```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
73
74         ********************************************************************************
75         *  TITLE :      SAS GRAIN PRICE PROJECT ANALYSIS
76         *
77         *  DESCRIPTION: Final project for BIOS 7400 with Xiao Song, UGA, Spring 2022.
78         *               Simple analysis of grain price data.
79         *
80         *-------------------------------------------------------------------------------
81         *  JOB NAME:    analysis.SAS
82         *  LANGUAGE:    SAS v9.4 (on demand for academics)
83         *
84         *  NAME:        Zane Billings
85         *  DATE:        2022-04-22
86         *
87         ********************************************************************************;
88
89         FOOTNOTE "Job run by Zane Billings on &SYSDATE at &SYSTIME.";
90
91         TITLE 'ANALYSIS OF USDA HISTORICAL GRAIN PRICE DATA';
92
93         OPTIONS NODATE LS=95 PS=42;
94
95         LIBNAME HOME '/home/u59465388/SAS-Grain-Prices';
NOTE: Libref HOME was successfully assigned as follows:
      Engine:        V9
      Physical Name: /home/u59465388/SAS-Grain-Prices
96
97         ODS GRAPHICS / WIDTH = 6in HEIGHT = 6in;
98
99         ********************************************************************************;
100        * Show the descriptor portion of the dataset;
101        ********************************************************************************;
102
103        TITLE2 "CONTENTS OF GRAINS DATASET";
104
105        PROC CONTENTS DATA = HOME.GRAINS;
106        RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.09 seconds
      user cpu time       0.09 seconds
      system cpu time     0.00 seconds
      memory              3453.28k
      OS Memory           33196.00k
      Timestamp           04/25/2022 10:09:39 PM
      Step Count                        51  Switch Count  0
      Page Faults                       0
      Page Reclaims                     366
      Page Swaps                        0
      Voluntary Context Switches        5
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           32


107
108        ********************************************************************************;
109        * Plot outcome time series;
110        ********************************************************************************;
111
112        FOOTNOTE; * Remove the footnote so it isn't on the graphs;
```

```
113
114         * Plot the time series of log grain price over time. This makes a separate
115         time series line for each grain.;
116         TITLE2 "PRICE PER BUSHEL OF GRAINS OVER TIME";
117         PROC SGPLOT DATA = HOME.GRAINS;
118         SERIES X = YEAR Y = LPE / GROUP = GRN;
119         RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           2.78 seconds
      user cpu time       0.16 seconds
      system cpu time     0.03 seconds
      memory              20592.71k
      OS Memory           53672.00k
      Timestamp           04/25/2022 10:09:42 PM
      Step Count                        52  Switch Count  1
      Page Faults                       0
      Page Reclaims                     6898
      Page Swaps                        0
      Voluntary Context Switches        2147
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1168

NOTE: There were 571 observations read from the data set HOME.GRAINS.

120
121         * Color the points of the time series by the President's political party. The
122         default colors are already red and blue so we don't need to change them!
123         Also plots a gray line underneath the points, since the JOIN option for
124         the SCATTER statement will not connect the points in order.;
125         TITLE2 "GRAIN PRICES AND PRESIDENT'S POLITICAL PARTY OVER TIME";
126         PROC SGPANEL DATA = HOME.GRAINS;
127         PANELBY GRN;
128         SERIES X = YEAR Y = LPE / LINEATTRS = (COLOR = "GRAY") SMOOTHCONNECT;
129         SCATTER X = YEAR Y = LPE / GROUP = PARTY
130         MARKERATTRS = (SYMBOL = CIRCLEFILLED);
131         RUN;

NOTE: PROCEDURE SGPANEL used (Total process time):
      real time           1.23 seconds
      user cpu time       0.29 seconds
      system cpu time     0.03 seconds
      memory              6709.75k
      OS Memory           58460.00k
      Timestamp           04/25/2022 10:09:43 PM
      Step Count                        53  Switch Count  2
      Page Faults                       0
      Page Reclaims                     2171
      Page Swaps                        0
      Voluntary Context Switches        3703
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           1952

NOTE: There were 571 observations read from the data set HOME.GRAINS.

