

Paper 267-28

Application of the LOESS Procedure for Monitoring and Detecting Critical Movements in the US Automobile Market

Keiko I. Powers, J. Michael Jay, and Tie Gao
J. D. Power and Associates
Westlake Village, CA, USA

ABSTRACT

A nonparametric regression method, LOESS, is a procedure that performs local regression, and it allows us to derive a non-linear fitted line to closely represent a systematic pattern of various dynamic phenomena. The method works particularly well when dealing with rather noisy data, and therefore, the estimation method is ideal in investigating critical changes in the current automobile market in the United States. Historical time series data of auto-related key indicators, such as the sales volume, vehicle price, profit margin, etc often tend to exhibit a volatile pattern, making it difficult to distinguish the actual market trend from 'noise' fluctuations in the data. We set up a SAS data monitoring system with LOESS to analyze time series data of these auto-related key indicator variables. We estimated fitted lines of these indicators using LOESS, which in turn were used to examine various changes in the automobile market with respect to a set of decision criteria. These criteria include such rules as whether or not the current value is outside the historical minimum/maximum range. The LOESS procedure is a very useful tool in executing the monitoring system of the dynamics of the automobile market.

INTRODUCTION

Many users of SAS[®] for statistical analysis are familiar with the REG procedure or GLM procedure for regression-type modeling approaches, but the LOESS procedure is still fairly new to us. According to Cohen (1999), the LOESS procedure was initially introduced with SAS Release 7 as an experimental procedure and became a firmly incorporated product with Release 8. The LOESS procedure performs a nonparametric method for estimating regression surfaces. It allows great flexibility because no assumptions about the parametric form of the regression surface are needed. We can use the LOESS procedure for situations in which we do not know an appropriate parametric form of the regression surface. Furthermore, the procedure is suitable when there are outliers in the data and a robust fitting method is necessary.

This nonparametric procedure turned out to be a perfect analytical tool for assessing marketing dynamics of the automobile industry in the United States. Briefly, we were interested in systematically monitoring various features associated with new vehicle sales, such as vehicle price, gross profit, etc. In particular, our attempt was to compare the current status of, say, vehicle price, to its past historical pattern and quantitatively determine if the current status indicates a substantial deviation from the past behavior. For example, if the monitoring system indicates that the vehicle price of Chevrolet Malibu is higher than the historical maximum, then the LOESS based monitoring system will automatically report such a change. We have developed a SAS-based program to process automobile sales data and then to quantitatively compare the current status to the historical data. As described in detail below, the LOESS approach plays a major role in systematically monitoring the

historical dynamics of many key indicators of automobile sales trends.

BACKGROUND

The US automobile market has been a cutthroat market with many domestic (e.g., GM, Ford, Chrysler) and non-domestic brands (e.g., Toyota, Honda, BMW, Mercedes-Benz), and it is becoming increasingly competitive with more brands entering the market and gaining their market share (e.g., Hyundai). For many marketing managers and analysts in the automobile industry, gathering accurate and relevant information as quickly as possible is becoming a critical issue in staying competitive in the market. At the same time, with today's information-rich environment, summarizing the huge amount of data to an interpretable set of data has become even more critical. For these reasons, we recognized imperative needs to establish a market dynamics monitoring system that will systematically analyze and summarize the real-time automobile data.

The conceptual framework of the current approach was inspired by the data monitoring system developed by Ro et. al.(2000). Ro and his colleagues created a systematic and automatic approach to geometric pattern recognition based on the idea of 'technical analysis', a chart-based pattern recognition approach that has been a part of financial practices for many decades. With this approach, they first decomposed the historical pattern of the stock market performance into several 'chunks' of line shapes. These figures were then examined against a set of various geometric pattern detection criteria. Our current approach also utilizes a systematic and automatic pattern recognition system, but by taking advantage of the LOESS procedure and incorporating a different set of detection criteria.

The essential part of the current monitoring system is to selectively report deviations/changes in the 'current' status of key indicators from their past behavior on a weekly basis. So, for example, when the price of, say, Ford Taurus, has shown a sharp increase, the monitoring system is supposed to detect the change and alert us about the change. A weekly report is then prepared to show a collection of various changes in the current market situation that are detected by the monitoring system. Such real time information highlighting key movements is very critical to marketing managers in efficiently deciding their sales strategies.

