Paper SAS 3471-2019

Using SAS® Forecast Server and the SASEFRED Engine to Enhance Your Forecast

Catherine LaChapelle, SAS Institute

ABSTRACT

The SASEFRED interface engine is the best-kept secret in SAS/ETS® software. It dramatically reduces the amount of time and effort required to include economic indicator variables in your time series analysis. Using the SASEFRED engine in SAS® Enterprise Guide®, I can directly query the economic database of the Federal Reserve Bank of St. Louis. This public database contains over 529,000 economic time series aggregated from 86 sources. In this paper, I forecast wine demand and enrich my predictions via the inclusion of economic variables such as "Retail Sales: Beer, Wine, and Liquor Stores" and "Producer Price Index by Industry: Beer, Wine, and Liquor Stores." Although I use a retail example, this technology is relevant to all industries. The diversity of economic variables provided in this database ensures that it is useful to virtually every time series analysis and industry. This specific example leverages SAS Enterprise Guide and SAS® Forecast Server as interfaces. However, this functionality works on SAS® 9.4 as well as on SAS® Viya® technology.

INTRODUCTION

The SASEFRED interface engine is a powerful tool that allows you to seamlessly integrate external economic variables into your forecasting project. In this paper I introduce the engine, discuss the FRED data source, and show an example of how to pull all this data together in a SAS project. By including external macroeconomic variables, I can enhance the accuracy of my forecast and answer additional questions about my models and data.

WHY ENRICH TIME SERIES DATA?

As a forecaster, there are many factors I must consider when I am trying to build an accurate model to predict future values of my dependent variable. My goal is to accurately identify the signal, while ignoring the noise in the data. Although the signal of my data might be represented strongly by my dependent variable, in most cases there are independent variables which also add relevant information. For example, if I am trying to predict product sales in a given month, it is likely that the promotion I am running on that product is relevant. By including these independent variables in my time series model, I can create more accurate forecasts.

Another benefit of including independent variables is the ability to perform scenario analysis. Scenario analysis allows me to test the "what if" of a situation. For example, if I am predicting wine sales and I have a promotion variable, I can test different scenarios. Using the previous example, I could examine what would happen if I discounted a particular bottle of wine by 20% this month instead of 10%. I might want to look at whether my sales increase proportionally and what is the impact on my profit. This example uses a business variable I probably already have contained in my database. If we expand from this type of scenario, we can look at macroeconomic factors at play. In this paper's example, I include

several wine-specific variables contained within the FRED database. These variables give me the ability to look at the larger context surrounding my business problem, not just the ones contained within my business.

Typically, when I build a model as a statistician, I am wary of adding too many independent variables for fear of introducing too much extraneous noise to my model. One of the great benefits of using SAS® Forecast Server and SAS® Visual Forecasting as forecasting tools is the automated variable selection feature within the software. In both products, when I am creating a forecasting project I am prompted to assign each independent variable to a classification for use. The option **Use if Significant** provides a lot of value and automation for the forecaster. This option automatically tests each independent variable for significance against each time series evaluated in the project. For example, if you have a hierarchical forecasting project with 1,000 individual series, SAS Forecast Server and SAS Visual Forecasting will test each of these series individually for significance with each independent variable. This automated feature selection makes it more practical to pull in all of the possible independent variables, such as economic information from FRED, and test them at scale.

HOW SASEFRED WORKS

The SASEFRED interface engine allows you to directly query the Federal Reserve Economic Database of St. Louis. The Federal Reserve Bank of St. Louis is in the Eighth District of the Federal Reserve System in the United States. The charter of the Research Division of the Bank is to advise the Bank president on relevant factors that influence economic policy. These areas of research include money and banking, macroeconomics, and international and regional economics. Currently there are over 529,000 individual time series published by FRED at varying levels of time series aggregation. These time series are aggregated from 87 distinct data sources. Due to the breadth of time series in this database, there is relevant economic information for virtually every industry.

This database is published and maintained by the United States government and thus is free for anyone to access. It does not require a paid subscription to access the data, but requires only the generation of a free API key for the query. More information about the creation of an API key can be found in the SASEFRED documentation on the FRED website.

CODE EXAMPLE- WINE SALES FORECASTING

In this paper I apply the SASEFRED engine to capture wine-related macroeconomic variables and enhance my existing wine sales forecasting data set. I use SAS Enterprise Guide as an interface to run the program, prepare my data, and then join my tables together into a final data set for forecasting. Figure 1 depicts an overview of the process flow in SAS Enterprise Guide.

Figure 1 is an overview of the data manipulation process in SAS® Enterprise Guide.

