**Regression**

**Daily**

Significant correlation in 2020, 1990 and 1981:

* 2020: positive - gold price rises 1.3% as s&p500 rises 1%.Model accounts for 30% of change in gold that's explained by change in S&P500.
* 1990: negative - gold price falls 0.29% as s&p500 rises 1%. Model accounts for 7% of change in gold that's explained by change in S&P500.
* 1981: positive - gold price rises 0.41% as s&p500 rises 1%. Model accounts for 5% of change in gold that's explained by change in S&P500.

**Weekly**

Significant correlation sustained in 1990 and 1981. Correlation lost in 2020:

* 2020: No relationship between gold and S&P500 returns when using monthly returns.
* 1990: Negative relationship is sustained, stronger, and a better fit - gold price falls 0.45% as s&p500 rises 1%. Model accounts for 24% of change in gold that's explained by change in S&P500.
* 1981: Positive relationship is sustained, stronger, and a better fit - gold price rises 0.99% as s&p500 rises 1%. Model accounts for 26% of change in gold that's explained by change in S&P500.

**Monthly**

Significant correlation lost in 1990 and 1981. New signficant correlation lost in 1980:

* 1990, 1981: No relationship between gold and S&P500 returns when using monthly returns.
* 1980: positive - gold price rises 2.24% as s&p500 rises 1%. Model accounts for 86% of change in gold that's explained by change in S&P500.

**Prices**

For the S&P500, the biggest variance in price during a single recessionary period was observed in 2007, with a difference of 124% between the maximum and minimum prices reached during this period. The smallest variance was observed in 1980, 21%

For gold, the biggest variance in price during a single recessionary period was observed in 1973, with a difference of 117% between the maximum and minimum prices reached during this period. The smallest variance was observed in 1969, 12%

The biggest difference in change occurred in 2007.

**Returns**

In five of the recessionary periods, an inverse relationship appeared to hold between gold and S&P500 returns: gold experienced positive returns while S&P500 returns were negative (1969, 1973, 2001, 2007 and 2020

This relationship was violated in 1980, 1981, and 1990, where both gold and S&P500 experienced positive returns overall.

The size of the returns was highly variable, ranging from positive returns of +1% and +87% for gold, and returns of -38% to +9% for S&P500. Gold returns were positive during all periods.

**Correlation analysis**

Significant correlation in 2020, 1990 and 1981:

“In 2020 and 1981, gold returns using Pearson’s correlation analysis were positively correlated with S&P500 returns (*r* = 0.55 in 2020, r = 0.22 in 1981, *p* < 0.001)”

“In 1990, gold returns using Pearson’s correlation analysis were negatively correlated with S&P500 returns (*r* = -0.25 in 2020, r = *p* < 0.001)”

**Explanation**

**Many variables determine price of gold**

Relationship between gold and S&P500 was only ever at most significant for three periods (daily returns), and otherwise uncorrelated. Even when relationship was significant, the weakness of regression model and statistical results suggest that many influential variables were missing from my analysis. These could include bonds, etc.

**Returns are random**

The theory of random walk states that returns are random and cannot be used to predict values. This appears to be true for the majority of the results, with no correlations found between the returns. However, for two of the recessionary periods studied (1990 and 1981), both daily and weekly returns returned significant relationships in the same direction.

**Inverse relationships**

Despite the commonly held belief that gold and stocks should be inversely correlated, the data show mixed results. At times, it does appear that as stock prices fall, gold prices increase (1990). However, a significant positive relationship was observed for two periods (1981 and 2020), contradicting this belief.

**Economic recession vs financial crisis**

This project deals with financial data during periods of economic recessions. Economic recessions are characterized by long-term trends (), while financial crisis occur abruptly and tend to last days to weeks (sustained negative returns).

Financial crises tend to precede economic recessions. The classification of an economic recession may come months after the most dramatic drop in the stock market.

Breakpoint analysis could be used to identify where in the recessionary period the financial data changes direction to narrow down the time series within which to evaluate the relationship between gold and S&P500 to return more meaningful results.

Normality - Shapiro-Wilk’s test.