DATA AND PROCEDURE

Power Information Network (PIN), a division of J. D. Power and Associates, has been collecting and archiving point-of-sale data from automobile retailers in the United States. The sales data are electronically sent to the PIN computer system daily. The data consist of new and used vehicles purchased by end consumers since 1996 and cover 26 key markets (e.g., New York, Florida, Detroit/Chicago area, Texas, California, etc) with over 6,000 retailers currently providing data to PIN. The total cumulative number of transactions now exceeds 16 million

records. The transaction data can be aggregated and summarized to produce historical data, which allow us to analyze key market indicators to monitor over time the automobile market in the USA. These indicators include various features, such as vehicle price, profit margin, consumer incentives, or days to turn (i.e., how many days it takes to sell).

Due to the volatile nature of the automobile market, the obtained weekly time series data on these key indicators are very hard to interpret because it is difficult to separate important changes from data-related noise (see the Chart 1. in Appendix III showing the sample data).

As you can see, there are many small spikes in the time series data for gross profit. It is hard to determine if these small fluctuations are in fact reflecting changes in the market condition or if they are merely the results of noise in the data. The critical step, therefore, was to come up with an estimation method that allows us to derive a best fitted line in representing the time-related changes. For these purposes, the traditional linear regression approaches were not suitable. Because the main purpose of the study was to compare the 'current' status to the past historical pattern, a linear fitted line did not help us understand the 'local' time-to-time changes in the historical behavior.

On the other hand, the LOESS procedure is an excellent tool in deriving a non-linear 'locally' fitted line (see the Chart 2. in Appendix III showing the fitted line with the original data).

Briefly, PROC LOESS first requires a step to determine the level of the closeness of the fitted line to the data points by iteratively testing different values of 'smoothing parameter'. For example, our SAS program has the following setup:

```
proc loess data=los;
  ods output FitSummary=PinSummary;
  model &var=nweek/smooth =0.04 to 1.0 by
  0.01 dfmethod=exact;
run;
```

The smaller the smoothing parameter is, the more closely the fitted line traces the data points. Once the best fitted line is determined with LOESS, the data value of the 'current' week is compared against the past historical LOESS values based on a set of predetermined criteria. These criteria include whether or not the 'current' values fall outside the maximum-minimum range of the historical data (based on the LOESS estimated time series), or if there is a continuous upward or downward trend for several consecutive weeks (see Appendix I for a complete list of the criteria). When the 'current' value exceeds any of these criteria, it is reported as an 'out-of-normal' value.

RESULTS AND DISCUSSION

For testing our monitoring system, the LOESS procedure was applied to the following key indicators related to vehicle sales performance using data for fullsize pickups.

1. Vehicle price
2. Vehicle gross profit
3. Total gross profit
4. Customer cash rebate
5. Rebate penetration %: % of transactions with customer cash rebate > 0
6. Captive share of finance: % of captive loaner within finance transactions
7. Captive share of lease: % of captive loaner within lease transactions
8. Percent finance: % of finance transactions

9. Percent lease: % of lease transactions
10. APR: lease or finance APR
11. Monthly payment: monthly payment of finance or lease
12. Days to turn: days in dealer lot

The Chart 3 in Appendix III. gives a sample output chart that shows the results of the LOESS-based monitoring system. The left side of the chart (before the vertical reference line) is used to establish the historical pattern of gross profit for this vehicle. The smaller right-hand side of the chart shows the results of the LOESS-based monitoring. The highlighted points indicate weekly time periods that display possible critical 'out-of-normal' movements of gross profit for this full-size pickup. The chart also provides the dates and the reference numbers of the criteria (e.g., F1 or F8, see Appendix I for the detailed descriptions of the criteria) that the current data value exceeded the cutoff point of.

CONCLUSION

This SAS LOESS-based data checking program is still in an early experimental /development stage, and further refinement is needed before it can be fully implemented into our comprehensive automobile market monitoring system. However, the initial results indicate that this data mining technique based on the LOESS procedure will be a valuable tool in systematically examining various key indicators related to the automobile marketing dynamics.

REFERENCES

- Cohen, R. A. (1999), "An Introduction to PROC LOESS for Local Regression," Proceedings of the 24th SAS® Users Group International Conference, Paper 273.
- Lo, A. W., Mamaysky, H., and Wang, J. (2000), "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," *The Journal of Finance*, 55(4), 1705-1765.

APPENDIX I

A set of criteria was derived to decide what pattern is considered a 'critical' change in the indicator status. After investigating many key indicators for various vehicle models and consulting automobile experts, the following set of decision criteria is established. The rule numbers in parentheses are used in Appendix III – Chart 3 for identification purposes.