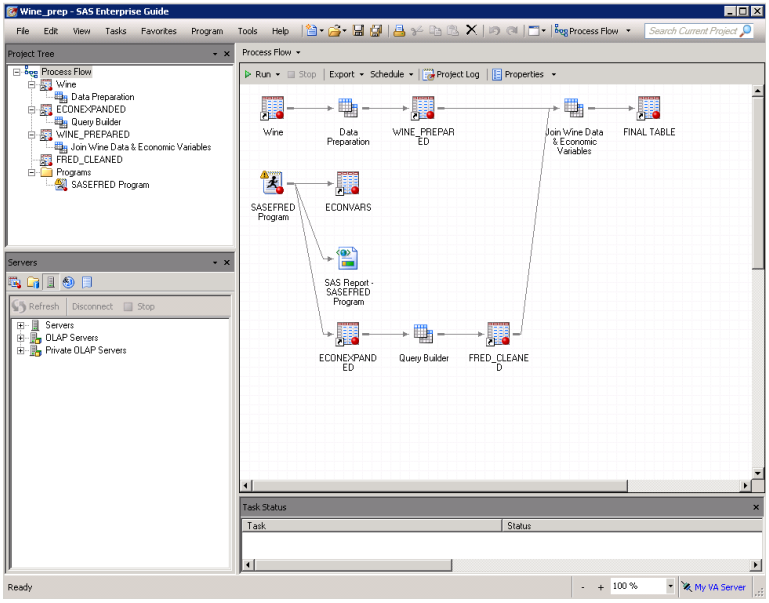

Figure 1. SAS Enterprise Guide Project Overview

The SASEFRED engine runs in a SAS LIBNAME statement. The sample code that I use for this example is included below. It is easy to copy and paste this code into your own program. You can then simply change out the relevant fields such as your file directory where you would like to save the generated table and your own individual API key in place of the x's I have here as a placeholder.

To determine which time series I would like to request for my analysis, I go to the FRED website and search for a relevant term for the data I am trying to model. Each series is tagged, which makes it easier to narrow my search. I can filter the results by concept, geography, geography type, frequency, source, release, and seasonal adjustment.

In this example, I searched for "wine," and chose three time series that seemed like they could add relevant additional information to my model. Because the series names are a string of letters and numbers and are often difficult to decipher, I recommend adding a key in a commented section of your code. Every time I add a series to my code, I add the description of the series to the comments so that I can easily reference it later.

```
title 'Retrieve Economic Indicator Variables';
libname _all_ clear;
libname fred sasefred "D:\FRED"
   OUTXML=fredex01
   AUTOMAP=replace
   MAPREF=MyMap
   XMLMAP="D:\FRED\fredex01.map"
   APIKEY='xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'
   IDLIST='PCU31213031213009, MRTSSM4453USN, PCU4453104453101'
   START='2014-02-23'
   END='2016-10-23'
   FREQ='m'
   OUTPUT=1
```

```
   AGG='avg'
   FORMAT=xml;

data econvars;
   set fred.fredex01 ;
run;
   /* Retail Sales: Beer, Wine, and Liquor Stores (MRTSSM4453USN)
   Producer Price Index by Industry: Wineries: Wines, Dessert,
   Effervescent, and Wine Coolers(PCU31213031213009)
   Producer Price Index by Industry: Beer, Wine, and Liquor Stores:
   Retailing of Beer, Wine, and Liquor(PCU445310445310l1) */
```

Figure 2 is a sample of the first 30 observations from the SASEFRED code example.

	date	realtime_start	realtime_end	PCU31213031213009	MRTSSM4453USN	PCU445310445310l1
1	2014-01-01	2019-03-21	2019-03-21	100.7	3381	122.9
2	2014-02-01	2019-03-21	2019-03-21	100.7	3310	124.7
3	2014-03-01	2019-03-21	2019-03-21	101.5	3652	121.5
4	2014-04-01	2019-03-21	2019-03-21	101.5	3702	122.2
5	2014-05-01	2019-03-21	2019-03-21	101.5	4165	122.1
6	2014-06-01	2019-03-21	2019-03-21	101.5	4036	124.3
7	2014-07-01	2019-03-21	2019-03-21	101.5	4242	125.2
8	2014-08-01	2019-03-21	2019-03-21	101.5	4219	125.9
9	2014-09-01	2019-03-21	2019-03-21	101.5	3814	124.1
10	2014-10-01	2019-03-21	2019-03-21	101.6	4090	127.3
11	2014-11-01	2019-03-21	2019-03-21	101.6	4103	126.2
12	2014-12-01	2019-03-21	2019-03-21	101.7	5572	125.9
13	2015-01-01	2019-03-21	2019-03-21	101.6	3572	128.3
14	2015-02-01	2019-03-21	2019-03-21	101.7	3482	127.7
15	2015-03-01	2019-03-21	2019-03-21	101.7	3857	126.2
16	2015-04-01	2019-03-21	2019-03-21	101.7	3867	127
17	2015-05-01	2019-03-21	2019-03-21	101.7	4335	125.2
18	2015-06-01	2019-03-21	2019-03-21	101.7	4217	125.4
19	2015-07-01	2019-03-21	2019-03-21	101.7	4532	126.8
20	2015-08-01	2019-03-21	2019-03-21	101.6	4260	127.9
21	2015-09-01	2019-03-21	2019-03-21	101.6	4073	127.1
22	2015-10-01	2019-03-21	2019-03-21	101.7	4273	127
23	2015-11-01	2019-03-21	2019-03-21	101.8	4266	126
24	2015-12-01	2019-03-21	2019-03-21	101.8	5816	125.4
25	2016-01-01	2019-03-21	2019-03-21	103.3	3610	124.5
26	2016-02-01	2019-03-21	2019-03-21	103.3	3773	125
27	2016-03-01	2019-03-21	2019-03-21	103.3	4051	123.6
28	2016-04-01	2019-03-21	2019-03-21	103.3	4114	123.5
29	2016-05-01	2019-03-21	2019-03-21	104.2	4403	124.7
30	2016-06-01	2019-03-21	2019-03-21	104.2	4474	124.8

