AN ALGORITHM FOR ARIMA DYNAMIC REGRESSION VARIABLE SELECTION

Ross Bettinger, Analytical Consultant, Seattle, WA

ABSTRACT

It is frequently the case with Big Data that you have a gazillion exogenous predictor time series that you need to analyze so you can build a dynamic regression time series model.[1] The %ARIMA_SELECT macro can help you to determine which exogenous predictors[2] are strongly correlated with your response variable and which are not. The %ARIMA_SELECT macro is designed to rank-order individual exogenous time series according to a user-specified performance measure, and can be a valuable tool for reducing the dimensionality of your data. It can also perform variable selection using forward selection or "all subsets" selection for model building.

KEYWORDS

ARIMA, forward selection, backward elimination, all subsets selection, time series, methodology, model building, dynamic regression, exogenous predictor, big data, dimension reduction

INTRODUCTION

An event of interest occurs and is assumed to be related to other variables that are observed or measured in the same time frame as the event's occurrence. Modelers hypothesize that the response variable is related to one or more predictor variables and that future values of the response variable can be accurately forecast by appropriate mathematical formulations of these predictors.

Variable selection is a critical component of model building for several reasons. Too few variables in a model may result in underperformance in the sense that not all factors relevant to an outcome are included, so that important relationships between the response variable and predictors are not brought into the model. This produces omitted variable bias, and the misspecification error produces erroneous predictions. Too many variables in a model may introduce spurious correlations between exogenous variables and thus confound the model's forecasts. Also, too many predictors may consume excessive CPU cycles when time is a critical constraint.[3] We may apply the principle of Ockham's razor: "*Numquam ponenda est pluralitas sine necessitate* [Plurality must never be posited without necessity]"[4] and choose the fewest number of predictor variables that adequately account for the most accurate predictions. There is a measure of subjectivity in the process that may be addressed by business logic in addition to statistically-informed exploration.

DISCUSSION OF VARIABLE SELECTION METHODOLOGY

Common variable selection techniques include forward selection, backward elimination, stepwise selection, and variations on selection of subsets of variables.

- In forward selection, no candidate predictors are initially included in a model. Instead, a model is built for each candidate predictor, and the single variable that explains the most variance (as determined by its F statistic) is added to the model if its significance level (p-value) is less than an inclusion threshold value[5] (e.g., p < SLENTRY). Once a variable has been included, it remains in the model.

---

[1] Dynamic regression models are time series models that include exogenous predictor variables into a standard time series model. These external variables allow the modeler to include dynamic effects of causal factors such as advertising campaigns, weather-related natural disasters, economic events, and other time-dependent relationships into a time series framework.

[2] The terminology used in time series modeling differs from regression modeling. The time series response variable is the dependent variable in the regression context. Similarly, independent time series variables are called dynamic regressors, exogenous variables, or predictors.

[3] For example, credit card authorizations must be approved or denied according to service level agreements (SLA) of a few seconds after a card has been swiped through a card reader, including transit time to and from an authorization server. A fraud detection model may be very accurate in detecting fraud, but if it does not produce a score within the SLA constraints, it is not useful.

[4] http://en.wikiquote.org/wiki/William_of_Occam

[5] The threshold value states the risk that a modeler is willing to accept that a specified percentage of true hypotheses will be falsely rejected. For example, the PROC REG default significance level for entry of a variable into a model (SLENTRY) is 0.50, and to stay in a model (SLSTAY) it is 0.10. Threshold values are specified as proportions, e.g., .05, .01, .001, &c., indicating that the acceptable rate of rejection of H_0 is 1 in 20, 1 in 100, 1 in 1000, &c. Smaller p-values represent increasing risk aversion.

After the first predictor has been selected, it is included in the model and removed from the list of candidates. Candidate predictors must satisfy the same conditions previously applied, and the process is repeated until either there are no more candidates or the remaining candidates do not satisfy the inclusion criterion.

The choice of p-value is somewhat arbitrary, based on the assessment of the risk of falsely rejecting a variable that ought to have been included in a model (Type I error is the incorrect rejection of a true null hypothesis, a "false positive" result).

- In backward elimination, all variables are initially included in a model. Variables that do not at least minimally contribute to explaining variance (as determined by their F statistics) are removed, one by one, until all remaining variables produce F statistics whose significance level is inferior to a retention threshold value (e.g., p < SLSTAY). Once a variable has been removed from the model, it is removed from the set of candidate predictors.

- Stepwise selection is a hybrid of forward selection and backward elimination algorithms. Variables are included in a model individually as in forward selection, but once included, are in competition with new candidates for inclusion. A variable that has been included at an earlier step may be removed at a later step if the candidate variable, in concert with all other variables already included in a model, explains more variance than the previously-included variable. There are significance levels for entry into the model and for staying in the model.

