PREDICTING FINAL RETAIL SALES ESTIMATES FROM ADVANCE REPORTS

Howard Hogan, Patrick J. Cantwell, and M. Cristina Cruz, Bureau of the Census*
Bureau of the Census, Washington, DC 20233

Key words: Prediction, linear models, exploratory data analysis.

1. Introduction

On the ninth working day of each month, the Bureau of the Census releases its first estimates of retail sales for that month as measured by the Advance Monthly Retail Sales Survey (simply, the Advance). The Advance provides timely estimates of retail sales for the entire United States, for all durable and nondurable goods, and for many important detailed kinds of business, such as automotive dealers, building materials, and department stores. The most important estimated quantities are the change in sales from the prior month (expressed as a percent increase or decrease) and, to a lesser extent, the level of sales for the month. The accuracy of these numbers is critical because of their influence on economic policy and the financial markets.

At the time the Advance estimate is made available, the Census Bureau also releases a preliminary estimate for the prior month, and a final estimate for the month before that. The latter two estimates are derived from the Monthly Retail Trade Survey (MRTS, sometimes called the "full monthly survey"). Thus, within 60 days after its release, we revise the Advance estimate with the preliminary estimate, and further revise the preliminary with the final estimate.

The preliminary estimate is considered to be a better estimate than the Advance, and the final better than the preliminary. First, the MRTS sample (about 12,000 units) is much larger than that of the Advance (3400); thus the Advance estimate suffers from greater sampling variance. Second, because the Advance estimate is released so quickly, some sample respondents have not had enough time to complete their books or derive a good estimate of their sales for the month. This increases nonresponse and may influence the quality of the reports we receive. Compared to the MRTS--where sample units are allowed more time to provide their sales for the same data month--it appears that more Advance reporters give an estimate of their sales rather than the preferred book value. If so, this introduces a higher level of *response variance* in the Advance estimates.

In addition, it is possible that early reporting causes a *bias* in the estimates. Recent papers (Bienias, Davie, Hogan, and Konschnik 1996; Davie 1996) have found no strong evidence of a bias due to early reporting *in the Advance*. Yet several papers comparing the reporting periods *in the MRTS* have demonstrated a statistically significant downward bias in the preliminary estimate relative to the final estimate for many kinds of business. (See Waite 1974; Cantwell, Caldwell, Hogan, and Konschnik 1995.) In the latter case, there is no proof that the bias in MRTS reporting is due to the shorter time allowed for the preliminary reports. Further, the reporting schedule for the Advance is different from that for the MRTS. However, weighing the information from various studies, we must allow for the possibility of a bias--perhaps downward--due to early reporting in the Advance survey.

This paper describes efforts to build an explicit model to predict the final estimate based on the Advance survey using less heavily edited data. Because of the time constraints to produce the Advance estimates, the prediction process cannot require as much careful analysis as we would like. The models considered make use of methods such as using auxiliary information and shrinking estimates to reduce variability. We also address how best to handle known biases and how to combine detailed kinds of business into larger aggregates.

In Section 2 we briefly describe the design of the Advance survey and the current method of estimating levels and changes. In Section 3, various models are discussed--what types of models, which variables are included, why they may help predict the final estimate. In Section 4 the method for evaluating the models is presented. Using only data from past months and currently reported sales from the Advance survey, we predict the final estimate (available two months later in real time). The differences between the predicted and the actual final estimates allow us to compare the currently used Advance estimator to the various competing models presented in this paper.

Our initial results in Section 5 indicate that gains are realized in estimating the month-to-month change for some kinds of business. This is also the case when the numbers are aggregated to the U.S. total for retail sales.

2. Estimation in the Advance Survey

Based on responses to the Advance survey, the Census Bureau computes retail sales estimates each month for 16 detailed kinds of business and a number of aggregates. The principal kinds of business for which we publish estimates are building materials, automotive, furniture, other durable goods, general merchandise stores (including department stores), food stores, gasoline service stations, apparel and accessory stores, eating and drinking places, drug and proprietary stores, and other nondurable goods. The most important aggregates are total durables, total nondurables, and total U.S. sales.

The measure of month-to-month change probably sparks the most interest with users of the data. Here, for each of the 16 detailed kinds of business, we use a simple ratio estimator based on sample units that report for *both* months. For any such case i, let $y_{t,i}$ and $y_{t-1,i}$ represent its sales for the current (t) and prior (t-1) months, respectively, and let w_i represent its sample weight in the Advance survey. First, the data are edited for errors in reporting or data transfer and for apparent inconsistencies. Then, for this kind of business, we estimate the change ("advance trend") from month t-1 to month t as

$$AT_{t,k} = 100\left(\frac{\sum_{i \in KB\,k} w_i\, y_{t,i}}{\sum_{i \in KB\,k} w_i\, y_{t-1,i}} - 1\right) \tag{1}$$

To estimate the level of sales, we multiply the ratio of month-to-month change by the preliminary estimate for level from the prior month, $PL_{t-1,k}$, obtained from the larger Monthly Retail Trade Survey (MRTS) for this kind of business:

$$AL_{t,k} = PL_{t-1,k} \times \frac{\sum_{i \in KB\,k} w_i y_{t,i}}{\sum_{i \in KB\,k} w_i y_{t-1,i}} \tag{2}$$

At the time we compute the Advance estimates for the current month (t), the preliminary estimates are available for the prior month (t-1). When estimating the Advance *level* for an aggregate S, $AL_{t,S}$, we first sum the level estimates for its component detailed kinds of business as determined in equation (2). The month-to-month change is then the ratio $AL_{t,S}$ / $PL_{t-1,S}$, where the denominator is the preliminary estimate of level for the same aggregate.

The Census Bureau publishes estimates of level of sales in two forms: (i) unadjusted, and (ii) adjusted for seasonal variations, trading-day differences, and holidays. (We do not try to account for change in the price of goods.) For month-to-month change, the Bureau publishes only adjusted estimates.

3. Modeling the Final Estimate of Sales

3.1 Defining the Problem

The main purpose of the Advance survey is to measure the month-to-month change in retail trade. (Once this is obtained, a level estimate is easily derived.) The final estimate of change, available two months later, gives the best available measure of this number. Therefore, in this sense one can view the Advance estimate as a *forecast* of the final estimate. In this section, we develop several models for fitting the final estimate given data from the Advance survey or other sources. In Section 5 we will evaluate some of these models by trying to predict the final *before it becomes available*, and analyzing the differences between the predictions and the actual final estimates.

A naïve model for forecasting the final estimate of month-to-month change from the Advance would be

$$FT_{t,k} = AT_{t,k} + \text{error}_{t,k},$$

where $FT_{t,k}$ is the final trend (percent change) and $AT_{t,k}$ is the Advance trend--each for month t and detailed kind of business k. More generally, one could write

$$FT_{t,k} = f(AT_{t,k}) + \text{error}_{t,k},$$

where f represents any function. One of the simplest is, of course, the linear:

$$FT_{t,k} = a + b \times AT_{t,k} + \text{error}_{t,k}.$$

Such an approach has long been recognized in the econometrics literature. (See Theil 1971, p. 34.) Note that this approach can be useful even when the forecast is relatively unbiased with respect to the realized, that is, there are negligible errors of *central tendency*. The approach can also correct for errors due to *unequal variation*. Since the Advance estimate is based on a small subsample of the final estimate, we should not be surprised to see the variations of the two differ.

3.2 The Data Set

As mentioned above, each month raw data are collected from the respondents. During this process, Census Bureau clerks and analysts attempt to edit the most glaring errors. For example, if a company reports yearly sales to date rather than monthly sales, a clerk attempts to correct the information. If a group of stores has been acquired (or sold), so that this month's sales do not correspond to last month's sales, a clerk contacts the respondent and tries to correct the error. We refer to the data file at this point as *the analysis file*, and the estimates obtained from it as the Advance analytic estimates.

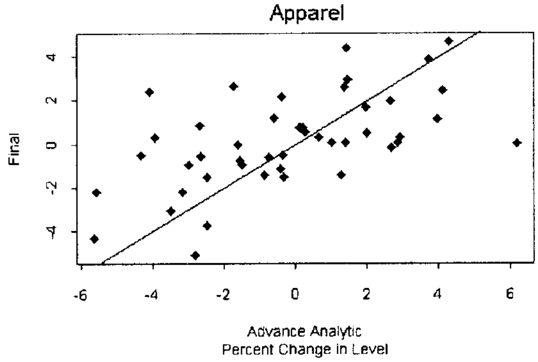
The analysis file no doubt still contains reporting and processing errors. It also may contain correct, but highly unusual observations. For example, a beach town souvenir t-shirt vendor, sampled with high weight, may have sold an unusually large number of shirts during the spring break from schools in March. After weighting, this case might unduly influence the Advance sales estimate for apparel stores or even the U.S. total. Currently, subject matter experts review the analysis file and early estimates. During this process, they often find it necessary to suppress highly influential values either because the case may be in error (but time does not allow verification) or because the case is so unusual that including it would give a misleading measure of the true change in sales in that kind of business.

There are two drawbacks to this process. First, it takes too long, considering the processing schedule for the Advance survey. Second, this process is necessarily subjective. The analysts must make quick decisions with incomplete data and without an explicit framework of measurement and prediction. However, let there be no doubt--the analysts do an excellent job. To see whether an explicit model can do as well, we began by saving the analysis files by kind of business for several months.

3.3 Developing a Model

At this point, it would help to look at some real data. All data considered in the remainder of the paper are seasonally adjusted. For simplicity, let us begin with only one kind of business--apparel and accessory stores (apparel, for short). Apparel stores account for about 5% of total retail trade in the U.S.

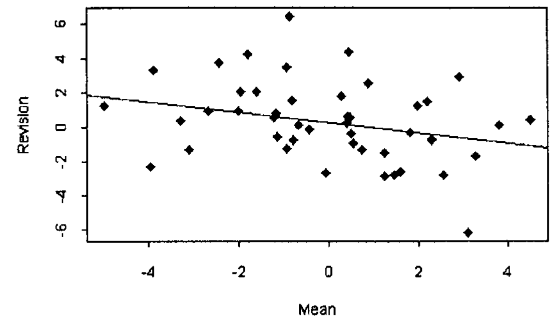
Advance Analytic vs. Final
Apparel



Each data point in the figure represents the estimated percent change in sales in apparel stores for a given month, from December 1992 to August 1996. A 45° reference line has been added. Points on the line are correctly forecast by the Advance analytic estimates. Clearly, for points above the 45° reference line, the Advance analytic estimate underestimates the final; for points below the line, it overestimates.

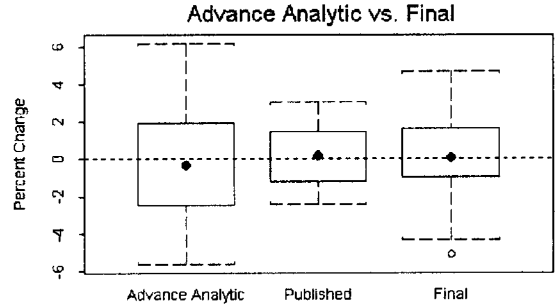
We can get a better look by rotating the plot so that the reference line is horizontal, constructing a mean-difference plot. Now the y variable represents the difference between the final estimate and the Advance analytic estimate ($FT_{t,k}$ - $AT_{t,k}$), while the x variable represents the average of the two estimates. Points on the x-axis now are those showing complete agreement between prediction and reality. A line through the points has been added to depict the tendency. It would seem from the plot that there may be a slight downward trend. That is, when the Advance analytic estimate is high, it seems to overestimate the final; when it is low, it seems to underestimate the final.

Revision vs. Estimated Trend
Apparel



The question of ***unequal variation*** between the Advance analytic estimates and the final estimates must also be addressed. We illustrate this for apparel with box plots. For reference, the plot of the published estimates is included.

Estimated Change
Advance Analytic vs. Final



Although the center points, representing the medians, are quite close, the Advance analytic and final estimators have notably different spreads. This is what we might expect, given that the Advance analytic estimate (i) is based on a small subsample of the final, and (ii) is much less edited and refined. Note that the spread of the published estimates is the smallest of the three. This demonstrates how the analysts' knowledge of the kinds of business and the companies helps keep out erroneous sales reports and unusual observations.