Figure 2. Output Table from SASEFRED Code

After I have identified the series I would like to query, the next step is to identify the temporal component of the request. The three components of the temporal component are the start date, the end date, and the frequency of the desired output. Because the same level of aggregation is not available for every series, it is important to review the frequency options for each time series on the FRED website. In this example, there are three series whose most granular frequency level is monthly, but the wine sales data I am trying to predict uses a weekly frequency. Figure 2 contains a sample of the output from the SASEFRED code.

The FREQ option in the SASEFRED engine allows you to convert a higher frequency series to a lower frequency series. For example, if it is published in a monthly format and I would like quarterly data I can request quarterly data in the FREQ statement and the interface engine will automatically do the conversion for me.

```
proc expand data=econvars out=econexpanded from=month to=week;
    id date;
    convert PCU31213031213009;
    convert MRTSSM4453USN;
    convert PCU4453104453101;
run;
```

Because I want to change the level of aggregation, I follow my query with PROC EXPAND code. The EXPAND procedure allows me to change the temporal aggregation of my input data set. Because I am going from a lower frequency (monthly) to a higher frequency (weekly), the EXPAND procedure uses a cubic spline of the input data to determine the new values of my series.

Figure 3 is a sample of the first 30 observations of the final merged table with all variables.

	Date	Region	Varietal	Sales	Promotion	PPI_Wineries	Retail_Sales_Wine	PPI_Wine_Stores
1	23FEB2014	Mid West	Bordeaux	144	0	101.37546337	3605.7617208	121.9511894
2	02MAR2014	Mid West	Bordeaux	165	0	101.51508202	3656.5664728	121.4550645
3	09MAR2014	Mid West	Bordeaux	312	0	101.57676297	3667.0951475	121.36877747
4	16MAR2014	Mid West	Bordeaux	243	0	101.57959509	3658.2830483	121.56527095
5	23MAR2014	Mid West	Bordeaux	269	0	101.54872827	3655.5728475	121.88119704
6	30MAR2014	Mid West	Bordeaux	347	0	101.5093124	3684.4072173	122.15320787
7	06APR2014	Mid West	Bordeaux	262	0	101.48454451	3766.6980135	122.23563683
8	13APR2014	Mid West	Bordeaux	197	0	101.47843697	3889.6703495	122.15651828
9	20APR2014	Mid West	Bordeaux	165	0	101.48411313	4020.8615052	122.04247746
10	27APR2014	Mid West	Bordeaux	197	0	101.49457134	4127.5827881	122.02127117
11	04MAY2014	Mid West	Bordeaux	248	0	101.50292459	4177.9484668	122.21750296
12	11MAY2014	Mid West	Bordeaux	254	0	101.50607242	4166.6002542	122.65160101
13	18MAY2014	Mid West	Bordeaux	261	0	101.50547782	4120.1495977	123.21844582
14	25MAY2014	Mid West	Bordeaux	216	0	101.50287546	4067.1112591	123.80544347
15	01JUN2014	Mid West	Bordeaux	352	0	101.5	4036	124.3
16	08JUN2014	Mid West	Bordeaux	222	0	101.49825898	4047.8896447	124.62127518
17	15JUN2014	Mid West	Bordeaux	177	0	101.49775147	4094.0902671	124.81544366
18	22JUN2014	Mid West	Bordeaux	241	0	101.49824941	4158.4710038	124.96043379
19	29JUN2014	Mid West	Bordeaux	225	0	101.49952475	4224.900991	125.13417392
20	06JUL2014	Mid West	Bordeaux	383	0	101.5013111	4277.8351305	125.40288716
21	13JUL2014	Mid West	Bordeaux	292	0	101.50296562	4307.482882	125.71780418
22	20JUL2014	Mid West	Bordeaux	341	0	101.50363178	4307.319932	125.96488716
23	27JUL2014	Mid West	Bordeaux	287	0	101.50245058	4270.8594564	126.02934913
24	03AUG2014	Mid West	Bordeaux	268	0	101.49858051	4191.7617811	125.79772994
25	10AUG2014	Mid West	Bordeaux	237	0	101.49270344	4076.5077044	125.2721674
26	17AUG2014	Mid West	Bordeaux	213	0	101.48818511	3954.1656	124.65846402
27	24AUG2014	Mid West	Bordeaux	248	0	101.48866448	3856.1030653	124.18315365
28	31AUG2014	Mid West	Bordeaux	278	0	101.4977805	3813.6876979	124.07277015
29	07SEP2014	Mid West	Bordeaux	352	0	101.51808652	3850.1853592	124.50180288
30	14SEP2014	Mid West	Bordeaux	315	0	101.54543636	3938.8636972	125.3235592

