Hans Elliott
EC422 - Economic Forecasting
March 2021

Retail Trends: Forecasting Consumer Spending for Retail Industry Client

Background

Due to the impact of the COVID-19 pandemic, consumers spent less overall in 2020 and shifted their spending to a modified mix of goods and services. Total expenditures and expenditures on services plummeted in early 2020 and then recovered slightly, but remain below pre-pandemic levels. Expenditures on goods outperformed those on services, and although they took a hit early in the pandemic, consumers are spending more on goods now than they were pre-pandemic. Research shows that the government provided unemployment benefits and stimulus packages were integral to reducing the decline in consumer spending. Within consumer spending on goods, certain types of goods outperformed others. Goods often used at home, such as games and other recreational goods, alcoholic beverages, and furnishing, performed well in 2020. Retail related goods - such as clothing, footwear, luggage and personal items - did not perform as well. After recovering from a major decline in early 2020, retail expenditures rose to about their pre-pandemic level by the end of 2020.[1]

Aladangady and Garcia write that, during recessions, "household spending on goods—particularly durables—and housing tends to fall sharply and remain weak for many quarters." Meanwhile, "services spending has generally responded little to business cycles." However, spending has behaved in the opposite way during the current recession, with social distancing due to pandemic concerns playing a key role in a recent shift in the composition of household spending, and low interest rates providing support for housing and durable goods. Services spending remains well below pre-pandemic levels, but spending on durables has spiked higher than pre-pandemic levels, which the authors claim is partially a result of households spending less on services and people spending more time at home. However, this elevated spending might not persist once households reach "their desired stock of durables."[2]

A report released by the New York Federal Reserve in January 2021 shows that, while household expenditures rose modestly over the last quarter of 2020, consumers' spending expectations for a year from now surged. The report shows that "consumers are slowly spending more as the economy reopens" and "people gradually return to work," so consumers are hopeful that their financial situations will continue to improve as the coronavirus pandemic comes under control. The median household expects spending to grow by 3% over the next year, which is 0.8% higher than expected spending was in August 2020. Notably, consumers expect spending on nonessential items to rise by a median 1.6% over the next year, so further increases in spending on retail goods should be expected.[3]

While we can be optimistic about increases in consumer spending as the economy recovers, the pandemic has impacted people in different ways, and many Americans will be focused on repaying debt as their financial situations improve. A recent survey found that millennials, women, Blacks, and Latinos are more likely to have a higher amount of debt than savings when compared to other populations. This is possibly due to the reality that "initial job losses as well ongoing closures and a growing string of business failures have been concentrated in the lower-wage sectors of services industries like travel, restaurants and retail that historically have provided employment to large populations of minority workers." Middle-income Americans also face more debt than savings, as these households did not have the same opportunities as higher-income households, who largely retained employment and, while being forced to spend less, have used their savings to pay off debt or to benefit from higher stock prices and

---

[1]https://www.kansascityfed.org/research/economic-bulletin/consumer-spending-declines-shifts-in-response-to-the-pandemic/

[2]https://www.federalreserve.gov/econres/notes/feds-notes/the-unusual-composition-of-demand-during-the-pandemic-20210114.htm

[3]https://www.reuters.com/article/us-usa-fed-consumer-expectations/u-s-consumers-expect-spending-to-rebound-strongly-in-one-year-ny-fed-survey-finds-idUSKBN29U1UR

housing values. Overall, these results suggest that higher-income consumers will likely increase spending on nonessential goods, including retail goods, significantly over the next year, but many other households will be slow to increase spending on these goods.[4]

Methodology

To evaluate the forecast of Retial Trend LLC's previous consultant, I reconstructed their ARMA(1,1) with a trend model that was used to forecast *personal consumption expenditures for goods* (FRED Code: *DGDSRC1*). I first examined Akaike information criterion (AIC) and Schwarz information criterion (SIC) scores for an array of ARMA (Autoregressive Moving Average) models that were based on the residuals of a linear regression, which removed the trend from the series. The lowest AIC score was for an ARMA(1,1) model, and the lowest SIC score was for an ARMA(3,3) model, but that model does not converge, so the consultant was right to pick an ARMA(1,1) model. I then evaluated the ARMA(1,1) model and the consultants forecast, which I recreated (see Results).

