choosing the difference option to remove the trend from the time series.**



**Observations from the Differenced Sequence Chart:**

This differenced approach is useful for analyzing the underlying behavior of the series without the influence of long-term trends.

1. **Variability**: The differenced values highlight periods of high variability, with noticeable spikes in both positive and negative directions.
2. **No Clear Pattern**: The differenced series does not display a consistent pattern or seasonality, suggesting irregular fluctuations in market value.
3. **Volatility Clusters**: There are periods where changes are relatively stable, followed by periods of heightened volatility, particularly around 2001-2002.
4. **Run the expert modeler on the series that you’ve selected and interpret the results**
* Explain the type of model selected and it's parameters, e.g., if a Holt model is selected, explain that it is an Exponential Smoothing model and the meaning of the alpha and gamma parameters. If an ARIMA model is selected --e.g., ARIMA(1,0,1) -- explain the components.
* **Hold Model With Forecast**

****

The chart demonstrates the application of Holt's model on the data, showing:

1. **Original Data**: The observed values from January 1999 to December 2003.
2. **Holt Fitted Values**: The model's fitted values, which closely follow the original data's trend.
3. **Forecast**: A 12-month projection beyond December 2003, indicating a continuation of the upward trend.

#### Forecasted Values:

* January 2004: 55,222
* February 2004: 56,817
* March 2004: 58,412
* April 2004: 60,007
* **Hold Model Parameters**
1. **Alpha (α)**: The **smoothing parameter for the level**.
	* Controls how much weight is given to the most recent observation relative to the past data.
	* Ranges between 0 and 1:
	* **Low α**: Smoother series but slower adaptation to changes.
	* **High α**: Reacts quickly to changes but may capture too much noise.
2. **Gamma (γ)**: The **smoothing parameter for the trend**.
	* Controls how much weight is given to the trend component over time.
	* Ranges between 0 and 1:
	* **Low γ**: A more stable trend estimate.
	* **High γ**: A trend estimate that adapts quickly to changes.
* **Interpret the Stationery R-squared, Mean Absolute Percent Error (MAPE), and Maximum Absolute Percent Error**



**Model Fit:**

1. **Stationary R-squared (0.121)**:
* Measures how well the model explains the variance in a stationary version of the data (adjusted for trend/seasonality).
* A value of 0.121 suggests the model has limited explanatory power for the stationary data.
1. **R-squared (0.999)**:
* Indicates the proportion of variance explained by the model for the original data (not stationary-transformed).
* A near-perfect score (0.999) implies the model fits the original data exceptionally well.
1. **MAPE (0.940)**:
* Mean Absolute Percent Error represents the average percentage error between actual and predicted values.
* At less than 1% (0.940), the model demonstrates excellent accuracy.
1. **MaxAPE (1.866)**:
* Maximum Absolute Percent Error is the largest percentage error in the dataset.
* The maximum error of 1.866% indicates the model performs well across all observations.

**Implications:**

* The model fits the original data exceptionally well (**R-squared = 0.999**) but struggles with explaining stationary variations (**Stationary R-squared = 0.121**).
* The forecast errors are minimal, as evidenced by low **MAPE (0.940%)** and **MaxAPE (1.866%)**.
* High **MaxAE (928.032)** could indicate occasional outlier deviations, possibly requiring further exploration.
1. **Assess the test assumption of no autocorrelation by interpreting the residual PACF plot.**

****

****

**Key Observations from Residual PACF**

1. **Lag 1 (0.064)**:
	* A small positive correlation exists at lag 1, suggesting minimal short-term dependency in residuals.
2. **Lag 2 (0.015)**:
	* Almost no correlation at lag 2, indicating little to no dependence between values two steps apart.
3. **Lags 3 to 6**:
	* Lag 4 (-0.222) and Lag 6 (-0.301) show stronger negative correlations, suggesting potential systematic error patterns that may require further investigation.
	* These spikes might indicate model underfitting or residual autocorrelation.
4. **Lag 7 to 12**:
	* Most lags show weak correlations, but Lag 8 (0.158) and Lag 10 (0.216) are moderately positive.
	* Residual patterns around these lags suggest some short-term relationships remain unexplained.
5. **Lags 13 to 24**:
	* Weak correlations dominate, with values near zero, indicating no significant long-term autocorrelation in residuals.

**Interpreting the Residual PACF**

* **Pattern**:
Ideally, residuals should resemble white noise, meaning most lags should have near-zero PACF values within the confidence interval (approximately ±1.96/√N for a 95% level).
	+ Significant spikes at lags (e.g., Lag 4 and Lag 6) may indicate model inadequacy.
	+ This could signal the need for additional predictors or adjustments (e.g., increasing lag terms in an ARIMA model).
1. **Using a sequence chart, visually display the results of the time series model and explain your findings. Plot the actual values, Predicted Values, and Confidence Intervals around the Predicted Values.**

****

* **Key Observations:**
* **Repeating Growth and Decline**:
	+ - Indicates a cyclical or seasonal pattern in the data.
		- The cycles may not have been explicitly modeled, causing the predicted values to follow the observed pattern too closely.
* **Overlapping Lines**:
	+ Suggests the model fits the data very well (low residuals) but might not account for any uncertainty or future variations (overfitting or lack of flexibility).
	+ Confidence intervals are too narrow, implying the model is overly confident about predictions.
* **Predicted Value, Upper Confidence, Lower Confidence, and Actual Value lines are almost overlapping:**
1. **Very Closed Confidence Intervals**:
	* Visually, the upper and lower confidence interval lines are very close to the predicted value line, it suggests:
		+ The model is highly confident in its predictions.
		+ The variance in the data is small, resulting in narrow confidence intervals.
2. **Lower Confidence Limit (LCL) line** is slightly lower than the other lines (Actual, Predicted, Upper Confidence Limit)
* **Narrow Confidence Intervals**

The confidence intervals (LCL and UCL) are very close to the predicted line.

This suggests that the model is confident in its predictions, with low residual variability.

* **Asymmetric Variability**

If the LCL is systematically lower, the model may expect greater downside risk (e.g., steeper declines) compared to upward trends.

* **Bias in Residuals**

If the residuals are not evenly distributed (e.g., negatively skewed), the LCL may drift lower compared to the UCL.

* **Overfitting or Misspecification**

If the model fits the data too closely (overfitting), it can fail to reflect realistic variability, leading to narrow or biased confidence intervals.

1. **Actual Values Agreed with Predictions**:
	* The actual values are closely following the predicted line, the model is effectively capturing the underlying trends and patterns in the data.
	* Low Residuals: