Predicting the US Equity Premium

Predicting the US equity premium.

- Intro
- How to do analysis in Goyal and Welch (2008)
- Using the pictures in Goyal and Welch to understand other predictability studies:
  - Cooper and Priestley (2008)

Introduction

There is a large literature on the predictability of (US) market indices, essentially asking:
Can we predict the equity market premium?
The literature is macro-finance in tone, trying to predict aggregate stock market indices, not individual stocks.

Is the equity market premium predictable? ctd

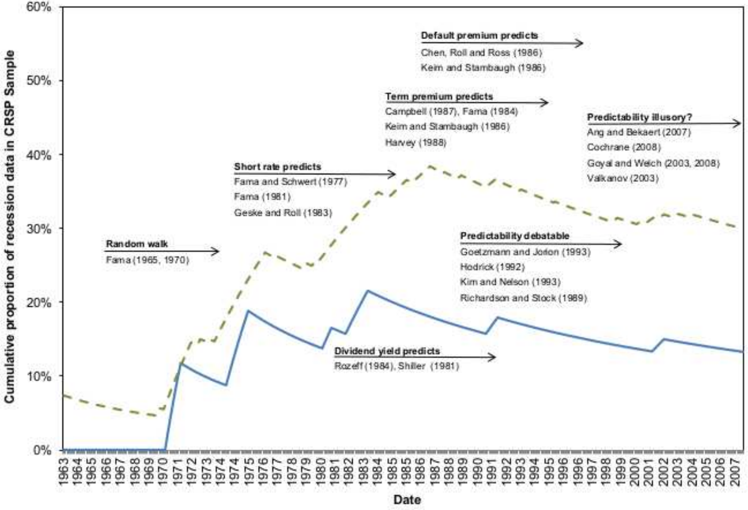


Fig. 2. The time-series of predictability research. The literature on stock return predictability follows closely the availability of recession data as a cumulative proportion of the total data in CRSP which originally started in 1962. Shown are the percentages of recession data as a percentage of the available data at a given date, as measured by NBER (solid line) and RSVAR (dashed line) dates. Both the NBER and RSVAR samples show similar profiles, although RSVAR recession probabilities represent a much larger proportion of the data. Many seminal, and first, papers on return predictability were published just after the peaking of the proportion of recession data to total available data in 1985 and are followed by a decline in the proportion of recession data thereafter. The citations are representative for expository purposes and are not intended to be indicative of initial research, nor a

Replicating Goyal Welch

Will replicate (some of) the analysis of Goyal and Welch (2008).
Their data is readily available, allows for comparison.
Use annual data.
It is available from the RFS web cite, or from Amit Goyal's web page. (There is also a set of updated data series.)

Replicating Goyal Welch

The following reads the data into zoo series

```
library(zoo)
dataGoyalWelchAnnual
    <- read.table("../data/PredictorData_annually.csv",
                 header=TRUE,sep=",",na.strings=c("NaN"))

dataGoyalWelchAnnual
     <- zoo(dataGoyalWelchAnnual[,2:ncol(dataGoyalWelchAnnu
            order.by=dataGoyalWelchAnnual[,1])
```

Replicating Goyal Welch

This is an overview of the data.

```
> head(dataGoyalWelchAnnual)
     Index  D12  E12 b.m tbl AAA BAA lty cay ntis      Rfree infl eqis
1871  4.74 0.26 0.40  NA  NA  NA  NA  NA  NA   NA 0.05733672   NA   NA
1872  5.07 0.30 0.43  NA  NA  NA  NA  NA  NA   NA 0.07189866   NA   NA
1873  4.42 0.33 0.46  NA  NA  NA  NA  NA  NA   NA 0.08661904   NA   NA
1874  4.54 0.33 0.46  NA  NA  NA  NA  NA  NA   NA 0.04854423   NA   NA
1875  4.37 0.30 0.36  NA  NA  NA  NA  NA  NA   NA 0.04380927   NA   NA
1876  3.58 0.30 0.28  NA  NA  NA  NA  NA  NA   NA 0.04106633   NA   NA
     corpr svar csp ik CRSP_SPvw CRSP_SPvwx
1871    NA   NA  NA NA        NA         NA
1872    NA   NA  NA NA        NA         NA
1873    NA   NA  NA NA        NA         NA
1874    NA   NA  NA NA        NA         NA
1875    NA   NA  NA NA        NA         NA
1876    NA   NA  NA NA        NA         NA
```