At this point it would seem reasonable to fit a linear regression of the final estimate based on the Advance analytic estimate. The model using least squares on the data from apparel stores is

$$FT_{t,k} \;=\; 0.23 \;+\; 0.44\,AT_{t,k} \;+\; \mathrm{error}_{t,k}\;.$$

In words, the best estimate of the final trend would not be the Advance analytic trend, but roughly half of that rate plus one quarter of a percentage point. One could rewrite the model as

$$FT_{t,k} \;=\; (1 - 0.44)\,(0.41) \;+\; 0.44\,AT_{t,k} \;+\; \text{error}_{t,k}\;,$$

that is, a convex linear combination of the Advance trend and an *a priori* speculation of 0.41% growth. Because the coefficient of $AT_{t,k}$, 0.44, is less than 1, using a prediction model of $0.23 + 0.44\;AT_{t,k}$ decreases the variance of the forecast while "shrinking" it toward the value 0.41%. The intercept term, 0.23, different from 0, (i) may be compensating for inherent bias in the estimate--possibly due to collection, early reporting, or other factors--or (ii) may simply be the most appropriate adjustment (based on the available data) to the Analytic estimate when it is multiplied by 0.44.

Large residuals remain when fitting this model. One idea is not only to use the estimate from the kind of business itself--an estimate typically based on a small number of sample cases--but also to "borrow strength" from all cases in other businesses. Trying this with apparel gives:

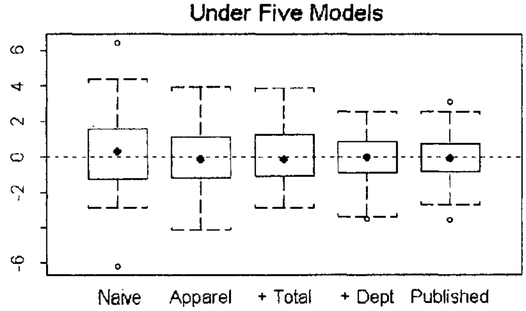
$$FT_{t,k} \;=\; 0.08 \;+\; 0.32\,AT_{t,k} \;+\; 0.70\,AT_{t,(k)} \;+\; \text{error}_{t,k}\;,$$

where $AT_{t,(k)}$ is the trend (percent change) for all Advance analytic sample cases not classified in apparel.

We also examined a linear model for apparel that includes the Advance analytic trend in Department Stores as a regressor. The motivation follows: (i) because a large part of the sales volume in Department Stores is clothing, the sales trends in these two kinds of business are highly correlated, and (ii) the sample reporting in MARTS for Department Stores represents a much larger portion of the total frame volume than does the sample reporting for apparel. In effect, we are allowing for a potential increase in bias in order to stabilize the estimate.

We now have several models under consideration: (1) a "naïve" model (Advance = final), (2) a linear model using only the trend from apparel stores, (3) a linear model that also includes the trend from all other sample cases, and (4) a linear model that includes the trend in Department Stores. We show box plots of the residuals (final - modeled estimate) under these models and for the published number:

Range of Revision
Under Five Models

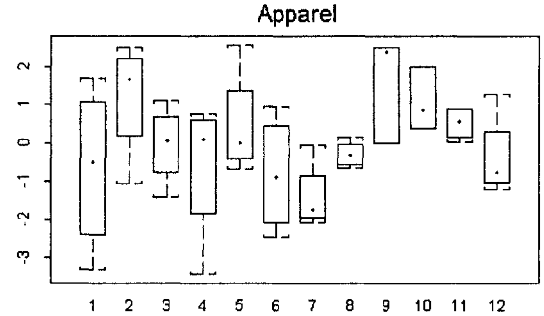


None of these models works as well as we might like. However, it appears that the models that allow the data to determine the coefficients of the input variables (all those but the first model) avoid very large prediction errors. In Section 5 we evaluate the performance of these and other models when "predicting" the final estimate for apparel and auto sales, as well as for the main aggregates.

Space does not permit us to discuss the other kinds of business. However, the results from apparel are somewhat typical. We should mention several models that were also investigated. We tried using the trend for "All other durable goods" as an explanatory variable for the detailed kinds of business under durable goods, and did similarly for the nondurables. This did not work as well as the "all other" category used above. We also tried using the change in the Consumer Price Index (CPI) for a particular (or related) kind of business. The problem here was that the *available* CPI is that from one month earlier. For example, the CPI available when estimating April's retail trade is the March CPI. This indicator--one month late--carried very little predictive power.