The salient point to note for all of these techniques is that the model described by the response variable and the set of independent predictors changes with each iteration of the algorithm. The number of variables changes and hence the degrees of freedom change. Variables chosen in one step are subject to different significance levels in a subsequent step. This nonconstant significance level may result in a false positive result (Type I error) so that a variable that ought to have been included in a model is rejected. The Bonferroni correction is one attempt to resolve this problem.[6] Further details on variable selection methodologies are available in [1].

Subset selection techniques require that models of successively-larger subsets of variables be built using variables that satisfy some selection criterion, e.g., maximum R^2. For example, a subset selection algorithm using the maximum R^2 criterion would sift through the set of candidate predictors and find the "best" one-variable model for which the variable produced the greatest R^2. So, for p predictors, the algorithm would create p models. Then, it would find the "best" two-variable, creating all combinations of two variables. This would require C_2^p models. If all p predictors were used, the final model would require p predictors, and a total of $2^p - 1$ models would be built. While the false positive rate within a subset is held constant, the computational burden quickly becomes excessive for all but small subsets. The curse of dimensionality precludes using this technique for large p.

The %ARIMA_SELECT macro does not use significance levels in selecting variables because in practice (as opposed to theory), the decision to include or exclude a variable is frequently only marginally related to its p-value. Corporate "best practices" may dictate which variables shall be used to build a model. For example, a modeling group's practice may be to use no more than 15 variables in a model. Or, specified variables from subsets of predictors must be present in a model. There may be rules based on business logic or domain knowledge that constrain the choice of predictors, even at the expense of model performance. While it is important that predictors be strongly correlated with the response variable, the significance level is not the sole criterion for selection, so it is not emphasized in the %ARIMA_SELECT algorithm.

USING THE %ARIMA_SELECT MACRO FOR DYNAMIC REGRESSION VARIABLE SELECTION[7]

The PROC ARIMA procedure does not have a variable selection capability. It can build only one model at a time and uses all of the time series predictors that the analyst has specified. Model building using PROC ARIMA can be manually intensive when there are many candidate time series to process. Unless a painstakingly rigorous time-intensive regimen is followed, the variables selected may not produce a "best" model.

The %ARIMA_SELECT macro uses a "wrapper" approach[8] to variable selection. PROC ARIMA is exercised over the set of predictors in systematic ways to 1) reduce the number of candidate time series and 2) build models using the

---

[6] See, e.g., http://en.wikipedia.org/wiki/Bonferroni_correction

[7] Parameters to the %ARIMA_SELECT macro are listed in the Appendix.

[8] A "wrapper" is a program that iteratively invokes other programs to perform specified tasks to achieve a goal. In this case, the %ARIMA_SELECT macro iteratively invokes PROC ARIMA and other SAS programs to perform variable selection.

reduced subset of selected predictors. Table 1 represents the selection methods available via %ARIMA_SELECT.

Algorithm	Description
Baseline	Builds models for individual candidate predictors, computes R^2 for baseline predictors, selects and rank-orders a candidate predictor if $R^2(baseline\ vars + predictor) > R^2(baseline\ vars)$
Forward Selection	Builds a model for each candidate predictor, selects best performer (which is then added to the model), iterates on remaining candidate predictors
Subset Selection	Builds subsets of all combinations of specified number of predictors
Baseline + Forward Selection	Applies forward selection to predictors chosen by baseline algorithm
Baseline + Subset Selection	Builds subsets of all combinations of predictors chosen by baseline algorithm

Table 1: %ARIMA_SELECT Variable Selection Options

BASELINE ALGORITHM

The baseline algorithm builds an ARIMA model using user-supplied parameters (e.g., orders of autoregressive (p), differencing (d), moving average (q) terms) to specify a model and optionally includes in the model any predictors that are specified as starting variables. The R^2 statistic is produced from the residuals and the total sum of squares. A predictor is selected if, for some ΔR^2, the predictor $R^2 >$ baseline $R^2 + \Delta R^2$. Thus, any predictor chosen will improve the performance of the model in which the predictor is included.

The baseline algorithm is applied to all time series in a modeling dataset or to a specified set of time series variables that have been specified by the analyst.

FORWARD SELECTION ALGORITHM

The forward selection algorithm uses the subset of time series variables specified by the analyst or all of the time series variables in a modeling dataset.

It builds *p* models from the *p* predictors in the initial set of candidates and chooses the predictor with the highest performance measure as specified by the analyst. The measure may be R^2, adjusted R^2, mean absolute percent error (MAPE), root mean squared error (RMSE), or Schwarz' Bayesian criterion (SBC). The chosen variable is then removed from the set of candidates, and the process is repeated on the next *p-1* predictors. The process is continued until the list of candidates is empty or the maximum number of predictors, as specified by the analyst, is reached. At most, $p(p + 1)/2$ models are built.

SUBSET SELECTION ALGORITHM

The subset selection algorithm uses the same initial set of variables as the forward selection criterion. It builds models with successively more variables, choosing them according to the same criterion as the forward selection algorithm.

The analyst must specify how many variables are in the first model and optionally in the last model. If the START parameter is 1, then the best one-variable model is built (*p* models), the best two-variable model (C_2^p models), the best three-variable model (C_3^p models),…, up to the STOP parameter value. If the STOP parameter is omitted, all of the variables will be used to build subsets. There may be at most $2^p - 1$ models built.[9]

BASELINE + FORWARD SELECTION ALGORITHM

The baseline + forward selection algorithm applies the baseline algorithm to the set of candidate predictors. The forward selection algorithm is then applied to the set of baseline-selected candidate predictors. This is an efficient use of the %ARIMA_SELECT macro because the initial predictors have been selected based on their ability to exceed the baseline $R^2 + \Delta R^2$, and hence are likely to be relatively few in number.

Candidate predictors are selected up to the maximum number indicated by the analyst (TOPN parameter value).

---

[9] One must be very judicious in using this option. There 31,557,600 seconds/year. If each model required only one second to run, compute performance statistics, and record results, then the maximum number of models that could be processed by the subsets option in one year would be log2(31,557,600) = 24.9 models (to one significant digit). We strongly recommend that the START/STOP parameters be used, and that the subsets option be used sparingly, unless you have a computer on the order of Asimov's Multivac. See "All The Troubles of the World" in [2].

BASELINE + SUBSET ALGORITHM

The baseline + subset selection algorithm uses the predictors selected by the baseline algorithm as the starting set for the subset selection algorithm. This usage strategy of the %ARIMA_SELECT macro is to be preferred to the subset algorithm *per se* because the initial set of candidate predictors is highly likely to be much more feasible with respect to computation time than for simply using all available time series variables.

Predictors in each subset are ordered according to the performance measure selected by the analyst.

CONSIDERATIONS

If the baseline algorithm was used (&SELECT= [BASE | BASE_FWD | BASE_SUB]), the incremental improvement of a predictor's R^2 over the baseline $R^2 + \Delta R^2$ and standard regression statistics (parameter estimate, t-statistic, two-sided Pr > |t|) are reported.

If the baseline algorithm was used and the &OUTFILE parameter contains the name of a file, the names of the variables selected are written to the specified file. Additional statistics (R^2, incremental improvement of a predictor's R^2 over the baseline $R^2 + \Delta R^2$, two-sided Pr > |t|) are also written to the file. The variables and measures can be used to select variables individually for custom-built models.

User-specified prewhitening of dynamic regressors (predictors) is not implemented.[10] In the interests of automating the variable selection process, only the unfiltered exogenous variable will be considered. While this limitation may cause a predictor to be rejected as a candidate for model building, it is a tradeoff in designing an efficient automation tool.

The largest number of predictors that can be processed for subset selection is 1023.[11]

EXAMPLE OF USE

The following example illustrates how to use the %ARIMA_SELECT macro to perform initial variable selection prior to more thorough investigation. The %ARIMA_SELECT macro is applied to a demonstration adapted from course notes for "Forecasting Using SAS® Software: A Programming Approach" [3].[12]

SCENARIO

An internet-based startup provides financial and brokerage services to customers and advertises its services in several different venues. Management wants to know how effective its advertising spend has been in generating profit for the firm. What is the advertising ROI? Data have been compiled for the purpose of determining the effectiveness of several marketing channels: direct mail, internet ads, print media, and TV and radio advertising. The weekly data span the 88 weeks from 28 September 1997 to 30 May 1999. Analysis of the data to answer Management's question is beyond the scope of this paper.

There are only four time series variables that represent individual media channels, and there is no information provided on frequency of sales campaigns or duration of campaigns per channel. So, we used the simple decay-effect Adstock transformation [4] to expand the number of variables per channel according to "half-life" of advertising effect upon level of sales. For example, a half-life of one week means that the effectiveness of an advertising message to promote sales diminishes 50% from its initial presentation after one week. We created five variables per channel. Direct mail, for instance, created variables DirectMail_Adstock1, DirectMail_Adstock2, DirectMail_Adstock3, DirectMail_Adstock4, and DirectMail_Adstock5, corresponding to campaign lengths of one to five weeks, respectively.