Linearity – Harvey-Collier

Homoscedasticity - Breusch-Pagan

Serial correlation - Durbin-Watson

<https://cran.r-project.org/web/packages/lmtest/lmtest.pdf>

Shapiro-Wilk

shapiro.test()

package “lmtest”

Durbin-Watson, autocorrelation

dwtest()

Breusch-Godfrey, serial correlation

bgtest()

Breusch-Pagan, heteroskedasticity

bptest()

Goldfeld-Quandt, heteroskedasticity

gqtest()

Harrison-McCabe, heteroskedasticity

hmctest()

Harvey-Collier, linearity

harvtest()

Rainbow test, linearity

raintest()

**1. Shapiro-Wilk**

**shapiro.test()**

> shapiro.test(X1969$`Daily gold returns`)

Shapiro-Wilk normality test

**DAILY**

**1969: W = 0.81226, p-value = 2.367e-16**

**1973: W = 0.96033, p-value = 4.225e-08**

**1980: W = 0.97738, p-value = 0.02756**

**1981: W = 0.96229, p-value = 7.829e-08**

**1990: W = 0.94413, p-value = 2.458e-06**

**2001: W = 0.80372, p-value = 4.579e-14**

**2007: W = 0.97125, p-value = 5.78e-07**

2020: W = 0.96546, p-value = 0.2188

**WEEKLY**  
1969: W = 0.98346, p-value = 0.7712

1973: W = 0.96973, p-value = 0.1171

1980: W = 0.97758, p-value = 0.8475

**1981: W = 0.92035, p-value = 0.0005143**

1990: W = 0.95669, p-value = 0.2225

**2001: W = 0.84967, p-value = 0.0004081**

2007: W = 0.99087, p-value = 0.8861

2020: W = 0.92918, p-value = 0.5087  
  
**MONTHLY**

1969: W = 0.9765, p-value = 0.9436

1973: W = 0.94625, p-value = 0.4328

1980: W = 0.91034, p-value = 0.4386

1981: W = 0.94258, p-value = 0.3818

1990: W = 0.92136, p-value = 0.441

2001: W = 0.94748, p-value = 0.6858

2007: W = 0.93035, p-value = 0.1969

2020: NA

**2. Harvey-Collier, linearity**

harvtest()

**DAILY**

1969: HC = 1.3529, df = 238, p-value = 0.1774

1973: HC = 1.2182, df = 345, p-value = 0.224

1980: HC = 1.267, df = 128, p-value = 0.2075

1981: HC = 0.52272, df = 346, p-value = 0.6015

1990: HC = 0.98469, df = 171, p-value = 0.3262

2001: HC = 0.4337, df = 172, p-value = 0.6651

2007: HC = 0.55421, df = 387, p-value = 0.5798

2020: HC = 0.64862, df = 40, p-value = 0.5203

**WEEKLY**

1969:HC = 1.1489, df = 41, p-value = 0.2573

1973: HC = 1.041, df = 61, p-value = 0.302

1980: HC = 0.59373, df = 21, p-value = 0.559

1981: HC = 0.43444, df = 61, p-value = 0.6655

1990: HC = 0.64576, df = 29, p-value = 0.5235

2001: HC = 0.97042, df = 29, p-value = 0.3399

2007: HC = 0.032713, df = 69, p-value = 0.974

2020: HC = 0.31356, df = 5, p-value = 0.7665

**MONTHLY**

1969: HC = 1.6672, df = 8, p-value = 0.134

1973: HC = 1.5927, df = 13, p-value = 0.1352

1980: HC = 0.66759, df = 3, p-value = 0.5522

1981: HC = 0.071191, df = 13, p-value = 0.9443

1990: HC = 0.90727, df = 5, p-value = 0.4059

2001: HC = 0.82234, df = 5, p-value = 0.4483

2007: HC = 0.88448, df = 15, p-value = 0.3904

2020: NA

**3. Breusch-Pagan, heteroskedasticity**

**bptest()**

> bptest(lm.daily.1969)