Rule 1 (F1)

- a) The 'new week' data value is higher than the maximum value of the historical data points.
- b) The 'new week' data value is lower than the minimum value of the historical data points

Rule 2 (F2)

- a) The 'new week' data value is higher than the maximum value of the 95% Confidence Interval (CI) upper bound values of the historical data points.
- b) The 'new week' data value is lower than the minimum value of the 95% Confidence Interval (CI) lower bound values of the historical data points.

Rule 3 (F3)

- a) The 'new week' data value is higher than the maximum value of the 99% Confidence Interval (CI) upper bound values of the historical data points.
- b) The 'new week' data value is lower than the minimum value of the 99% Confidence Interval (CI) lower bound values of the historical data points.

Rule 4 (F4)

a) The 'new week' data value is higher than the maximum value of the 99.99% Confidence Interval (CI) upper bound values of the historical data points.

b) The 'new week' data value is lower than the minimum value of the 99.99% Confidence Interval (CI) lower bound values of the historical data points.

Rule 5 (F5)

a) Focusing on the most recent periods, the past data values consecutively decline for at least 8 weeks. The 'new week' data value is higher than the 95% Confidence Interval (CI) upper bound value of the latest historical data point and is above the historical linear trend line.

b) Focusing on the most recent periods, the past data values consecutively increase for at least 8 weeks. The 'new week' data value is lower than the 95% Confidence Interval (CI) lower bound value of the latest historical data point and is below the historical linear trend line.

Rule 6 (F6)

a) Focusing on the most recent periods, the past data values consecutively decline for at least 8 weeks. The 'new week' data value is higher than the 99% Confidence Interval (CI) upper bound value of the latest historical data point and is above the historical linear trend line.

b) Focusing on the most recent periods, the past data values consecutively increase for at least 8 weeks. The 'new week' data value is lower than the 99% Confidence Interval (CI) lower bound value of the latest historical data point and is below the historical linear trend line.

Rule 7 (F7)

a) Focusing on the most recent periods, the past data values consecutively decline for at least 8 weeks. The 'new week' data value is higher than the 99.99% Confidence Interval (CI) upper bound value of the latest historical data point and is above the historical linear trend line.

b) Focusing on the most recent periods, the past data values consecutively increase for at least 8 weeks. The 'new week' data value is lower than the 99.99% Confidence Interval (CI) lower bound value of the latest historical data point and is below the historical linear trend line.

Rule 8 (F8)

a) The 'new week' data value is higher than the 99.99% CI upper bound value of the most recent historical data point.

b) The 'new week' data value is lower than the 99.99% CI lower bound value of the most recent historical data point.

Rule 9 (F9)

a) The 'new week' data value is higher than the most recent historical data point by over 25%.

b) The 'new week' data value is lower than the most recent historical data point by over 25%.

Rule 10 (F10)

a) The difference between the 'new week' data value and the latest historical data value is higher than the 95% CI upper bound value. Here, the CI was computed based on the difference values between the two adjacent historical data points.

b) The difference between the 'new week' data value and the latest historical data value is lower than the 95% CI lower bound value. Here, the CI was computed based on the difference values between the two adjacent historical data points.