ADSTOCK TRANSFORMATION

The mathematical formulation of the simple decay-effect Adstock model, equation 2 in [4], is

$$A_t = T_t + \lambda A_{t-1}, \text{ t = 1, ..., n}$$

where A_t is the Adstock at time t, T_t is the value of the advertising variable, e.g., $1000's of media coverage bought, and λ is the "decay" or lag weight parameter.[13] Equation 10 in [4] describes the relationship between half-life and λ. It

---

[10] Prewhitening a dynamic regressor is performed by defining a filter for the regressor in the identification and estimation phases of ARIMA modeling. Applying this filter to the regressor produces a white noise residual series that is cross-correlated with the white noise residual series of a similarly-filtered response variable.

[11] The value 2^{1023} produces an integer that is pretty close to a gazillion, IMHO. Surely, you have better things to do than wait while the Universe succumbs to entropic heat death before your modeling run completes.

[12] Chapter 5.3, Demonstration: Evaluating Advertiser Effectiveness

[13] The informed reader will recognize equation [2] as a first-order linear difference equation with constant coefficients.

is

$$\lambda^n = \frac{1}{2}$$

so that $\lambda = e^{\log(.5)/n}$. Then, by specifying the half-life of an advertising campaign to be n weeks and computing the resulting λ, we can substitute λ into equation 2 in [4] and model the effects of the campaign on sales.
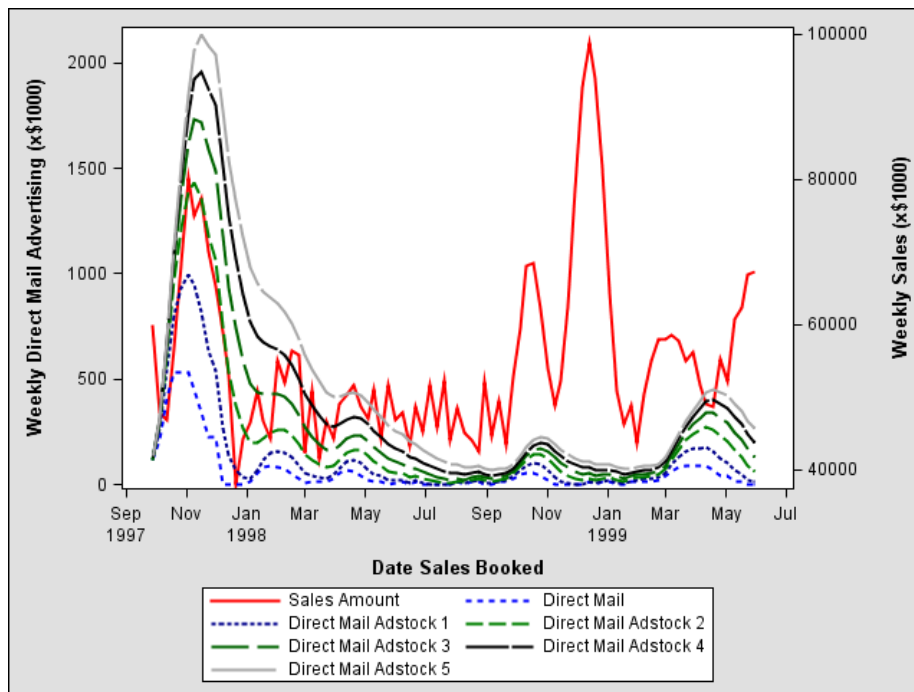
APPLYING THE ADSTOCK TRANSFORMATION

We computed λ for five campaigns to compute half-lives of 1, 2, 3, 4, and 5 weeks. The values of λ are given in Table 2.

Half-Life (Weeks)	Decay Parameter, λ
1	0.5000
2	0.7071
3	0.7937
4	0.8409
5	0.8706

Table 2: Half-Life Decay Parameter

We applied equation 2 to each of the advertising time series and created five additional variables per time series, one for each of the campaign half-lives in Table 2. Figure 1 shows the effect of applying the Adstock transformation to the DirectMail time series.

Figure 1: Sales Amount vs Direct Mail Advertising Spend

The Adstock transformations have varying effects according to the half-life of each campaign. For direct mail, the original correlation over time between SalesAmount and DirectMail is r=0.28 (p<.008). Applying the Adstock transformation with a half-life of one week increases the correlation to r=0.33 (p<.0017), which is a beneficial improvement. We see that the one-week campaign is most closely related to the Sales Amount, and that increasing campaign length serves to diminish the effect of advertising spend on direct mail. We also note that the longer the campaign length, the less closely-correlated is the Adstock-transformed direct mail time series to the original direct mail time series.

