SAS®

# GLOBAL FORUM 2020

MARCH 29 – APRIL 1

USERS PROGRAM

# Incorporating Auxiliary Information into Your Model Using Bayesian Methods in SAS® Econometrics

Matt Simpson, SAS Institute Inc.

Originally from Missouri

Iowa State University PhD in Stat + Econ

University of Missouri postdoc in Stat

Token Bayesian developer in SAS Econometrics (1.5 years)

Default presentation template user

Does not know where his prior comes from

# Bayesian methods enable you to take into account additional information through the prior

*But doing this is not easy*

Solution: Think real hard

…and use some nifty tools

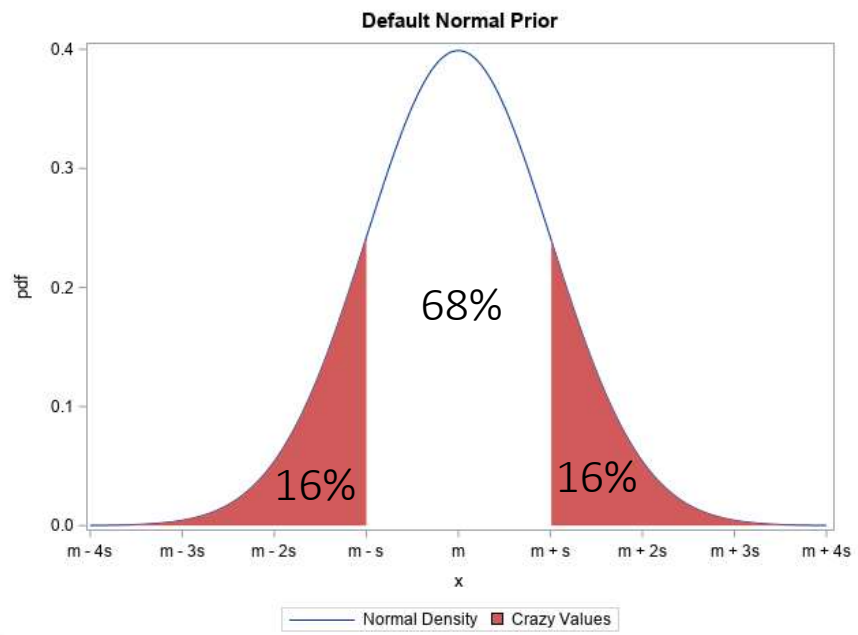
# Start with a Reasonable Default Prior

## Default priors:

- Weakly informative *for questions you care about*
- Spread out, but not too much

## Starting point: $\theta \sim \mathrm{N}(m, s^2)$

- $m =$ the value you expect, or $H_0$
- $m \pm s =$ the most extreme value you think realistically possible



# Default Prior Selection

## In Linear Regression

- Logged covariate in log-log models: $\beta_{\log} \sim \mathrm{N}\left(0,\ s^2_{\log}\right)$
- Standardized covariates: $\beta_{\mathrm{std}} \sim \mathrm{N}\left(0,\ s^2_{\mathrm{std}}\,\mathrm{Var}(y)\right)$
- Class/dummy covariate: $\beta_{\mathrm{class}} \sim \mathrm{N}\left(0,\ s^2_{\mathrm{class}}\,\mathrm{Var}(y)\right)$
- Untransformed covariate: $\beta_{\mathrm{raw}} \sim \mathrm{N}\left(0,\ s^2_{\mathrm{raw}}\frac{\mathrm{Var}(y)}{\mathrm{Var}(x)}\right)$
- Intercept: $\beta_0 \sim \mathrm{N}(0, s_0^2 \max[\mathrm{Var}(\beta)])$
- Error standard deviation: $\sigma \sim \mathrm{N}^{+}\left(0, s_\sigma^2\,\mathrm{Var}(y)\right)$

# Default Prior Selection

## In Linear Regression

Defaults:

- $s_{\log} = s_{\text{std}} = s_{\text{class}} = s_{\text{raw}} = 4$
- $s_0 = 100$
- $s_\sigma = 1$

For more information, see my SAS Global forum paper

- https://github.com/sascommunities/sas-global-forum-2020/tree/master/papers/4311-2020-Simpson

# Implementation in PROC QLIM

- Manually transform the data
- Compute sample summary statistics
- Plug into PROC QLIM:

```
proc qlim data = trucksales_transformed plots = none;
   class area_type;
   model log_sales = area_type log_pop_bachelors log_pop_below_bachelors
      log_median_income log_price log_cost_of_living
      log_mean_precip mean_summer_temp_cs mean_winter_temp_cs;
   bayes seed = 72834 ntu = 100 mintune = 20 maxtune = 20 nmc = 10000
      statistics = (summary interval prior);
   prior intercept ~ normal(mean = 8.88, var = 10000);
   prior log_pop_bachelors log_pop_below_bachelors log_median_income
      log_cost_of_living log_mean_precip log_price ~ normal(mean = 0, var = 16);
   prior mean_summer_temp_cs mean_winter_temp_cs
      area_type_rural area_type_sub ~ normal(mean = 0, var = 7.62);
   prior _sigma ~ normal(mean = 0, var = 0.48);
run;
```

# Semi-automatic Prior Selection

## Two-Step Process

1. Given a data set, produce:
   - Transformed data set
   - CLASS and MODEL statements
   - Prior distribution data set

autoprior_linear.sas

2. Given prior distribution data set,* produce:
   - Prior statements

autoprior_priorstmts.sas

*Can manually create the prior distribution data set

# Example: Default Priors

## autoprior_run.sas

```
%include 'trucksales_data.sas'; /* generate the data set */

/* generate the dataset of all variables to be used in the model
   and their transformation */
data trucksales_vars;
    input variable $20. type $ transform $ response;
    datalines;
sales                numeric log          1
pop_bachelors        numeric log          0
pop_below_bachelors  numeric log          0
median_income        numeric log          0
cost_of_living       numeric log          0
mean_summer_temp     numeric std          0
mean_winter_temp     numeric std          0
mean_precip          numeric log          0
price                numeric log          0
area_type            class   none         0
;
run;
```

```
%let variables = trucksales_vars;
%let dataset = trucksales;
%let log_s = 4;
%let std_s = 4;
%let none_s = 4;
%let class_s = 4;
%let intercept_s = 100;
%let sigma_s = 1;
%let intercept_mean = 0;
%let sigma_mean = 0;
```

# Example: Default Priors

## autoprior_run.sas

```
%include 'autoprior_linear.sas';
```

TABLE: Work.Data_transformed

| area_type | log_sales | log_pop_bachelors | log_pop_below_bachelors | log_median_income | log_cost_of_living | log_mean_precip | log_price | std_mean_summer_temp | std_mean_winte |
|---|---|---|---|---|---|---|---|---|---|
| rural | 5.926926026 | 9.152393412 | 10.61575085 | 10.437639002 | 5.170483995 | 3.258096538 | 10.169690593 | 0.0581498424 | -0.511952108 |
| sub | 5.0937502008 | 9.6066974611 | 10.467493689 | 11.013698977 | 4.8520302639 | 3.258096538 | 10.196157166 | 0.4458154586 | 0.4655220835 |
| sub | 4.9904325868 | 9.3561708202 | | | | | 0501722 | 0.4458154586 | -0.26758356 |
| rural | 5.8861040315 | 9.4983723832 | | | | | 9001991 | 0.6396482667 | -0.511952108 |
| urban | 4.1431347264 | 9.4940897214 | | | | | 7744086 | 0.4458154586 | 0.7098906313 |
| rural | 6.0776422433 | 9.1085291058 | | | | | 8129771 | -1.298679814 | 0.5877063574 |
| sub | 5.0106352941 | 9.469237093 | | | | | 1998248 | 1.221146691 | 1.1986277269 |
| rural | 5.9427993751 | 9.3695638775 | | | | | 0501722 | -0.329515774 | -0.756320655 |
| sub | 5.0369526024 | 9.3287454364 | | | | | 9967119 | -0.329515774 | -0.145399286 |
| sub | 4.8828019226 | 9.1724307989 | | | | | 4246271 | 0.0581498424 | -1.122873477 |