I chose a new model based on the discovery that the series (I used the log of *DGDSRC1*) has a unit root, meaning that it is nonstationary. I then tested the first difference of the series for a unit root and found stationarity. This finding inspired me to use an ARIMA (Autoregressive integrated moving average) model. I examined AIC and SIC scores for an array of ARMA models based on the first difference of the series, and found that the lowest AIC score was for an ARMA(2,2) model, and the lowest SIC score was for an ARMA(4,4) model. The ARIMA(4,1,4) model does not converge, so I settled with an ARIMA(2,1,2) model to create my improved forecast (see Results for an evaluation of the model).

Once I had both the recreated forecast and my new forecast, I compared them visually and by using the Diebold-Mariano (DM) test. I chose the DM test over the Granger-Newbold test because, after testing the errors for autocorrelation, I found they had Ljung Box p-values of 0.00 (so they do show autocorrelation), which violates the GN test's assumption that the errors are uncorrelated. As seen in the Results section, the forecast evaluations show that my new forecast is preferred to the old forecast. Therefore, I used the ARIMA(2,1,2) model to create a 5 year forecast from 2021:2 to 2026:2. I first recreated the model for the period 1984:1 to 2021:1, and then made the forecast.

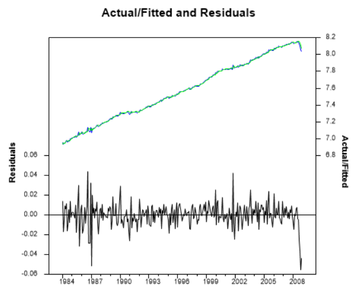
Results

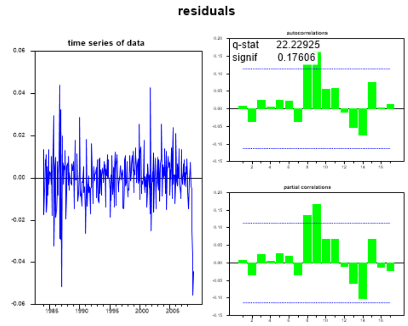
The Old Model and Forecast (ARMA (1,1) model)

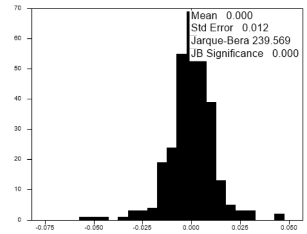
Below are the results of the previous consultant's model, an ARMA(1,1) model with a trend over the period 1984:1 to 2008:12. The model has a centered R squared of 0.99. Its AR and MA roots are below 1, although its AR root is close (0.998). The residuals have a Ljung-Box p-value of 0.17, so we fail to reject the null hypothesis that there is no evidence of time dependency. Looking at the residuals, they appear to be white noise. However, the residuals have a Jarque-Bera significance level of 0.00, so we must reject that test's null hypothesis, meaning the residuals are not independently and normally distributed. There appears to be some outlier observations and a large spike at the center of the distribution.

	Variable	Coeff	Std Error	T-Stat	Signif
1.	CONSTANT	13.3356114	201.3353489	0.06624	0.94723480
2.	AR{1}	0.9982833	0.0244986	40.74860	0.00000000
3.	MA{1}	-0.3177058	0.0632089	-5.02628	0.00000087
4.	TREND	-0.0020751	0.0881292	-0.02355	0.98123076

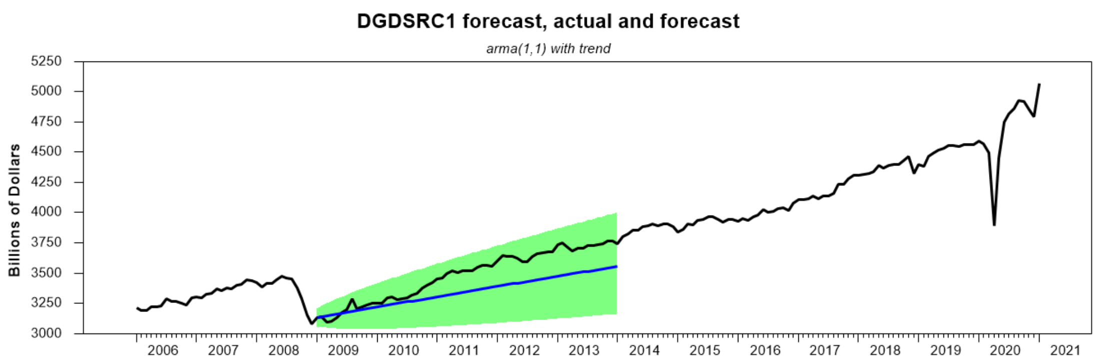
[4]https://www.nbcnews.com/business/economy/americans-are-saving-more-during-pandemic-there-s-still-huge-n1257252







The recreated forecast using the ARMA(1,1) model is shown below. Just as the client explained, this model dramatically under forecasts personal consumption expenditures on goods.