Figure 3. Final Merged Table with Dependent and Independent Variables

The final step in the process, as seen in Figure 1, is to merge the table of economic indicator variables, with the wine sales data. I use the Query Builder in SAS Enterprise Guide to join the tables and do the final data preparation. Figure 3 contains the first 30 observations of the final data preparation step in this process. When my final table is prepared for forecasting, I am prepared to use my forecasting engine of choice. With SAS Forecast Server and SAS Visual Forecasting, I can leverage the automatic independent variable selection feature. This feature tests the statistical significance of my independent variables and includes only those variables that are significant, ensuring that I will not overfit my time series model by including these independent variables.

ADDITIONAL ETS INTERFACE ENGINES

In addition to the SASEFRED interface engine, there are 10 other ETS interface engines with similar functionality to SASEFRED. Each of these additional interface engines connects to a different externally published database.

Interface Engine	Database
SASECRSP	Center for Research in Security Prices
SASEFRED	Federal Reserve Economic Database of St. Louis (FRED)
SASEFAME	Fame
SASEHAVR	Haver Analytics DLX (Data Link Express)
SASENOAA	National Oceanic and Atmospheric Administration (NOAA)
SASEOECD	Organisation for Economic Co-Operation and Development (OECD)
SASEQUAN	Quandl
SASERAIN	World Weather Online
SASEWBGO	World Bank Group Open (WBGO) data website
SASEXCCM	CRSP/Compustat Merged (CCM) Database
SASEXFSD	FactSet OnDemand data service

Table 1. Full List of SAS/ETS Interface Engines

Some of these databases require a subscription, but others, such as the FRED data, are free to the public. Development work on these interface engines is ongoing, and SAS continues to add support for new engines. Table 1 lists of all the current SAS/ETS interface engines and the databases to which they are connected.

CONCLUSION

The SASEFRED interface engine is a powerful tool in every forecaster's arsenal. With a few simple lines of code in a LIBNAME statement, you can quickly and easily integrate econometric time series data into a SAS table. Because of the powerful SAS forecasting engines, I can pull many times series variables from FRED that might be relevant, and let SAS statistical models test their significance for me.

REFERENCES

"Economic Research." Federal Reserve Bank of St. Louis. http://research.stlouisfed.org. Accessed March 20, 2019.

"API Keys." Federal Reserve Bank of St. Louis. https://research.stlouisfed.org/docs/api/api_key.html. Accessed March 25, 2019.

SAS Institute Inc. 2018. *SAS/ETS 15.1 User's Guide*. Cary, NC: SAS Institute Inc. Available https://support.sas.com/documentation/onlinedoc/ets/151/etsug.pdf. Accessed March 20, 2019.

ACKNOWLEDGMENTS

Thank you to my manager, Vinicius Vivaldi, for your unending support. This paper would not have been possible without Kelly Fellingham, the developer in SAS R&D responsible for all the ETS interface engines. A big thanks to the technical editor of this paper, David Bass.

RECOMMENDED READING

SAS Institute Inc. 2018. *SAS/ETS 15.1 User's Guide*. Cary, NC: SAS Institute Inc. Available https://support.sas.com/documentation/onlinedoc/ets/151/etsug.pdf. Accessed March 20, 2019.

CONTACT INFORMATION

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

Catherine LaChapelle
SAS Institute
212-413-2573
Catherine.LaChapelle@sas.com
www.linkedin.com/in/CatherineLachapelle