It is not obvious from this exactly what series is what.
in particular it is not obvious that Index is an index of stock market prices.
Doing the first few figures, those involving dividends and earnings.

```
Sp <- dataGoyalWelchAnnual$Index
D12 <- dataGoyalWelchAnnual$D12
E12 <- dataGoyalWelchAnnual$E12
rf  <- dataGoyalWelchAnnual$Rfree
Rf  <- 1+rf
Rm <- log(Sp+D12)-lag(log(Sp),-1)

eRm <- na.omit(Rm-log(Rf))
dp  <- na.omit(log(D12)-log(Sp))
dy  <- na.omit(log(D12)-lag(log(Sp),-1))
ep  <- na.omit(log(E12)-log(Sp))
```

Goyal and Welch are concerned with comparing the outcome of a prediction exercise

$$er_{m,t} = a + bpred_{t-1} + e_t$$

where $pred_{t-1}$ is some variable thought to predict the equity market premium,
to a "naive" forecast using just the historical mean of the equity market premium.
The metric is the difference in aggregated squared prediction errors.

Their Goal: Produce a picture like the following:

Which is the difference in squared forecast errors between a forecast using a predictive variable
(in this case dividend price ratio)

Function doing In Sample calculation

```
library(dyn)

in_sample_diff_calc_annual <- function(predictor){
                                  # for annual data start when having 21 years
  first_obs <- 21
  data <- merge(eRm,predictor,all=FALSE)
  names(data) <- c("eRm","predictor")
  demeaned2 <- zoo((as.numeric(data$eRm - mean(data$eRm))^2),
                   order.by=index(data$eRm))

  regr <- dyn$lm(data$eRm ~lag(data$predictor,-1))
  res2 <- as.numeric(regr$residuals^2)
  res2 <- zoo(res2,order.by=index(data$eRm)[-1])
                                  # there is one less residual than mean differe
  tmp <- merge(demeaned2,res2,all=FALSE)
  diff <- cumsum(tmp$demeaned2)-cumsum(tmp$res2)
  diff <- diff - as.numeric(diff[first_obs])
                                  # to align at zero on the first oos observation
  return (diff)
}
```

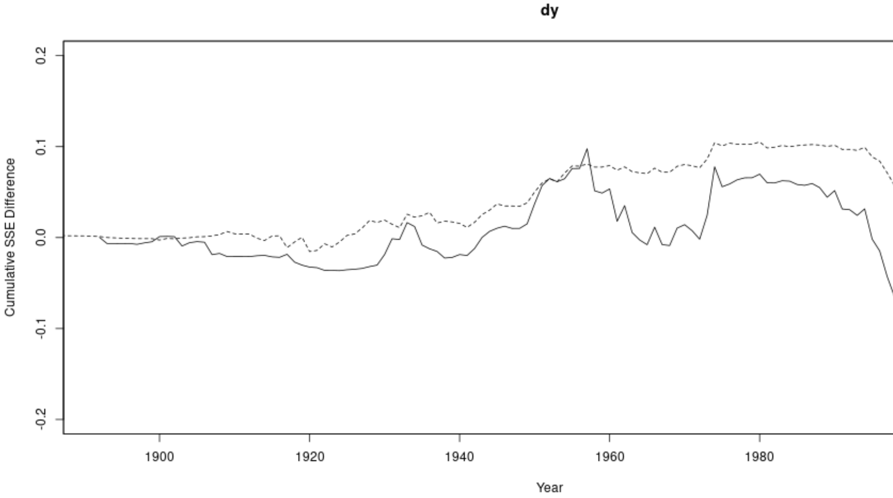
Function doing Out Of Sample calculation:

```
out_of_sample_diff_calc_annual <- function (pred) {
                                       # for annual data start when having 21 y
  first_obs <- 21
  data <- merge(eRm,pred,all=FALSE)
  names(data)<-c("eRm","pred")
  head(data)
  se_NULL <- NULL
  se_ALT <- NULL
  n <- length(data$eRm)
  for (t in first_obs:n){
    se0 <- (data$eRm[t]-mean(data$eRm[1:(t-1)]))^2
    se_NULL <- rbind(se_NULL,zoo(se0,index(data$eRm)[t]))

    prem <- data$eRm[2:(t-1)]
    pred <- data$pred[1:(t-2)]
    regr <- lm(prem~pred)
    npred <- data.frame(pred <- data$pred[t-1])
    pr <- predict.lm(regr, npred)
    se1 <-( data$eRm[t] - pr)^2
    se_ALT <- rbind(se_ALT,zoo(se1,index(data$eRm)[t]))
  }
  cse_diff <- cumsum(se_NULL)-cumsum(se_ALT)
```