Looking at the series, it appears that the log of *DGDSRC1* may be nonstationary with a trend. Therefore, the previous consultant produced a bad forecast because they used a deterministic trend to model a stochastic series.

Testing the series for a unit root using the Dickey-Fuller test with a trend, I found that the series is nonstationary (the T-Statistic is greater than the critical values). I then tested the first difference of the series with a DF test with no constant or trend and found that it was stationary (as seen on the right).

Dickey-Fuller Unit Root Test, Series LOG_SPEND

Sig Level	Crit Value
1%(**)	-3.99236
5%(*)	-3.42635
10%	-3.13610
T-Statistic	-0.15149

Dickey-Fuller Unit Root Test, Series DIFF_LSPEND

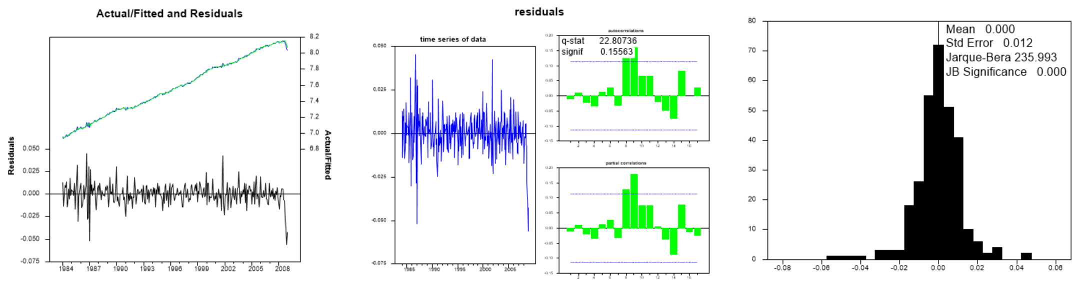
Sig Level	Crit Value
1%(**)	-2.57254
5%(*)	-1.94064
10%	-1.61621
T-Statistic	-4.18545**

Due to this finding, I decided to use an ARIMA model with one difference to create a better forecast.

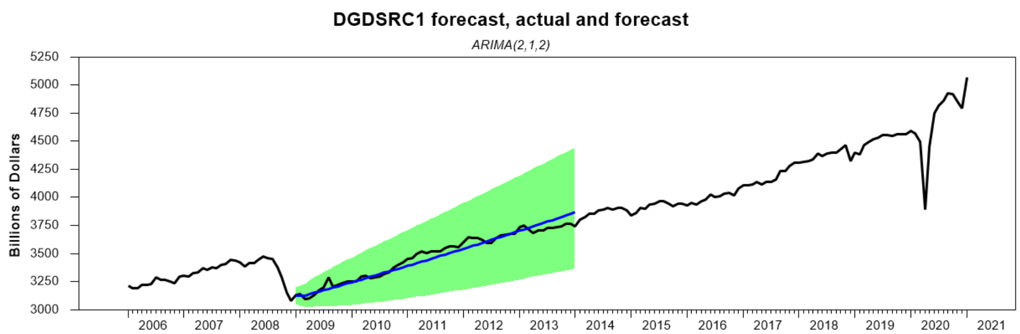
A New Model and Forecast (ARIMA(2,1,2) model)

As detailed in the Methodology section, my model selection process led me to an ARIMA(2,1,2) model. This model has a centered R squared of 0.99, and its roots are well below 1. The residuals have a Ljung-Box p-value of 0.15, so we fail to reject the null hypothesis that there is no evidence of time dependency. Looking at these residuals, they also appear to be white noise. These residuals also have a Jarque-Bera significance level of 0.00, so we must again reject the null hypothesis, meaning the residuals are not independently and normally distributed.