132
133         * Make a boxplot of log price vs. president's political party. This ignores the
134         time series information, but can tell us if either party has more high or
135         low years compared to the other.;
136         TITLE2 "LOG PRICE DISTRIBUTION BY PRESIDENT'S POLITICAL PARTY";
137         PROC SGPANEL DATA = HOME.GRAINS;
138         PANELBY GRN;
139         HBOX LPE / GROUP = PARTY;
```

```
140        RUN;

NOTE: PROCEDURE SGPANEL used (Total process time):
      real time           0.62 seconds
      user cpu time       0.17 seconds
      system cpu time     0.02 seconds
      memory              4159.90k
      OS Memory           57268.00k
      Timestamp           04/25/2022 10:09:44 PM
      Step Count                        54  Switch Count  26
      Page Faults                       0
      Page Reclaims                     1630
      Page Swaps                        0
      Voluntary Context Switches        1110
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           2072

NOTE: There were 571 observations read from the data set HOME.GRAINS.

141
142        ODS GRAPHICS / WIDTH = 6in HEIGHT = 9in;
143
144        * Make a scatterplot of the log price vs each covariate, ignoring the time
145        series component of the data. There is not an easy way to connect the points
146        like a phase portrait using PROC SGPLOT.;
147        * I divided this into two plots so they would fit on one page nicer. In the final
148        manuscript they could be put side by side.;
149        TITLE2 "SCATTERPLOTS OF PRICE VS COVARIATES";
150        PROC SGSCATTER DATA = HOME.GRAINS;
151        PLOT LPE * (ACR HVT LNR PRD YLD) / REG
152        COLUMNS = 2 GROUP = GRN;
153        RUN;

NOTE: PROCEDURE SGSCATTER used (Total process time):
      real time           0.73 seconds
      user cpu time       0.29 seconds
      system cpu time     0.02 seconds
      memory              7200.78k
      OS Memory           61360.00k
      Timestamp           04/25/2022 10:09:45 PM
      Step Count                        55  Switch Count  1
      Page Faults                       0
      Page Reclaims                     2192
      Page Swaps                        0
      Voluntary Context Switches        2330
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           5696

NOTE: There were 571 observations read from the data set HOME.GRAINS.

154
155        PROC SGSCATTER DATA = HOME.GRAINS;
156        PLOT LPE * (INFL PWR TEMP VALUE) / REG
157        COLUMNS = 2;
158        RUN;

NOTE: PROCEDURE SGSCATTER used (Total process time):
      real time           0.53 seconds
      user cpu time       0.30 seconds
      system cpu time     0.02 seconds
      memory              6859.84k
      OS Memory           63664.00k
      Timestamp           04/25/2022 10:09:45 PM
```

```
      Step Count                        56  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1948
      Page Swaps                        0
      Voluntary Context Switches        391
      Involuntary Context Switches      10
      Block Input Operations            0
      Block Output Operations           4512

NOTE: There were 571 observations read from the data set HOME.GRAINS.

159
160        ********************************************************************************;
161        * Plots of covariates across time;
162        ********************************************************************************;
163
164        * Plot the time series of each covariate, to assess how they change. I split
165        this one into two plots to prevent the plots being too small as before.;
166        TITLE2 "CHANGE IN COVARIATES ACROSS TIME";
167        PROC SGSCATTER DATA = HOME.GRAINS;
168        PLOT (ACR HVT LNR PRD YLD) * YEAR /
169        COLUMNS = 2 GROUP = GRN JOIN MARKERATTRS = (SIZE = 0);
170        RUN;

NOTE: PROCEDURE SGSCATTER used (Total process time):
      real time           0.49 seconds
      user cpu time       0.19 seconds
      system cpu time     0.03 seconds
      memory              5676.53k
      OS Memory           63152.00k
      Timestamp           04/25/2022 10:09:46 PM
      Step Count                        57  Switch Count  1
      Page Faults                       1
      Page Reclaims                     1882
      Page Swaps                        0
      Voluntary Context Switches        2021
      Involuntary Context Switches      0
      Block Input Operations            24
      Block Output Operations           1592

NOTE: There were 571 observations read from the data set HOME.GRAINS.

171
172        PROC SGSCATTER DATA = HOME.GRAINS;
173        PLOT (INFL PWR TEMP VALUE) * YEAR /
174        COLUMNS = 2 JOIN MARKERATTRS = (SIZE = 0);
175        RUN;

NOTE: PROCEDURE SGSCATTER used (Total process time):
      real time           0.28 seconds
      user cpu time       0.11 seconds
      system cpu time     0.01 seconds
      memory              3918.65k
      OS Memory           60516.00k
      Timestamp           04/25/2022 10:09:46 PM
      Step Count                        58  Switch Count  1
      Page Faults                       0
      Page Reclaims                     802
      Page Swaps                        0
      Voluntary Context Switches        281
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1008

NOTE: There were 571 observations read from the data set HOME.GRAINS.
```

```
176
177         ********************************************************************************;
178         * Univariate analyses;
179         ********************************************************************************;
180
181         * Univariate analysis of main outcome (log price) by grain type;
182         TITLE2 "UNIVARIATE SUMMARY OF GRAIN DATA OVER TIME";
183
184         ODS GRAPHICS / WIDTH = 4in HEIGHT = 4in;
185
186         PROC UNIVARIATE DATA = HOME.GRAINS PLOTS;
187         VAR LPE;
188         CLASS GRN;
189         RUN;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           1.10 seconds
      user cpu time       0.57 seconds
      system cpu time     0.03 seconds
      memory              5072.87k
      OS Memory           60340.00k
      Timestamp           04/25/2022 10:09:47 PM
      Step Count                        59  Switch Count  0
      Page Faults                       0
      Page Reclaims                     1842
      Page Swaps                        0
      Voluntary Context Switches        1178
      Involuntary Context Switches      13
      Block Input Operations            0
      Block Output Operations           2968


190
191         ********************************************************************************;
192         * Bivariate analyses of price and covariates, ignoring time;
193         ********************************************************************************;
194
195         TITLE2 "BIVARIATE CORRELATIONS ACROSS NUMERICAL VARIABLES";
196         * Correlations -- check to see which covariates are correlated with the outcome,
197         and which are correlated with each other and should not be modeled
198         together.;
199         PROC CORR PEARSON SPEARMAN DATA = HOME.GRAINS;
200         VAR LPE ACR HVT LNR PRD YLD INFL PWR TEMP VALUE;
201         BY GRN;
202         RUN;

NOTE: PROCEDURE CORR used (Total process time):
      real time           1.39 seconds
      user cpu time       1.39 seconds
      system cpu time     0.00 seconds
      memory              2451.81k
      OS Memory           60336.00k
      Timestamp           04/25/2022 10:09:49 PM
      Step Count                        60  Switch Count  0
      Page Faults                       0
      Page Reclaims                     305
      Page Swaps                        0
      Voluntary Context Switches        3
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           496


203
```

```
204         TITLE2 "SUMMARY STATISTICS BY PRESIDENTIAL PARTY AND GRAIN";
205         * Mean difference in LPE by party -- proc corr does not have a point biserial
206         option, so we can check the difference/overlap in means and standard errors
207         to assess if party seems to impact log price for any of the grains.;
208         PROC MEANS DATA = HOME.GRAINS MEAN STDERR MEDIAN RANGE NWAY;
209         VAR LPE;
210         CLASS PARTY;
211         BY GRN;
212         RUN;

NOTE: There were 571 observations read from the data set HOME.GRAINS.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.05 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2081.59k
      OS Memory           61620.00k
      Timestamp           04/25/2022 10:09:49 PM
      Step Count                        61  Switch Count  4
      Page Faults                       0
      Page Reclaims                     324
      Page Swaps                        0
      Voluntary Context Switches        17
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           24


213
214         ********************************************************************************;
215         * Simple and multiple OLS regression models;
216         ********************************************************************************;
217
218         TITLE2 "SIMPLE LINEAR REGRESSION MODELS";
219         * Write a macro to fit all regression models of the form
220         MODEL COVAR GRN COVAR * GRN
221          without having to type out all of the PROC GLM statements. This model will
222          be parametrized without an intercept, and will generate all appropriate
223          diagnostic plots for the model.;
224
225         %MACRO ALLSIMPLE(DAT = , RESP = , PRED = );
226         %LET N = %SYSFUNC(COUNTW(&DAT));
227         %DO I = 1 %TO &N;
228         PROC GLM DATA = &DAT PLOTS = ALL;
229         CLASS GRN;
230         MODEL &RESP = %SCAN(&PRED, &I) | GRN / NOINT;
231         RUN;
232         %END;
233         %MEND;
234
235         %ALLSIMPLE(
236         DAT = HOME.GRAINS,
237         RESP = LPE,
238         PRED = ACR PRD INFL TEMP PWR YEAR
239         );

NOTE: Due to the presence of CLASS variables, an intercept is implicitly fitted.  R-Square has
      been corrected for the mean.

