A graph showing a graph of sales

Description automatically generated with medium confidence

A graph showing a graph of a graph

Description automatically generated with medium confidence

A graph showing a graph of the rate

Description automatically generated with medium confidence

A graph showing a graph of a graph

Description automatically generated with medium confidence

A graph showing the growth of a stock index

Description automatically generated

A graph showing the growth of a stock market

Description automatically generated

A graph showing the price of a stock market

Description automatically generated  
**Summary of Models:**

**1. OLS Model (Multiple Linear Regression)**

**Formula:**

TOTALSA∼PAYEMS+UNRATE+USSTHPI+CUSR0000SETB01\text{TOTALSA} \sim \text{PAYEMS} + \text{UNRATE} + \text{USSTHPI} + \text{CUSR0000SETB01}TOTALSA∼PAYEMS+UNRATE+USSTHPI+CUSR0000SETB01

**Training Set Performance (Model Summary):**

* **Residuals:**
  + Min: -3.4708
  + 1Q: -0.5195
  + Median: 0.0841
  + 3Q: 0.4786
  + Max: 2.9105
* **Coefficients:**
  + (Intercept): 0.1488800
  + PAYEMS: -0.0002509
  + UNRATE: -0.9546211 (significant, p = 0.00402)
  + USSTHPI: -0.0188996
  + CUSR0000SETB01: 0.0058102
* **R-squared:** 0.1983
* **Adjusted R-squared:** 0.1808
* **F-statistic:** 11.32, p-value: 3.145e-08

**2. ARMA Model for UNRATE**

* **ARMA Model:** Fitted ARMA model to the UNRATE variable using auto.arima().
* **Forecast (Next 4 Quarters):**
  1. 0
  2. 0
  3. 0
  4. 0

**3. ARMA Model for PAYEMS**

* **ARMA Model:** Fitted ARMA model to the PAYEMS variable using auto.arima().
* **Forecast (Next 4 Quarters):**
  1. 365.0405
  2. 406.3754
  3. 406.3754
  4. 406.3754

**4. ARMA Model for USSTHPI**

* **ARMA Model:** Fitted ARMA model to the USSTHPI variable using auto.arima().
* **Forecast (Next 4 Quarters):**
  1. 3.347528
  2. 1.007490
  3. 9.890191
  4. 12.486363

**5. ARMA Model for CUSR0000SETB01**

* **ARMA Model:** Fitted ARMA model to the CUSR0000SETB01 variable using auto.arima().
* **Forecast (Next 4 Quarters):**
  1. -0.8407632
  2. 0.0000000
  3. 0.0000000
  4. 0.0000000

**Forecasted Total Vehicle Sales Using OLS Model:**

* -0.01087783
* 0.02785996
* -0.14001934
* -0.18908590

**6. ARMA Model for TOTALSA (Total Vehicle Sales)**

* **Model:** Fitted an ARMA model to the differenced TOTALSA series.
* **Forecast (Next 4 Quarters):**
  1. -0.2332294
  2. 0.0000000
  3. 0.0000000
  4. 0.0000000

**7. Dynamic Regression Model**

**Formula:**

TOTALSA∼TOTALSA.1+PAYEMS+UNRATE+USSTHPI+CUSR0000SETB01\text{TOTALSA} \sim \text{TOTALSA.1} + \text{PAYEMS} + \text{UNRATE} + \text{USSTHPI} + \text{CUSR0000SETB01}TOTALSA∼TOTALSA.1+PAYEMS+UNRATE+USSTHPI+CUSR0000SETB01

**Training Set Performance (Model Summary):**

* **Residuals:**
  + Min: -3.4630
  + 1Q: -0.4133
  + Median: 0.0656
  + 3Q: 0.4688
  + Max: 2.7531
* **Coefficients:**
  + (Intercept): 0.1587167
  + TOTALSA.1: -0.3054775 (significant, p = 3.24e-06)
  + PAYEMS: -0.0002827
  + UNRATE: -1.0684912 (significant, p = 0.000749)
  + USSTHPI: -0.0183416
  + CUSR0000SETB01: 0.0064147
* **R-squared:** 0.2895
* **Adjusted R-squared:** 0.2699
* **F-statistic:** 14.75, p-value: 3.947e-12

**Forecasted Total Vehicle Sales (Dynamic Model):**

* -0.2205235
* -0.1838954
* -0.3468187
* -0.3944367

**RMSE for Test Set:**

1. **OLS Model RMSE:** 16.1483
2. **Reduced-Form Model RMSE:** 16.1289
3. **Dynamic Model RMSE:** 16.3566

This summary compares the OLS, Reduced-Form, and Dynamic models and evaluates their performance using RMSE values on the test set. The OLS model performed slightly better than the dynamic model in terms of RMSE, while the reduced-form model showed comparable results. However, none of the models were particularly strong given the relatively high RMSE values.

**Part B**

**Model Descriptions and Performance Summary**

**Model 1: lm(sales ~ sb + snb + tv)**

* **Test RMSE**: 317.00
* **Overview**: This is a linear regression model with branded search (sb), non-branded search (snb), and TV ads (tv). It performs well, with a relatively low test RMSE. The results suggest that branded and non-branded search and TV ads explain sales variation but more factors could enhance prediction accuracy.