Pearson Correlation Coefficients, N = 88 Prob > |r| under H0: Rho=0		
	SalesAmount	**DirectMail**
DirectMail	0.28113	1.00000
Weekly Direct Mail Advertising (x$1000)	0.008	
DirectMail_Adstock1	0.32942	0.96152
Campaign Half-Life = 1 Week	0.0017	<.0001
DirectMail_Adstock2	0.3193	0.86582
Campaign Half-Life = 2 Weeks	0.0024	<.0001
DirectMail_Adstock3	0.27833	0.77596
Campaign Half-Life = 3 Weeks	0.0086	<.0001
DirectMail_Adstock4	0.2304	0.70043
Campaign Half-Life = 4 Weeks	0.0308	<.0001
DirectMail_Adstock5	0.18297	0.63680
Campaign Half-Life = 5 Weeks	0.088	<.0001

Table 3: Effect of Direct Mail Adstock Transformations

MODEL BUILDING USING BASELINE + FORWARD SELECTION

We can use the %ARIMA_SELECT macro to determine the marginal effect of Adstock transformations in modeling the relationship between SalesAmount and each media channel. Applying the Adstock transformation to each original media channel adds five variables to each of the candidate predictors. We will examine the effect of the transformation for each channel, exclusive of the others.

For direct mail, we can run the following code:

```
%let DM = DirectMail_Adstock1 DirectMail_Adstock2 DirectMail_Adstock3
          DirectMail_Adstock4 DirectMail_Adstock5 ;

%ARIMA_SELECT( SASDATA.Saledata_Adstock, TIMEID=DATE
             , DEPVAR=SalesAmount, INDVAR=&DM, STARTVAR=DirectMail
             , DELTA_RSQ=.001
             , NOTES=n, PRINT=n
             , MEASURE=adj_rsq, SELECT=BASE_FWD, START=1, STOP=, TOPN=10
             , D=, NLAG=24, I_OTHER_OPT=
             , MAXITER=50, METHOD=ml, P=5, Q=, E_OTHER_OPT=
             , INTERVAL=week
             , F_OTHER_OPT=%str( align=BEGINNING )
             , OUTFILE=
             , ODSFILE=Path\Filename
             )
```

Since the selection parameter, &SELECT, indicates baseline + forward selection, the %ARIMA_SELECT macro first applies the baseline algorithm to produce a set of baseline candidate predictors, and then applies the forward selection algorithm to the set of baseline candidate predictors.

The baseline algorithm produces the following results:

Baseline Candidate Predictors						
Baseline R^2 = 0.081388						
Candidate Predictor	**Candidate R^2**	**Improvement Over Baseline R^2**	**Param Estimate**	**Std Error**	**t Value**	**Prob > |t|**
DirectMail_Adstock1	0.1253	0.0440	42.8780	20.2270	2.1200	0.0340
DirectMail_Adstock2	0.1020	0.0207	11.0960	7.5660	1.4670	0.1425
DirectMail_Adstock3	0.0882	0.0069	4.2790	4.8200	0.8880	0.3746
DirectMail_Adstock4	0.0818	0.0004	1.3890	3.6630	0.3790	0.7045
DirectMail_Adstock5	0.0805	-0.0009	-0.2250	3.0250	-0.0740	0.9407

Table 4: Baseline Candidate Predictors

We see that increasing campaign length has a diminishing effect on the relationship between SalesAmount and DirectMail. These results are consistent with those of Table 3. The R^2 statistic is used to determine which candidates add a positive marginal contribution to the baseline model. Each candidate variable R^2 is compared to the baseline $R^2 + \Delta R^2$ and if it produces an improvement, it is selected into the baseline candidate set of predictors.

The forward selection algorithm uses the set of individually-selected baseline predictors and selects predictors based on adjusted R^2. The first variable selected, DirectMail_Adstock1, has the highest adjusted R^2 of the four candidate predictors, so it is selected first. Then, given that the model includes DirectMail_Adstock1, the next predictor selected is DirectMail_Adstock4 since it has the highest adjusted R^2 of the remaining three candidate predictors. The algorithm continues selecting predictors in order of highest adjusted R^2 until the list is exhausted.

The results are summarized in terms of several performance measures. Note that, since the DirectMail_Adstock5 time series did not make a positive contribution above the baseline $R^2 + \Delta R^2$, it was not included in later processing. Variables that do not exceed the threshold criterion are culled from further analysis so that the number of predictors is minimized to include only those that add explanatory power to the model.