NOTE: PROCEDURE GLM used (Total process time):
      real time           1.56 seconds
      user cpu time       0.67 seconds
      system cpu time     0.06 seconds
      memory              10490.46k
      OS Memory           68324.00k
```

```
      Timestamp           04/25/2022 10:09:50 PM
      Step Count                        62  Switch Count  23
      Page Faults                       0
      Page Reclaims                     14145
      Page Swaps                        0
      Voluntary Context Switches        2338
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           8944



NOTE: Due to the presence of CLASS variables, an intercept is implicitly fitted.  R-Square has
      been corrected for the mean.
240
241         * Fit the same model that was used as before, but with party as a covariate.
242         Party needs to be in the class statement, and is the only categorical
243         variable, so it wasn't worth modifying the above macro to use party
244         correctly and I did it manually.;

NOTE: PROCEDURE GLM used (Total process time):
      real time           1.49 seconds
      user cpu time       0.83 seconds
      system cpu time     0.06 seconds
      memory              10236.87k
      OS Memory           70628.00k
      Timestamp           04/25/2022 10:09:52 PM
      Step Count                        63  Switch Count  23
      Page Faults                       0
      Page Reclaims                     13704
      Page Swaps                        0
      Voluntary Context Switches        2877
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           12328


245         PROC GLM DATA = HOME.GRAINS PLOTS = ALL;
246         CLASS GRN PARTY;
247         MODEL LPE = GRN | PARTY;
248         RUN;

249
250         TITLE2 "1866 FULL MODEL";
251         * 1866 FULL MODEL: this model includes all non-correlated predictors that were
252         measured in 1866.;

NOTE: PROCEDURE GLM used (Total process time):
      real time           1.30 seconds
      user cpu time       0.72 seconds
      system cpu time     0.06 seconds
      memory              10358.00k
      OS Memory           71140.00k
      Timestamp           04/25/2022 10:09:53 PM
      Step Count                        64  Switch Count  24
      Page Faults                       0
      Page Reclaims                     13359
      Page Swaps                        0
      Voluntary Context Switches        4414
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           10136


253         PROC MIXED DATA = HOME.GRAINS PLOTS = ALL;
```

```
254         CLASS GRN PARTY;
255         MODEL LPE =
256         GRN HVT PRD INFL PWR YEAR PARTY
257         GRN*HVT GRN*PRD GRN*INFL GRN*PWR GRN*YEAR GRN*PARTY /
258         NOINT SOLUTION
259         ;
260         RUN;

NOTE: 10 observations are not included because of missing values.
NOTE: PROCEDURE MIXED used (Total process time):
      real time           1.90 seconds
      user cpu time       0.93 seconds
      system cpu time     0.11 seconds
      memory              11623.09k
      OS Memory           74212.00k
      Timestamp           04/25/2022 10:09:55 PM
      Step Count                        65  Switch Count  44
      Page Faults                       0
      Page Reclaims                     22547
      Page Swaps                        0
      Voluntary Context Switches        3161
      Involuntary Context Switches      65
      Block Input Operations            0
      Block Output Operations           10544


261
262         TITLE2 "1880 FULL MODEL";
263         * FULL MODEL WITH TEMP (1880 MODEL): this model is the same as the previous
264         model, but also includes the temperature anomaly information. Consequently,
265         it only uses data from 1880 onwards (even less for sorghum).;
266         PROC MIXED DATA = HOME.GRAINS PLOTS = ALL;
267         CLASS PARTY GRN;
268         MODEL LPE =
269         GRN HVT PRD INFL PWR YEAR PARTY TEMP
270         GRN*HVT GRN*PRD GRN*INFL GRN*PWR GRN*YEAR GRN*PARTY TEMP*PARTY/
271         NOINT SOLUTION
272         ;
273         RUN;

NOTE: 52 observations are not included because of missing values.
NOTE: PROCEDURE MIXED used (Total process time):
      real time           1.57 seconds
      user cpu time       0.86 seconds
      system cpu time     0.10 seconds
      memory              10613.28k
      OS Memory           74980.00k
      Timestamp           04/25/2022 10:09:56 PM
      Step Count                        66  Switch Count  44
      Page Faults                       0
      Page Reclaims                     21805
      Page Swaps                        0
      Voluntary Context Switches        3178
      Involuntary Context Switches      5
      Block Input Operations            0
      Block Output Operations           10240


274
275         ************************************************************************;
276         * GLS multiple regression analysis;
277         ************************************************************************;
278
279         * Take the better fitting (by AIC) of the two previous models, and run a model
280         that can account for correlation using generalized least squares.