Doing the calculation for dp as a predictor (the previous picture)

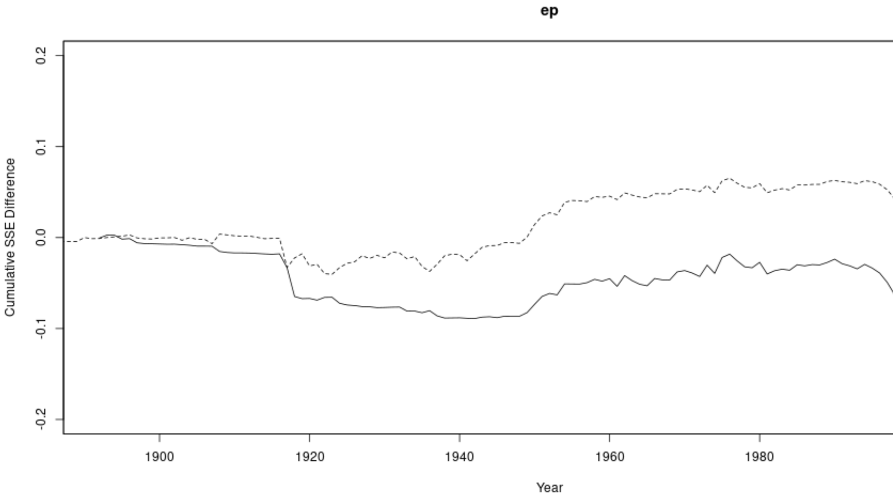
```
is_diff<- in_sample_diff_calc(dp)
oos_diff<- oos_calculation(dp)

postscript("../R_plots/annual_prediction_performance_dp.eps
           horizontal=FALSE,width=10,height=5)
plot(oos_diff,ylim=c(-0.2,0.2),main="dp",xlab="Year",
      ylab="Cumulative SSE Difference",type="l")
lines(is_diff,ylim=c(-0.2,0.2),type="l",lty=2)
dev.off()
```

Using dy



dy
Cumulative SSE Difference
0.2
0.1
0.0
-0.1
-0.2
1900
1920
1940
1960
1980
Year

Using ep


ep
Cumulative SSE Difference
0.2
0.1
0.0
-0.1
-0.2
1900
1920
1940
1960
1980
Year

Replicating Cooper and Priestley (2008)

Show how constructing the similar pictures to Goyal and Welch can add to understanding.
Look at a similar article
Cooper and Priestley (2008):
output gap, a measure of the difference between the *capacity* for output relative to *actual* output
predicts equity market premium.
Will replicate (and extend) their results.

Replicating Cooper and Priestley (2008)

Construct several measures of Output gap

- Monthly date on industrial producion:
  y_t – log of the industrial production index.
  Measure the output gaps as deviations from trends.
  - The residual in a quadratic trend regression