Summary of Forward Selection									
Performance Measure = Adjusted R^2									
Sel Seq	**Best Predictor**	**R^2**	**Adj R^2**	**MAPE**	**RMSE**	**SBC**	**Param Estimate**	**Std Error**	**t Value**	**Prob > |t|**
1	DirectMail_Adstock1	0.13	0.08	0.132	11456.0	1908.5	42.88	20.23	2.12	0.0340
2	DirectMail_Adstock4	0.20	0.15	0.128	11049.0	1905.6	-17.92	6.44	-2.78	0.0054
3	DirectMail_Adstock3	0.23	0.18	0.127	10864.0	1906.1	142.92	70.23	2.04	0.0418
4	DirectMail_Adstock2	0.30	0.24	0.123	10441.0	1902.6	-1122.81	396.6	-2.83	0.0046

Table 5: Direct Mail Performance Measures

We ran the same code on the Internet, PrintMedia, and TV and Radio time series and observed the following summaries of forward selection:

Sel Seq	Best Predictor	R^2	Adj R^2	MAPE	RMSE	SBC	Param Estimate	Std Error	t Value	Prob > |t|
			Summary of Forward Selection							
			Performance Measure = Adjusted R^2							
1	Internet_Adstock5	0.7258	0.7126	8.41%	6413.9	1806.6	-7.026	1.099	-6.394	<.0001
2	Internet_Adstock2	0.7770	0.7634	7.87%	5818.7	1792.8	33.549	7.660	4.380	<.0001
3	Internet_Adstock1	0.7826	0.7665	7.54%	5780.6	1795.0	-63.305	42.372	-1.494	0.1352
4	Internet_Adstock4	0.7971	0.7793	7.68%	5619.2	1793.5	-168.184	69.799	-2.410	0.0160
5	Internet_Adstock3	0.8013	0.7812	7.56%	5595.4	1796.2	910.48	695.088	1.310	0.1902

Table 6: Internet Performance Measures

Sel Seq	Best Predictor	R^2	Adj R^2	MAPE	RMSE	SBC	Param Estimate	Std Error	t Value	Prob > |t|
			Summary of Forward Selection							
			Performance Measure = Adjusted R^2							
1	PrintMedia_Adstock5	0.245	0.209	13.4%	10644	1896.0	-22.178	3.936	-5.635	<.0001
2	PrintMedia_Adstock4	0.315	0.274	12.7%	10196	1891.9	139.287	47.554	2.929	0.003
3	PrintMedia_Adstock1	0.319	0.268	12.6%	10233	1896.0	-59.858	95.419	-0.627	0.531
4	PrintMedia_Adstock2	0.319	0.260	12.6%	10293	1900.4	75.746	301.544	0.251	0.802
5	PrintMedia_Adstock3	0.334	0.267	12.1%	10242	1903.0	-3158.96	2370.17	-1.333	0.183

Table 7: Print Media Performance Measures

Sel Seq	Best Predictor	R^2	Adj R^2	MAPE	RMSE	SBC	Param Estimate	Std Error	t Value	Prob > |t|
			Summary of Forward Selection							
			Performance Measure = Adjusted R^2							
1	TVRadio_Adstock5	0.0896	0.0457	14.0%	11688	1912.1	-1.624	0.798	-2.035	0.0418
2	TVRadio_Adstock4	0.2284	0.1814	13.5%	10825	1902.2	33.974	8.509	3.993	<.0001
3	TVRadio_Adstock3	0.2777	0.2242	13.3%	10537	1901.0	-86.953	36.660	-2.372	0.0177
4	TVRadio_Adstock2	0.2866	0.2242	13.2%	10536	1904.5	84.637	84.898	0.997	0.3188
5	TVRadio_Adstock1	0.3471	0.2809	12.3%	10143	1901.3	-266.293	97.645	-2.727	0.0064

Table 8: TV and Radio Performance Measures

MODEL BUILDING USING BASELINE + ALL-SUBSETS SELECTION

One way to use the %ARIMA_SELECT macro in all-subsets selection mode is to simply put all of the time series into the model and let the macro do the work. The shortcomings of this approach have been discussed above, so we must try to be more clever. We can use the %ARIMA_SELECT macro judiciously to select the best single predictor in each of the four media groups, and then use the four variables chosen in a much more manageable all-subsets execution.

We used the baseline algorithm to select the best predictors from each of the media channels. The original media predictor was included in the set of candidate predictors. These predictors will be used as candidate predictors to demonstrate all-subsets selection.