**Model 2: lm(sales ~ sb + snb + tv + factor(dow))**

* **Test RMSE**: 352.61
* **Overview**: Adding the day of the week (dow) slightly increases the RMSE. This suggests that day-of-week variation may not improve the model’s prediction significantly.

**Model 3: lm(sales ~ lag(sales, 1) + sb + snb + tv)**

* **Test RMSE**: 366.81
* **Overview**: Including lagged sales increases RMSE compared to Model 1, implying overfitting. While lagged sales improve training performance, they reduce prediction accuracy on the test set.

**Model 4: lm(sales ~ lag(sales, 1) + sb + snb + tv + factor(dow))**

* **Test RMSE**: 418.62
* **Overview**: This model shows a higher RMSE, which indicates overfitting when both lagged sales and day-of-week seasonality are considered together.

**Model 5: auto.arima(train$sales, xreg = cbind(train$sb, train$snb, train$tv))**

* **Test RMSE**: 339.58
* **Overview**: This ARIMA model with external regressors (branded, non-branded search, and TV) performs better than some linear models, showing that ARIMA can capture time series dynamics more effectively.

**Model 6: dynlm(sales ~ lag(sales, 1))**

* **Test RMSE**: 724.24
* **Overview**: This dynamic linear model with only lagged sales performs poorly due to ignoring important variables like marketing spend.

**Model 7: dynlm(sales ~ sb)**

* **Test RMSE**: 770.52
* **Overview**: Branded search alone doesn’t explain sales variance sufficiently, resulting in poor performance.

**Model 8: dynlm(sales ~ snb)**

* **Test RMSE**: 301.11
* **Overview**: This is one of the best-performing models, highlighting non-branded search as a critical driver of sales.

**Model 9: dynlm(sales ~ tv)**

* **Test RMSE**: 1239.44
* **Overview**: TV ads alone offer little predictive power, with this model having the highest RMSE among all models.

**Model 10: dynlm(sales ~ lag(sales,1) + lag(sb,1) + lag(snb,1) + lag(tv,1) + factor(dow))**

* **Test RMSE**: 597.56
* **Overview**: The added complexity of this dynamic model doesn’t improve performance as expected, with a high RMSE indicating poor generalization.

**Key Variables Analysis**

**Model 1: lm(sales ~ sb + snb + tv)**

* **Non-Branded Paid Search (snb)**: The most important variable, with a strong positive influence on sales (coefficient = 0.0356, p < 2e-16).
* **Branded Paid Search (sb)**: Significant but with a smaller impact than snb (coefficient = 0.0592, p = 0.0364).
* **TV Advertising (tv)**: Not significant (p = 0.235), suggesting limited influence on sales. A screenshot of a computer program

  Description automatically generated

**Model 2: lm(sales ~ sb + snb + tv + factor(dow))**

* **Non-Branded Paid Search (snb)**: Still highly significant (coefficient = 0.0329, p < 2e-16).
* **Branded Paid Search (sb)**: Less significant when considering day-of-week effects (p = 0.0608).
* **TV Advertising (tv)**: Almost significant (p = 0.0617), indicating a potential small effect.
* **Day of the Week (dow)**: Tuesday and Wednesday show significant negative effects on sales. A screenshot of a computer

  Description automatically generated

**Model 3: lm(sales ~ lag(sales, 1) + sb + snb + tv)**

* **Lagged Sales**: Highly significant (p < 0.0001), showing that past sales strongly influence future sales.
* **Non-Branded Paid Search (snb)**: Still significant (p < 2e-16), reinforcing its importance.
* **Branded Paid Search (sb)**: Less significant when past sales are considered (p = 0.0929). A screenshot of a math problem

  Description automatically generated
* **TV Advertising (tv)**: Again not significant (p = 0.795).

**Model 4: lm(sales ~ lag(sales, 1) + sb + snb + tv + factor(dow))**

* **Lagged Sales**: Significant with a strong carryover effect (p < 0.0001).
* **Non-Branded Paid Search (snb)**: Remains significant (p < 2e-16).
* **Branded Paid Search (sb)**: No longer significant (p = 0.358).
* **TV Advertising (tv)**: Still not significant (p = 0.259).
* A screenshot of a computer

  Description automatically generated

**Model 5: auto.arima() with external regressors**

* **Branded Paid Search (sb)**: Significant (coefficient = 0.1722, p < 0.05), showing more influence than in linear models.
* **Non-Branded Paid Search (snb)**: Still important (coefficient = 0.0324, p < 0.05).
* **TV Advertising (tv)**: Not significant (p = 0.2).
* A close-up of a number

  Description automatically generated

Model 6 :

Lag(sales,1) is significant rest not   
A white screen with black text

Description automatically generated

**Conclusion and Comparison**

* **Best Performing Models**:
  + **Model 1**: (branded + non-branded search + TV ads) offers a strong baseline with an RMSE of **317.00**.
  + **Model 8**: (non-branded search only) performs even better, with an RMSE of **301.11**, highlighting the critical role of non-branded search.
* **Worst Performing Models**:
  + **Model 9**: (TV ads only) performs the worst, with an RMSE of **1239.44**, showing TV's limited impact when considered alone.
  + **Model 6**: (lagged sales only) also performs poorly with an RMSE of **724.24**, as it ignores important marketing factors.
* **Complex Models (Models 10)**: These models add complexity but do not perform significantly better, indicating overfitting