We also considered, for specific kinds of business, using different models in different months. Because the dynamics of the market affect a business one way in December, for example, and another way in January, perhaps the appropriate models should use different inputs in the two months. To gain some insight, we show boxplots of the residuals (final - modeled) for apparel *by month*. Using the linear model that includes Department Stores, the residuals are displayed separately for January (1, in the figure), February (2), etc. Note that we have only three or four residuals for each month.

Residuals vs. Month
Apparel



Model with Department Stores

We have only started pursuing this approach. Clearly, more data are needed before inferences are made and alternatives eliminated

4. Evaluating the Predicted Final Estimates

Our goal is to release an Advance estimate of month-to-month change that "predicts" the final estimate in retail sales in each of the various detailed kinds of business and the major aggregates. When fitting models to data from the same period of time, using additional explanatory variables can only induce a fit that is better--even if only marginally so. But analyzing the fit *on the same data* used to select the model does not indicate how well the model will work on new data, or how much a specific explanatory variable really will contribute. Now we wish to evaluate how well the models predict in "real time." That is, we predict the final estimates based on data (mainly from the analysis files) from the past m months *that would have been available at the time*. Then we examine the residuals between these predictions and the actual final estimates *computed later*. How successfully the model works can be measured by comparing the predicted value to the final estimate.

In predicting the final month-to-month percent change, $FT_{t,k} = FL_{t,k} / FL_{t-1,k}$, we have access to Advance estimates up through month t, preliminary estimates up through month t-1, and final estimates up through month t-2. For the various kinds of business, we used a set of m consecutive months (typically, 24) to model the relation between the final and the Advance analytic estimates (and any other data), and then predicted $FT_{t,k}$, where t was 2 months beyond the end of the set used for modeling. For example, we started by determining model parameters based on data from the months December 1992 (12/92) through 11/94. We then predicted the final estimate for 01/95 by inserting the Advance estimate for 01/95 (and the preliminary estimate for 12/94, if used in the model) into the model. The residual--that is, the difference between the predicted and published final for 01/95--is used to evaluate the success of the model.

Next we predicted $FT_{t,k}$ for 02/95 based on data from 01/93 through 12/94, then $FT_{t,k}$ for 03/95 based on data from 02/93 through 01/95, etc. This "moving window" approach for determining model parameters allows us (i) to test the model with the data that would actually be available, and (ii) to use the most recently available input data while ignoring data more that m+2 months old. Our selection of 24 for m is a compromise. As m increases, the model parameters are estimated with greater stability. Yet, because conditions in the economy and in the specific kind of business change over time, a smaller value of m ensures added emphasis on more recent input data. In fact, because we incorporate a new Advance sample every two to three years, it is important to keep the moving window short enough to quickly pick up potentially different aspects of the relation between the final and the Advance estimates.

For aggregates such as U.S. total, we *could* have directly modeled the estimates of month-to-month percent change. Instead, we modeled and predicted the final change estimates *only* for the detailed kinds of business. We then converted these to estimates of levels, and summed over the components to obtain aggregate estimates of level. Finally, these were converted to estimates of month-to-month percent change. We felt that modeling at the aggregate level and raking down to the detailed components would not take advantage of the dynamics that occur for specific detailed kinds of business. Further, we hoped that improving the prediction at lower levels would translate into good prediction for the aggregates.

5. Summary and Analysis of Our Results

We began analyzing the various models by predicting the final trend estimates over the 27 months, January 1995 through March 1997, and computing the residuals between the predicted trends and the actual final trends. We then found the average over the 27 months of (i) the prediction residuals, (ii) the absolute values of the residuals, and (iii) the squares of the residuals.

For apparel stores, auto sales, and many other kinds of business, we examined these summary statistics for the Advance analytic estimates and for estimates from many models under consideration. These models were all linear in form--applied with and without resistant regression. In fact, with our data there appears to be little difference in the mean absolute or squared residual between the usual least squares regression and resistant regression. But we wanted to distinguish between (i) obtaining a good fitting model on a set of data, where outliers or unusual values *are gainfully included* to evaluate the fit, and (ii) predicting future results, where the data used to determine the model parameters do not enter into the prediction evaluation. As our goal coincides with the latter, we decided to limit the remaining analysis to models using resistant regression.