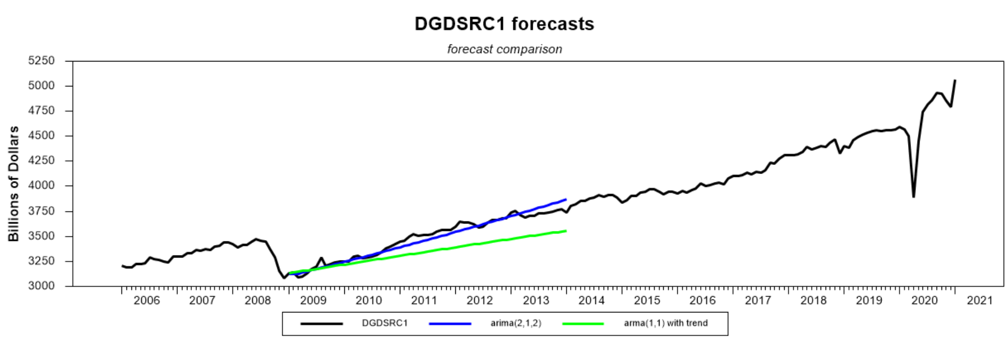
	Variable	Coeff	Std Error	T-Stat	Signif
1.	CONSTANT	0.003661449	0.000522886	7.00238	0.00000000
2.	AR{1}	0.239084530	0.510194578	0.46861	0.63969136
3.	AR{2}	-0.292741151	0.185226474	-1.58045	0.11507546
4.	MA{1}	-0.525020715	0.510805394	-1.02783	0.30487217
5.	MA{2}	0.328441235	0.253538297	1.29543	0.19618455



Using this model, I created a new forecast over the period 2009:1 through 2014:1, which can be seen below.



Model Comparison

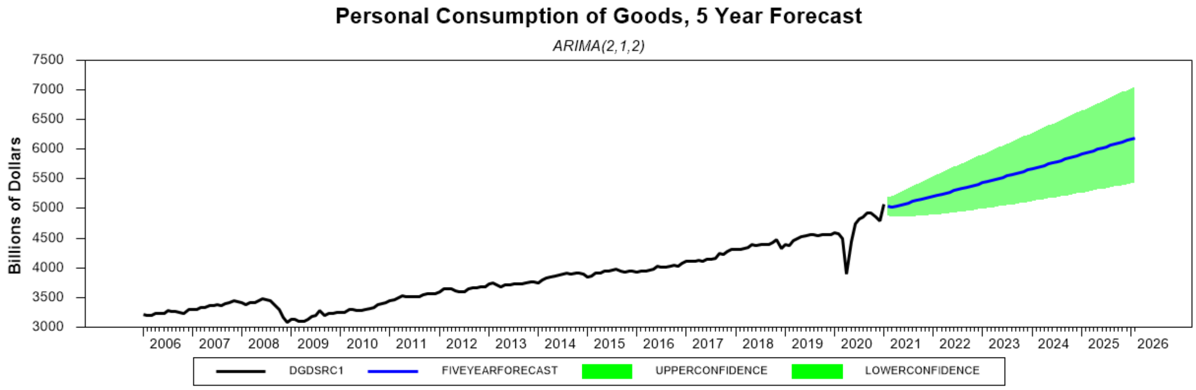


Visually, it is evident that the new forecast is more accurate than the old forecast. It does not under forecast or over forecast the series, but rather follows the trend in the actual data quite closely. The root mean square error of the new forecast is only 0.013 compared to the first forecast's error of 0.045. In addition, the Diebold-Mariano test produced the following results, which indicate to reject the old forecast (fore1) in favor of our new forecast (fore2):

```
Diebold-Mariano Forecast Comparison Test
   Forecasts of LOG_SPEND over 2009:01 to 2014:01
   Forecast     MSE       Test Stat P(DM>x)
   FORE1     0.00209088     3.9561 0.00004
   FORE2     0.00019125    -3.9561 0.99996
```

Five Year Out-of-Sample Forecast

Given the relative success of the new model in forecasting consumer spending on goods, I used the ARIMA(2,1,2) model to create a forecast for the next five years of personal consumption of goods.