    $$y_t = a + b \cdot t + c \cdot t^2 + v_t$$

  - The residual in a linear trend regression

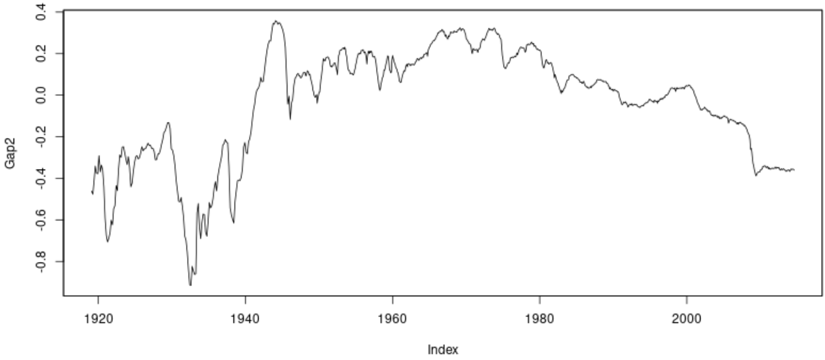
    $$y_t = a + b \cdot t + v_t$$

- Data on GDP.
  Subtracts the actual ex post GDP from an estimated of the *potential* GDP for the US estimated by the Congressional Budget Office.

We download data from FRED, the data service of the St. Louis Federal Reserve. The industrial production (INDPR) is a monthly index. The GDP (GDP) and the Potential GDP (NGDPPOT) are both quarterly series.
The three estimated output gap series are shown in next figures.

Gap1

Gap2



Gap CBO

As an estimate of the equity risk premium we use the monthly series RMRF provided by Ken French.
Quarterly premia are calculated by adding the monthly premia.
To test for predictability we calculate an *in sample* predictive regression

$$eR_m = \alpha + \beta Gap_{t-2}$$

Note that one lags the predictive variable two periods, to make sure it is observed at the time the forecast is made.

Doing first regression

```
lnIP <- log(mIndProd)
lnIP <- window(lnIP,start=c(1947,11))
t <- 1:length(lnIP)
t2 <- t^2
regr <- lm(lnIP ~ t + t2)
Gap1 <- regr$residuals
data <- merge(eRm,lag(Gap1,-2),all=FALSE)
EqtyPrem <- data$eRm
gap1 <- data[,2]
regr1 <- lm(EqtyPrem ~ gap1)
```

In sample regressions of predictability

	Dependent variable:		
	EqtyPrem		
	(1)	(2)	(3)
gap1	-3.664^{***}		
	(1.232)		
gap2		-1.832^{**}	
		(0.716)	
GapCBO			-0.004^{**}
			(0.002)
Constant	0.704^{***}	0.748^{***}	1.552^{***}
	(0.162)	(0.163)	(0.546)
Observations	926	926	258
Adjusted R^2	0.008	0.006	0.013

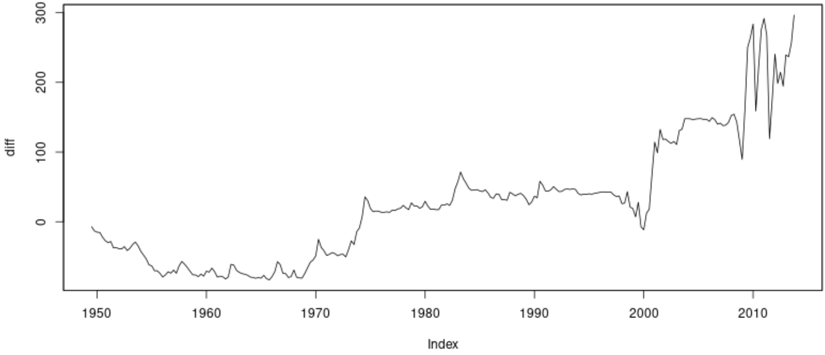
Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

For all three estimates of Gap we find significant in-sample predictability.
To gain some understanding of what is driving the results, we use the approach of Goyal and Welch (2008), comparing the forecasting of the equity premium using this variable with a simple estimate (historical mean).
Calculates the *cumulative squared difference* of the prediction errors, and takes the difference.
Following figures shows the results.
Note that such plots were not done in the original article.

Predictability gain over simple mean (in sample)
Panel A: Gap1

Predictability gain over simple mean (in sample)
Panel B: Gap2

Predictability gain over simple mean (in sample) Panel C: Gap CBO



Ilan Cooper and Richard Priestley. Time-varying risk premiums and the output gap. *Review of Financial Studies*, 2008.

Amit Goyal and Ivo Welch. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008.

Sam James Henkel, J Spencer Martin, and Federico Nardari. Time-varying short-horizon precitability. *Journal of Financial Economics*, 99(3):560–580, 2011.