APPENDIX II

The SAS data monitoring system was programmed using a macro routine with PROC LOESS. Here is the LOESS portion of the SAS codes that allows us to derive the fit line to which the decision criteria are applied.

```
/**/ loess  /**/
```

```
%macro modid(mod);
```

```
%do i=1 %to &mod;
  data los; set base;
    if modelid = &i;
```

```
%macro loess(var);
```

```
proc loess data=los;
  ods output FitSummary=PinSummary;
  model &var=nweek/smooth=0.04 to 1.0 by 0.01
  dfmethod=exact;
run;
```

```
ods listing;
```

```
/*---SmoothSelect Macro---*/
```

```
options nonotes;
ods listing close;
```

```
data temp;
  set PinSummary(keep = Label1 nValue1 smoothingParameter
  where=(Label1 in
    ('Number of Observations',
    'Residual Sum of Squares',
    'Trace[L]',
    'Equivalent Number of Parameters',
    'Delta1',
    'Delta2',
    'Lookup Degrees of Freedom')));
```

```
run;
```

```
proc transpose data=temp(drop=Label1) out=temp;
  by smoothingParameter;
run;
```

```
data temp(drop=_NAME_); set temp;
  rename Col1 = n
    Col2 = rss
    Col3 = traceL
    Col4 = delta1
    Col5 = delta2
    Col6 = nu1
    Col7 = lkdf;
```

```
data SmoothCriteria(keep = SmoothingParameter aicc1);
  set temp;
  sigmaHat=rss/n;
  aicc1=n*(log(sigmaHat) + (delta1/delta2)*(n+nu1)/(lkdf-2) );
```

```
proc sort data=SmoothCriteria(where=(aicc1^=.)
  out=Aicc1Results;
  by aicc1;
run;
```

```
ods listing;
```

```
proc sort data=Aicc1Results; by aicc1;
```

```
data _null_;
  set Aicc1Results; if _n_=1 then call symput('bestp',
  SmoothingParameter); else delete;
run;
```

```
proc loess data=los;
```

```

ods output OutputStatistics=&var.best99(rename =
(lowercl=lowercl99
uppercl=uppercl99))
  fitsummary = &var.bestsum;
  model &var=nweek/smooth = &bestp
  clm alpha=0.01;
run;

proc loess data=los;
ods output OutputStatistics=&var.best999(rename =
(lowercl=lowercl999
uppercl=uppercl999))
  fitsummary = &var.bestsum;
  model &var=nweek/smooth = &bestp
  clm alpha=0.001;
run;

proc loess data=los;
ods output OutputStatistics=&var.best95(rename =
(lowercl=lowercl95
uppercl=uppercl95))
  fitsummary = &var.bestsum;
  model &var=nweek/smooth = &bestp
  clm alpha=0.05;
run;

data best&var(keep= nweek &var &var.loessF &var.lowercl95
&var.lowercl99 &var.lowercl999
&var.uppercl95 &var.uppercl99 &var.uppercl999);
  merge &var.best999 &var.best99 &var.best95;
  by nweek;
  rename depvar=&var pred=&var.loessF
  lowercl95=&var.lowercl95 lowercl99=&var.lowercl99
  lowercl999=&var.lowercl999
  uppercl95=&var.uppercl95 uppercl99=&var.uppercl99
  uppercl999=&var.uppercl999;
run;

options ps=100;
options ls=64 nodate nonumber;

%mend;
%loess(md_vehgross);
%loess(md_manureba);
%loess(md_daystotu);
%loess(md_monthlyp);
%loess(md_totalgro);
%loess(md_vehprice);
%loess(mn_apr);
%loess(mn_capfinan);
%loess(mn_caplease);

%loess(mn_perfinan);
%loess(mn_perlease);
%loess(mn_rebatepe);

data loess&i;
  merge bestmd_daystotu
  bestmd_manureba
  bestmd_monthlyp
  bestmd_totalgro
  bestmd_vehgross
  bestmd_vehprice
  bestmn_apr
  bestmn_capfinan
  bestmn_caplease

```

```

bestmn_perfinan
bestmn_perlease
bestmn_rebatepe
;
by nweek;
modelid=&i;
run;

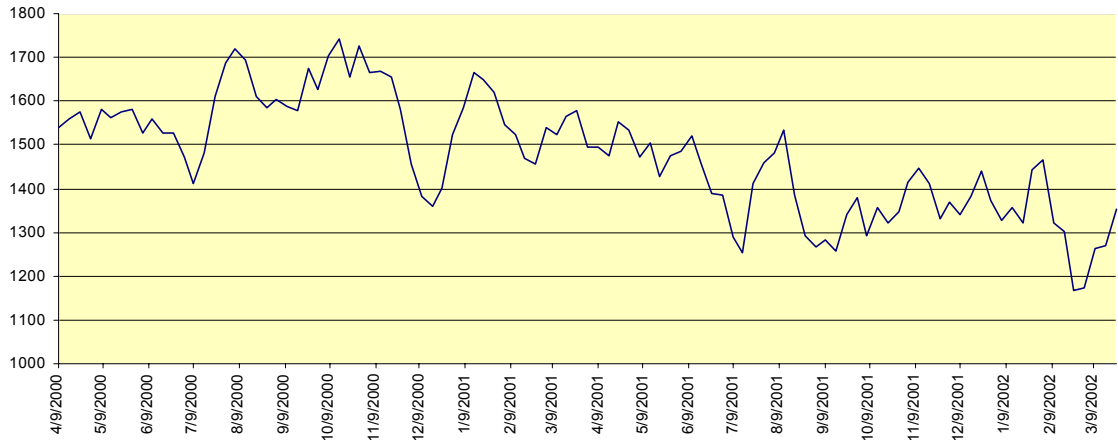
%end;

%mend modid;
%modid(&totmod);