The best predictor for each media channel was selected and included as a candidate predictor. The code to run the all-subsets algorithm is shown below.

```
%let CANDIDATES = DirectMail_Adstock1 Internet PrintMedia_Adstock5 TVRadio ;

%ARIMA_SELECT( SASDATA.Saledata_Adstock, TIMEID=DATE
            , DEPVAR=SalesAmount, INDVAR=&CANDIDATES, STARTVAR=
            , DELTA_RSQ=
            , NOTES=n, PRINT=n
            , MEASURE=adj_rsq, SELECT=SUB, START=1, STOP=, TOPN=10
            , D=, NLAG=24, I_OTHER_OPT=
            , MAXITER=50, METHOD=ml, P=5, Q=, E_OTHER_OPT=
            , INTERVAL=week
            , F_OTHER_OPT=%str( align=BEGINNING )
            , ODSFILE=Path\File
            )
```

Results of the all-subsets selection run are shown in Tables 9 and 10, and graphically in Figure 2.

%ARIMA_SELECT Subsets		
# of Vars	Sub set #	Variables
1	1	TVRadio
	2	PrintMedia_Adstock5
	4	Internet
	8	DirectMail_Adstock1
2	3	PrintMedia_Adstock5 TVRadio
	6	Internet PrintMedia_Adstock5
	9	DirectMail_Adstock1 TVRadio
	10	DirectMail_Adstock1 PrintMedia_Adstock5
	12	DirectMail_Adstock1 Internet
3	11	DirectMail_Adstock1 PrintMedia_Adstock5 TVRadio
	13	DirectMail_Adstock1 Internet TVRadio
	14	DirectMail_Adstock1 Internet PrintMedia_Adstock5
4	15	DirectMail_Adstock1 Internet PrintMedia_Adstock5 TVRadio

Table 9: All Subsets of Predictors

# of Vars	Subset #	R^2	Adj R^2	MAPE	RMSE	SBC
		Summary of All-Subsets Selection				
		Performance Measure = ADJ_R_SQR				
1	4	0.6435	0.631	7.74%	7270.3	1825.3
	8	0.1086	0.077	13.60%	11497.0	1905.7
	1	0.0487	0.015	14.30%	11877.0	1911.4
	2	0.0398	0.006	15.10%	11932.0	1912.2
2	12	0.9478	0.945	3.44%	2797.4	1661.6
	6	0.6581	0.642	7.48%	7162.6	1826.0
	10	0.2097	0.172	12.80%	10889.0	1899.6
	3	0.1443	0.103	13.30%	11331.0	1906.7
	9	0.1110	0.068	13.70%	11549.0	1909.9
3	14	0.9488	0.946	3.44%	2789.0	1664.5
	13	0.9478	0.945	3.44%	2814.2	1666.1
	11	0.2098	0.162	12.80%	10955.0	1904.1
4	15	0.9488	0.945	3.43%	2804.3	1668.9

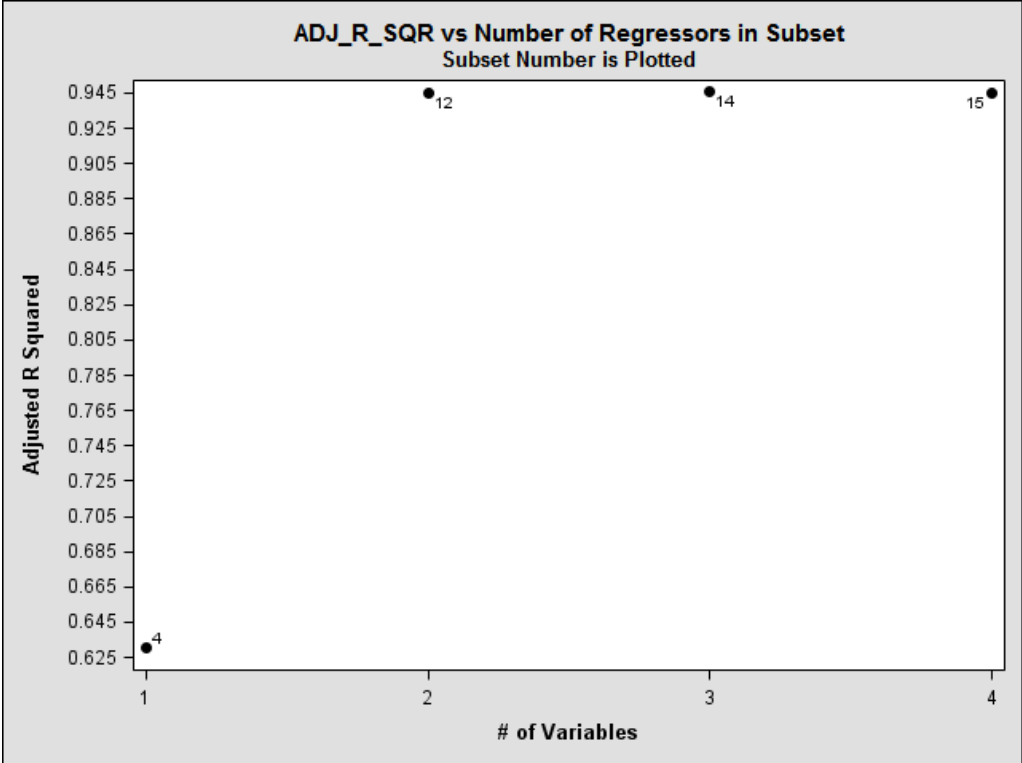**Table 10: Summary of All-Subsets Selection**