TABLE: Work.Prior

| parameter | distribution | hyper1 | hyper2 |
|---|---|---|---|
| log_pop_bachelors | normal | 0 | 16 |
| log_pop_below_bachelors | normal | 0 | 16 |
| log_median_income | normal | 0 | 16 |
| log_cost_of_living | normal | 0 | 16 |
| std_mean_summer_temp | normal | 0 | 7.6118757141 |
| std_mean_winter_temp | normal | 0 | 7.6118757141 |
| log_mean_precip | normal | 0 | 16 |
| log_price | normal | 0 | 16 |
| area_type_rural | normal | 0 | 7.6118757141 |
| area_type_sub | normal | 0 | 7.6118757141 |
| area_type_urban | normal | 0 | 7.6118757141 |
| intercept | normal | 0 | 160000 |
| _sigma | normal | 0 | 0.4757422321 |

# Example: Default Priors

## autoprior_run.sas

```
/* Creates the prior statements */
/* input: &prior - a macro variable naming a data set
                   containing all prior information */
/* output: &prior_stmts - list of all pric
%include 'autoprior_priorstmt.sas';

/* see the resulting priors */
proc qlim data = data_transformed plots =
   &model /* includes class statement and
   bayes seed = 72834 ntu = 2 mintune = 1
      statistics = prior;
   &prior_stmts
run;
```

**Prior Summaries**

| Parameter | Distribution | Hyper1 | Hyper2 | Min | Max |
|---|---|---|---|---|---|
| Intercept | Normal | 0 | 160000 | -Infty | Infty |
| log_pop_bachelors | Normal | 0 | 16 | -Infty | Infty |
| log_pop_below_bachelors | Normal | 0 | 16 | -Infty | Infty |
| log_median_income | Normal | 0 | 16 | -Infty | Infty |
| log_cost_of_living | Normal | 0 | 16 | -Infty | Infty |
| std_mean_summer_temp | Normal | 0 | 7.611876 | -Infty | Infty |
| std_mean_winter_temp | Normal | 0 | 7.611876 | -Infty | Infty |
| log_mean_precip | Normal | 0 | 16 | -Infty | Infty |
| log_price | Normal | 0 | 16 | -Infty | Infty |
| area_type_rural | Normal | 0 | 7.611876 | -Infty | Infty |
| area_type_sub | Normal | 0 | 7.611876 | -Infty | Infty |
| _Sigma | Normal | 0 | 0.475742 | 1E-11 | Infty |

# Example: Informative Priors

## autoprior_run.sas

```
/* now make an informative prior for price
data new_prior;
     set prior;
     if parameter = 'log_price' then do;
          hyper1 = -1;
          hyper2 = 0.5**2;
          end;
run;

%let prior = new_prior;

%include 'autoprior_priorstmt.sas';

/* see the resulting priors */
proc qlim data = data_transformed plots = r
    &model /* includes class statement and m
    bayes seed = 72834 ntu = 2 mintune = 1 m
       statistics = prior;
    &prior_stmts
run;
```

**Prior Summaries**

| Parameter | Distribution | Hyper1 | Hyper2 | Min | Max |
|---|---|---|---|---|---|
| Intercept | Normal | 0 | 160000 | -Infty | Infty |
| log_pop_bachelors | Normal | 0 | 16 | -Infty | Infty |
| log_pop_below_bachelors | Normal | 0 | 16 | -Infty | Infty |
| log_median_income | Normal | 0 | 16 | -Infty | Infty |
| log_cost_of_living | Normal | 0 | 16 | -Infty | Infty |
| std_mean_summer_temp | Normal | 0 | 7.611876 | -Infty | Infty |
| std_mean_winter_temp | Normal | 0 | 7.611876 | -Infty | Infty |
| log_mean_precip | Normal | 0 | 16 | -Infty | Infty |
| log_price | Normal | -1 | 0.25 | -Infty | Infty |
| area_type_rural | Normal | 0 | 7.611876 | -Infty | Infty |
| area_type_sub | Normal | 0 | 7.611876 | -Infty | Infty |
| _Sigma | Normal | 0 | 0.475742 | 1E-11 | Infty |

# What about Other Models?

## In PROC QLIM or PROC COUNTREG

Several options: See my paper and presentation for prior selection tips in other models

1. Modify autoprior_linear.sas
   - Tailor it to the model you want to fit

2. Create the prior data set yourself
   - Use the data set created by autoprior_linear.sas as a starting point

# Thank you!

Contact Information
Matt.Simpson@sas.com

Code available on Github:

https://github.com/sascommunities/sas-global-forum-2020/tree/master/demos/SD308-Simpson-AuxiliaryInfo


SAS® GLOBAL FORUM 2020