```

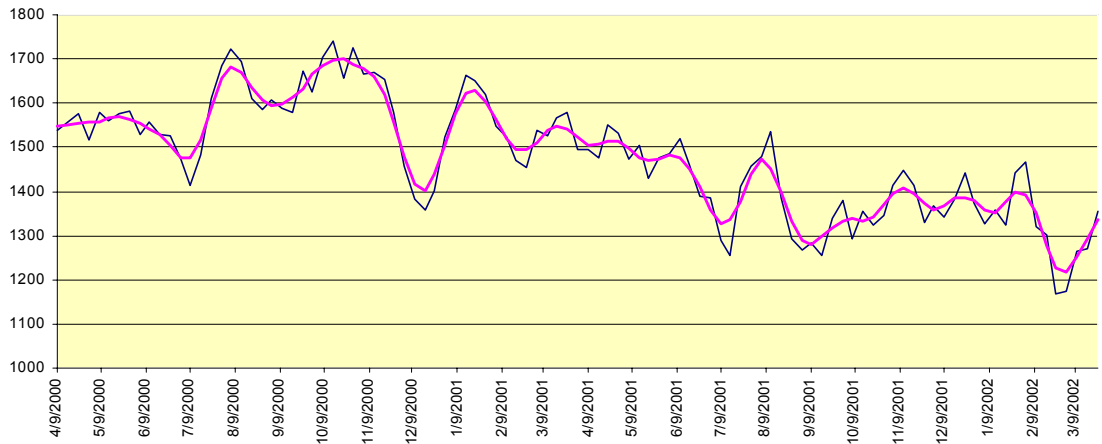
APPENDIX III – CHART 1

Weekly Data of Gross Profit for a Fullsize Pickup



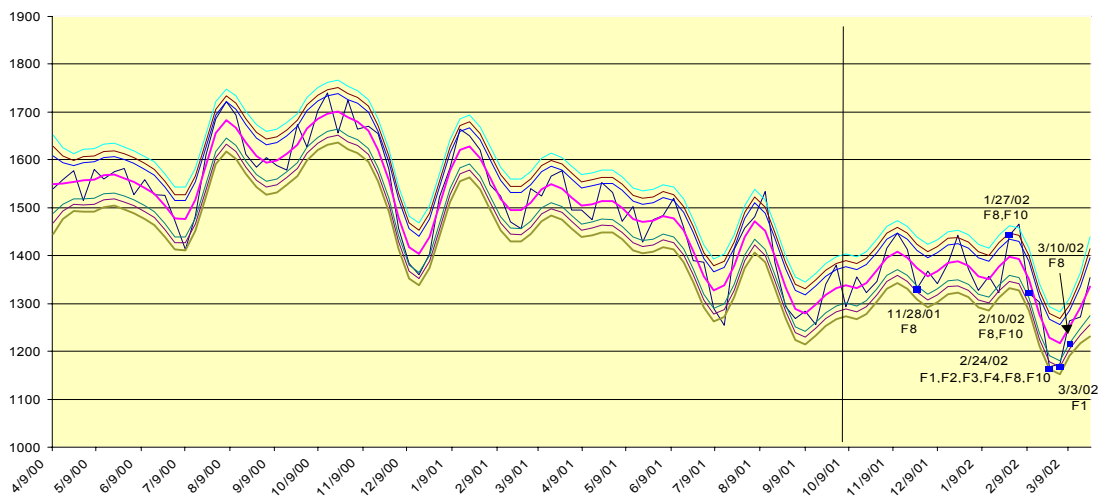
APPENDIX III – CHART 2

Weekly Data of Gross Profit for a Fullsize Pickup with LOESS Fitted Line



APPENDIX III – CHART 3

Gross Profit LOESS-Based Monitoring Results for a Fullsize Pickup



CONTACT INFORMATION

Your comments and questions are valued and encouraged.
Contact the authors at:

Keiko I. Powers, Ph.D.
J. D. Power and Associates
2625 Townsgate Road
Westlake Village, CA 91361
Work Phone: 805-418-8114
Fax: 805-418-8241
E-mail Address: Keiko.Powers@jdpa.com

J. Michael Jay
J. D. Power and Associates
2625 Townsgate Road
Westlake Village, CA 91361
Work Phone: 805-418-8216
Fax: 805-418-8241
E-mail Address: Michael.Jay@jdpa.com

Tie Gao
J. D. Power and Associates
2625 Townsgate Road
Westlake Village, CA 91361
Work Phone: 805-418-8122
Fax: 805-418-8241
E-mail Address: Tie.Gao@jdpa.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.