Figure 2: Performance Measure as a Function of Number of Regressors

We may readily conclude from Figure 2 that there is little, if any, performance improvement over more than two variables in a model, and that the exogenous predictor time series to use are represented by subset 12. The predictors that comprise subset 12 are DirectMail_Adstock1 and Internet.

SUMMARY

We have developed a set of selection methodologies for time series variables that involve repeated executions of PROC ARIMA. The baseline methodology produces a set of variables that are independently compared to the response variable. The forward selection methodology runs successive models that incorporate the best-performing variables of prior executions and hence builds marginal models. The all-subsets methodology computes subset models representing combinations of variables that exhaustively span the "predictor space" and indicates best-performing subsets, but at high computational expense for a large set of predictors. The performance measures for each execution are collected and used to intelligently focus attention on predictors that are most closely associated with the response variable of interest. The number of predictors may be limited to, e.g., the top 10, to comply with business logic constraints or corporate practices.

APPENDIX

Parameters for the %ARIMA_SELECT macro are displayed in Table A-1.

Note: parameters with equal sign ('=') are optional.

Parameter	Description	Value / Default
DSNIN	Name of input dataset	
TIMEID	Name of sequence variable	
DEPVAR	Name of time series response variable	
INDVAR=	list of independent dynamic regressors	
STARTVAR=	list of variables that are initially included in model	

Macro execution control parameters

DELTA_RSQ=	minimum increase in R^2 above baseline to include variable	$0 < R^2 < 1$
NOTES=	flag to turn off putting NOTE: statements to log (timesaver)	Default: N
MEASURE=	flag to specify measure to use for sort order of predictors	ADJ_RSQ, MAPE, RMSE, RSQ Default: RSQ
PRINT=	flag to suppress printing ARIMA results to output (timesaver)	Default: N
SELECT=	independent variable selection criterion	BASE, FWD, SUBSET, BASE_FWD, BASE_SUB Default: BASE
START=	number of vars in initial subset of independent vars	Default: 1
STOP=	number of vars in final subset of independent vars	Default: # of time series
TOPN=	# of variables to use for forward selection or all-subsets selection	Default: 10

ARIMA IDENTIFY options

D=	differencing lag(s)	
NLAG=	# of lags to consider in computing auto-and cross-correlations	Default: min(24, floor(#_of_obs/4))
I_OTHER_OPT=	other IDENTIFY statement options, e.g., ESACF, MINIC, SCAN, NOMISS	

ARIMA ESTIMATE options

MAXITER=	maximum # of iterations allowed	Default: 50
METHOD=	estimation method to use	Default: ml
P=	autoregressive specification of the model	
Q=	moving average specification of the model	
E_OTHER_OPT=	[optional] other ESTIMATE statement options	

ARIMA FORECAST options

INTERVAL=	[optional] time interval between observations	DAY, WEEK, MONTH, YEAR, &c.
F_OTHER_OPT=	other FORECAST statement options	

Output file names

OUTFILE=	Path\File of output file containing variable names and stats for case when &SELECT=BASE	
ODSFILE=	Path\File of HTML file produced that contains all %ARIMA_SELECT output for off-line inspection and documentation	

Table A-1: %ARIMA_SELECT Parameters

REFERENCES

1. Flom, Peter L. and David L. Cassell. (2007), "Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use", http://www.nesug.org/proceedings/nesug07/sa/sa07.pdf
2. Asimov, Isaac. (1959), Nine Tomorrows, Doubleday & Company, Garden City, New York.
3. Dickey, David and Terry Woodfield. (2011), "Forecasting Using SAS® Software: A Programming Approach Course Notes", SAS Institute, Cary, NC.
4. Joseph, Joy. (2006), "Understanding Advertising Adstock Transformations", Munich Personal RePEc Archive, MPRA Paper No. 7683, http://mpra.ub.uni-muenchen.de/7683, 12 March 2008 01:45 UTC.

BIBLIOGRAPHY

SAS Institute, Inc. (2013). SAS/ETS User's Guide, Version 12.1.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

> Ross Bettinger
>
> rsbettinger@gmail.com