Conclusion

As demonstrated by the results, the previous consultant produced a bad forecast because they modeled this stochastic series as a deterministic trend. A deterministic trend assumes the slope of a series will not change over time, but the results of the DF test demonstrate that the series is nonstationary. Modeling consumer spending on goods as a stochastic trend with an ARIMA model allowed for an improved forecast because stochastic trends can change over time. Once an ARIMA model was employed, the corresponding forecast was visually and statistically more accurate than the forecast that used a deterministic trend.

The forecast for the next five years predicts a continuous increase in personal consumption expenditure on goods, rising steadily from about 5,038 billion dollars in February 2021 to about 6,191 billion dollars in February 2026. This may be a fairly accurate forecast as pandemic-related restrictions continue to lift and the reopening of the economy improves consumers' financial situations. The mean standard error of the forecast is 349.5 billion dollars and looking at the confidence intervals there is certainly room for error in this forecast. However, if economic recovery continues on its expected course we can be fairly confident that consumer spending will steadily increase as the model predicts.

Even if spending on all goods follows the forecasted trend, it may be risky to use this forecast as an indicator of retail industry trends. As discussed in the Background section, spending on retail goods has not had the same response to the pandemic as spending on other goods. Rather, spending on retail goods has recovered slower from the pandemic's initial toll and currently sits around its pre-pandemic level. Other signs, such as the "retail apocalypse," suggest that forecasts for the retail industry specifically should be more pessimistic about consumer spending. The retail apocalypse, a phenomenon where an increasing number of brick-and-mortar retail stores are closing down, presents a major struggle to the retail industry and was likely exacerbated during the recent economic downturn. However, it is also possible that higher-income households can afford the extra costs of e-commerce (costs from technology or shipping for example). Considering that higher-income households make up a large portion of the increasing consumer spending on nonessential goods, the retail industry might see a boost in consumer spending that evolves into a gradual increase similar to that depicted by the forecast. Given all of these factors and the added economic uncertainty caused by the pandemic, it is difficult to put full faith in this forecast as an accurate basis for retail industry trends. However, it does seem that this is a strong model for overall consumption expenditures in the economy, and certainly an improvement compared to the previous consultant's model.