```

```
281         This model assumes exchangeable correlations between each of the time points.;
282
283         PROC MIXED DATA = HOME.GRAINS PLOTS = ALL;
284         CLASS PARTY GRN;
285         MODEL LPE = HVT PRD INFL PWR YEAR GRN GRN*HVT GRN*PRD /
286         NOINT SOLUTION CHISQ;
287         REPEATED;
288         RUN;

NOTE: 10 observations are not included because of missing values.
NOTE: Convergence criteria met.
NOTE: PROCEDURE MIXED used (Total process time):
      real time           1.37 seconds
      user cpu time       0.75 seconds
      system cpu time     0.08 seconds
      memory              10043.40k
      OS Memory           75232.00k
      Timestamp           04/25/2022 10:09:58 PM
      Step Count                        67  Switch Count  42
      Page Faults                       0
      Page Reclaims                     21132
      Page Swaps                        0
      Voluntary Context Switches        2039
      Involuntary Context Switches      28
      Block Input Operations            0
      Block Output Operations           9520


289
290         ********************************************************************************;
291         * Simple forecasting;
292         ********************************************************************************;
293
294         * Now instead of just using regression models, we can try to fit a more
295         flexible forecasting model using PROC ARIMA.
296         First, we need a time variable that is actually a SAS date, so we create
297         that first.;
298         DATA TS_DAT;
299         SET HOME.GRAINS;
300         T = MDY(1, 1, YEAR);
301         RUN;

NOTE: There were 571 observations read from the data set HOME.GRAINS.
NOTE: The data set WORK.TS_DAT has 571 observations and 17 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              959.09k
      OS Memory           68532.00k
      Timestamp           04/25/2022 10:09:58 PM
      Step Count                        68  Switch Count  2
      Page Faults                       0
      Page Reclaims                     171
      Page Swaps                        0
      Voluntary Context Switches        12
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


302
303         * Next we use the IDENTIFY modeling stage. We check up to 30 lags in the
304         first ARIMA modeling stage, and also explicitly test for stationarity at
305         the first 10 differences using the random walk with drift test. We
```

```
306         also use the SCAN method, which is a heuristic for identifying
307         candidate ARIMA models.;
308         PROC ARIMA DATA = TS_DAT;
309         IDENTIFY VAR = LPE NLAG = 30 SCAN STATIONARITY = (RW = 10);
310         BY GRN;
311         RUN;

NOTE: Interactivity disabled with BY processing.
312
313         * Next we use the ESTIMATE modeling stage. We fit several different ARIMA
314         models to the data in order to see which fits our time series the best,
315         and if any have white noise as the error term.;

NOTE: PROCEDURE ARIMA used (Total process time):
      real time           1.19 seconds
      user cpu time       0.72 seconds
      system cpu time     0.03 seconds
      memory              4831.75k
      OS Memory           70068.00k
      Timestamp           04/25/2022 10:09:59 PM
      Step Count                        69  Switch Count  0
      Page Faults                       1
      Page Reclaims                     2415
      Page Swaps                        0
      Voluntary Context Switches        1373
      Involuntary Context Switches      2
      Block Input Operations            1784
      Block Output Operations           4008


316         PROC ARIMA DATA = TS_DAT;
317         IDENTIFY VAR = LPE;
318         ESTIMATE P = 1;
319         ESTIMATE Q = 1;
320         ESTIMATE P = 1 Q = 1;
321         ESTIMATE P = 2;
322         BY GRN;
323         RUN;

NOTE: Interactivity disabled with BY processing.
324
325         * Finally, we use the best fitting model to make some simple forecasts in
326         the FORECAST modeling stage. We also identify outliers of the best
327         fitting model.;

NOTE: PROCEDURE ARIMA used (Total process time):
      real time           8.70 seconds
      user cpu time       4.60 seconds
      system cpu time     0.60 seconds
      memory              12713.62k
      OS Memory           79892.00k
      Timestamp           04/25/2022 10:10:08 PM
      Step Count                        70  Switch Count  192
      Page Faults                       0
      Page Reclaims                     114881
      Page Swaps                        0
      Voluntary Context Switches        12971
      Involuntary Context Switches      64
      Block Input Operations            0
      Block Output Operations           37408


328         PROC ARIMA DATA = TS_DAT;
329         IDENTIFY VAR = LPE;
330         ESTIMATE P = 1;
```

```
331         OUTLIER;
332         FORECAST LEAD = 10 INTERVAL = YEAR ID = T OUT = GRAIN_FC;
333         BY GRN;
334         RUN;

NOTE: Interactivity disabled with BY processing.
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
362
```