We considered models that include terms such as (i) the Advance analytic trend over all kinds of business *except* the business we are predicting, (ii) the preliminary trend, $PT_{t-1,k}$, from the previous month, (iii) (for estimating apparel only) the trend in Department Stores, and (iv) (for estimating auto sales only) the month-to-month percent change in the number of auto units sold.

A review of the residuals implies that most of these models improve on the performance of the Advance analytic estimates. Generally, there is no strong preference among these models. One exception is the model for autos where including the trend in auto units lowered the mean absolute and squared residuals considerably.

Finally, we compared the performance of the Advance analytic estimates, a simple linear model (using resistant regression), and the published estimates for five kinds of business or aggregates: apparel stores, auto sales, total durable goods, total nondurable goods, and the U.S. total in retail sales. See Table 1. Note that when aggregating component sales to totals, we used the simple linear model in all kinds of business except auto sales, where we used the model that also includes the trend in auto units.

One can see that in some kinds of business, such as apparel, the published estimate has fared better over this period (January 1995 to March 1997) than the models would have. Yet in others, such as auto sales, a properly selected model can improve on the published estimate. When aggregating to the totals, the strong performance of the models in the durable goods businesses helps the modeled estimates beat that of the published estimates, at least over this period. Further analyses are required to determine how well these and other models work when the economy experiences shifts in other directions.

References

BIENIAS, J.L., DAVIE, W., Jr., HOGAN, H., and KONSCHNIK, C.A. (1996). "An Analysis of the Advance Monthly Retail Sales Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 693-698.

CANTWELL, P.J., CALDWELL, C.V., HOGAN, H., and KONSCHNIK, C.A. (1995). "Examining the Revisions in Monthly Trade Surveys Under a Rotating Panel Design," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 567-572.

DAVIE, W., Jr. (1996). "Measuring the Effect of Early Reporting on Response Errors in an Economic Survey, " *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 699-704.

THEIL, H. (1971). *Applied Economic Forecasting*, North-Holland Publishing Co., Amsterdam.

U.S. Bureau of the Census (1997). *Current Business Reports, Monthly Retail Trade: Sales and Inventories*, March 1997, Publication no. BR/97-03, Washington, D.C.: Government Printing Office.

WAITE, P.J. (1974). "An Evaluation of Nonsampling Errors in the Monthly Retail Trade Sales Data," *Proceedings of the Section on Business and Economic Statistics, American Statistical Association*, 602-607.

* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau. We thank Julia Bienias and Tim Braam for constructing files for this research.

Table 1. Comparing Advance Analytic, Simple Linear, and Published Estimates

Final - Predicted 1/95 - 3/97	Average Residual	Average Absolute Residual	Average Squared Residual
APPAREL			
Adv. anal.	.65	1.79	4.77
Linear[1]	.18	1.34	2.90
Linear w/ Dept. stores[2]	.08	1.35	2.70
Published	.16	1.16	2.29
AUTOS			
Adv. anal.	.71	1.09	1.91
Linear[1]	-.37	.97	1.37
Linear w/ auto units[3]	-.35	.77	.93
Published	.52	.90	1.20
TOTAL DURABLES			
Adv. anal.	.46	1.15	2.21
Linear	-.26	.56	.48
Published	.27	.74	.97
TOTAL NONDURABLES			
Adv. anal.	.14	.55	.56
Linear	-.08	.33	.18
Published	.10	.33	.24
U.S. TOTAL			
Adv. anal.	.26	.64	.67
Linear	-.15	.35	.19
Published	.17	.40	.28

[1] $FT_{t,k} = b_0 + b_1 \times AT_{t,k} + \epsilon_{t,k}$

[2] $FT_{t,\mathrm{app}} = b_0 + b_1 \times AT_{t,\mathrm{app}} + b_2 \times AT_{t,\mathrm{dept.\ stores}} + \epsilon_{t,\mathrm{app}}$

[3] $FT_{t,\mathrm{autos}} = b_0 + b_1 \times AT_{t,\ \mathrm{autos}} + b_2 \times AT_{t,\mathrm{auto\ units}} + \epsilon_{t,\mathrm{autos}}$