This analysis was completed using RATS (Regression Analysis of Time Series) software. The code is below:

```
**** EC422: Hans Elliott ****
* Load in data
calendar(m) 1900:1
all 2026:12
data(format=fred) * * DGDSRC1
set trend = t

set log_spend = log(DGDSRC1)

* ARMA Forecast model selection:
linreg log_spend 1984:1 2008:12 resids1
#constant trend
@autocorr(header="Residuals") resids1
@bjautofit(pmax=5, qmax=5, crit=sbc) resids1
@bjautofit(pmax=5, qmax=5, crit=aic) resids1
   *AIC: ARMA(3,3) and SIC: ARMA(1,1)
   *strong AR component w one-sided decay in lag autocorrelation. partial autos suggest more than 1 lag

* Original ARMA(1,1) model and forecast 2009:1 2014:1:
*
boxjenk(constant,ar=1,ma=1,define=oldmod,regressors) log_spend 1984:1 2008:12 resids2
#trend
@armaroots(equation=oldmod) log_spend
@regactfit
@autocorr(header="residuals") resids2 1984:1 2008:12
@histogram2(stats,counts) resids2 1984:1 2008:12
   *2/4 sig cos, roots<1 but one is close, no residual autocorrelation, resids not normal
*forecast:
uforecast(equation=oldmod,nostatic,stderrs=stderrs1,from=2009:1,to=2014:1) fore1
   set forupper 2009:1 2014:1 = fore1+1.96*stderrs1
   set forlower 2009:1 2014:1 = fore1-1.96*stderrs1

   * convert forecast back to levels for graph (un-log)
set fore1a = EXP(fore1)
set foruppera = EXP(forupper)
set forlowera = EXP(forlower)

graph(style=line,vlabel="Billions of Dollars",header="DGDSRC1 forecast, actual and forecast", $
   subheader="arma(1,1) with trend",overaly=fan, ovcount=2, ovsame) 4
# DGDSRC1 2006:1 *
# fore1a * *
# foruppera * * 3
# forlowera * * 3

set error1 = log_spend - fore1
statistics error1
   *mean forecast error of 0.038
****************************************************************************
****************************************************************************
*
* Create a better model and forecast for the client:
```

```
*
  * Looking at log_spend, looks like it could be nonstationary with a trend
@dfunit(det=trend,method=aic) log_spend 1984:1 2008:12
   *indeed series has a unit root according to the df test (T-stat > critical values)

set diff_lspend = log_spend - log_spend{1}
   @dfunit(det=none,method=aic) diff_lspend 1984:1 2008:12
   * no second unit root. so use diffs = 1
*
* We will use an ARIMA model to create a better forecast
* Selection:
@autocorr(header="diff_lspend") diff_lspend 1984:1 2008:12
   *definitely a significant MA component based on how the autocorrs flip

@bjautofit(pmax=5, qmax=5, crit=sbc) diff_lspend 1984:1 2008:12
@bjautofit(pmax=5, qmax=5, crit=aic) diff_lspend 1984:1 2008:12
   * AIC: ARMA(4,4), SIC: ARMA(2,2)

   * try ARIMA(4,1,4) <- no convergence
 *ARIMA (2,1,2)
boxjenk(constant, ar=2, ma=2, diffs=1, define=modA) log_spend 1984:1 2008:12 resids3
@armaroots(equation=modA) diff_lspend
@regactfit
@autocorr(header="residuals") resids3 1984:1 2008:12
@histogram2(stats,counts) resids3 1984:1 2008:12
      *roots<1, no residual autocorrelation but resids not normal

* Forecast using ARIMA model
uforecast(equation=modA,stderrs=stderrs2,from=2009:1,to=2014:1) fore2
   set forupper2 2009:1 2014:1 = fore2+1.96*stderrs2
   set forlower2 2009:1 2014:1 = fore2-1.96*stderrs2

set fore2a = EXP(fore2)
set forupper2a = EXP(forupper2)
set forlower2a = EXP(forlower2)

graph(style=line,vlabel="Billions of Dollars",header="DGDSRC1 forecast, actual and forecast", $
   subheader="ARIMA(2,1,2)",overaly=fan, ovcount=2, ovsame) 4
# DGDSRC1 2006:1 *
# fore2a * *
# forupper2a * * 3
# forlower2a * * 3
   * much better forecast
set error2 = log_spend - fore2
statistics error2
   *mean error of 0.002 (compared to 0.038 from old forecast).
*
****************************************************************************
* Compare the forecasts
graph(style=line,vlabel="Billions of Dollars",key=below,header="DGDSRC1 forecasts", $
          klabel=||"DGDSRC1"|"arima(2,1,2)"|"arma(1,1) with trend"||,subheader="forecast comparison") 3
```

```
# DGDSRC1 2006:1 *
# fore2a * *
# fore1a * *
  * visually a much better forecast

* Forecast Comparisons
@gnewbold log_spend  fore1 fore2
    * says reject first forecast in favor of second
@autocorr error1
@autocorr error2
  * there is autocorrelation, so this test doesn't really work

@dmariano(lags=6,lwindow=newey) log_spend  fore1 fore2
  *sig. lvl says reject fore1 in favor of fore2
***************************************************************************
***************************************************************************
* 5 year forecast using ARIMA(2,1,2) model

boxjenk(constant, ar=2, ma=2, diffs=1, define=arima) log_spend 1984:1 2021:1 resids4
@armaroots(equation=modB) diff_lspend
@regactfit
@autocorr(header="residuals") resids4 1984:1 2021:1
@histogram2(stats,counts) resids4 1984:1 2021:1

set forstart = 2021:2
set forend = 2026:2

uforecast(equation=arima,from=2021:2,to=2026:2,stderrs=stderrs3) fore3
  set forupper3 2021:2 2026:2 = fore3+1.96*(stderrs3)
  set forlower3 2021:2 2026:2 = fore3-1.96*(stderrs3)

*convert back to level
set FiveYearForecast = EXP(fore3)
set UpperConfidence = EXP(forupper3)
set LowerConfidence = EXP(forlower3)


graph(style=line,vlabel="Billions of Dollars",header="Personal Consumption of Goods, 5 Year Forecast",
$
  subheader="ARIMA(2,1,2)",overaly=fan, ovcount=2, ovsame,key=below) 4
# DGDSRC1 2006:1 *
# FiveYearForecast * *
# UpperConfidence * * 3
# LowerConfidence * * 3

statistics FiveYearForecast
@uforeerrors(theil) fore1 log_spend
@uforeerrors(theil) fore2 log_spend
```