**DAILY**

1969: BP = 0.58852, df = 1, p-value = 0.443

1973: BP = 0.063721, df = 1, p-value = 0.8007

1980: BP = 1.7687, df = 1, p-value = 0.1835

**1981: BP = 12.689, df = 1, p-value = 0.0003679**

**1990: BP = 5.1269, df = 1, p-value = 0.02356**

2001: BP = 0.031596, df = 1, p-value = 0.8589

2007: BP = 2.7183, df = 1, p-value = 0.0992

2020: BP = 0.035228, df = 1, p-value = 0.8511

**WEEKLY**

1969: BP = 0.15429, df = 1, p-value = 0.6945

1973: BP = 1.6623, df = 1, p-value = 0.1973

1980: BP = 1.5741, df = 1, p-value = 0.2096

1981: BP = 1.2787, df = 1, p-value = 0.2581

1990: BP = 0.54306, df = 1, p-value = 0.4612

2001: BP = 0.62927, df = 1, p-value = 0.4276

2007: BP = 1.6085, df = 1, p-value = 0.2047

2020: BP = 0.57076, df = 1, p-value = 0.45

**MONTHLY**

1969: BP = 0.74248, df = 1, p-value = 0.3889

1973: BP = 0.035481, df = 1, p-value = 0.8506

1980: BP = 0.22167, df = 1, p-value = 0.6378

1981: BP = 0.0021128, df = 1, p-value = 0.9633

1990: BP = 0.011314, df = 1, p-value = 0.9153

2001: BP = 0.35773, df = 1, p-value = 0.5498

2007: BP = 1.4842, df = 1, p-value = 0.2231

2020: BP = NaN, df = 1, p-value = NA

**4. Durbin-Watson, serial correlation**

> dwtest()

**DAILY**  
**1969: DW = 1.301, p-value = 2.396e-08**

1973:DW = 1.9239, p-value = 0.2341

1980: DW = 2.2109, p-value = 0.887

1981: DW = 2.3152, p-value = 0.9984

1990: DW = 2, p-value = 0.4953

2001: DW = 2.3067, p-value = 0.9791

2007: DW = 1.9624, p-value = 0.3575

2020: DW = 2.5189, p-value = 0.9603

**WEEKLY**

1969: DW = 1.5335, p-value = 0.0571

1973: DW = 1.8195, p-value = 0.2358

1980: DW = 2.4095, p-value = 0.8444

1981: DW = 1.8493, p-value = 0.2687

1990: DW = 2.0893, p-value = 0.5991

2001: DW = 2.3232, p-value = 0.8185

2007: DW = 2.1834, p-value = 0.7828

2020: DW = 3.2228, p-value = 0.9681

**MONTHLY**

1969: DW = 1.4173, p-value = 0.1553

**1973: DW = 1.1612, p-value = 0.03175**

1980: DW = 1.3904, p-value = 0.2059

1981: DW = 2.0501, p-value = 0.5092

1990: DW = 2.7823, p-value = 0.8306

2001: DW = 2.3607, p-value = 0.7181

2007: DW = 2.205, p-value = 0.655

2020: DW = NaN, p-value = 1

**Breusch-Godfrey, serial correlation**

>bgtest()

**DAILY**

**1969: LM test = 29.351, df = 1, p-value = 6.039e-08**

1973: LM test = 0.48929, df = 1, p-value = 0.4842

1980: LM test = 1.4873, df = 1, p-value = 0.2226

**1981: LM test = 8.701, df = 1, p-value = 0.00318**

1990: LM test = 0.0064812, df = 1, p-value = 0.9358

**2001: LM test = 4.1811, df = 1, p-value = 0.04088**

2007: LM test = 0.11877, df = 1, p-value = 0.7304

2020: LM test = 3.2226, df = 1, p-value = 0.07263

**WEEKLY**

1969: LM test = 1.4502, df = 1, p-value = 0.2285

1973: LM test = 0.49106, df = 1, p-value = 0.4835

1980: LM test = 2.3137, df = 1, p-value = 0.1282

1981: LM test = 0.29983, df = 1, p-value = 0.584

1990: LM test = 0.082877, df = 1, p-value = 0.7734

2001: LM test = 1.1087, df = 1, p-value = 0.2924

2007: LM test = 0.6417, df = 1, p-value = 0.4231

2020: LM test = 3.1562, df = 1, p-value = 0.07564

**MONTHLY**

1969: LM test = 0.04634, df = 1, p-value = 0.8296

1973: LM test = 2.7677, df = 1, p-value = 0.09618

1980: LM test = 0.059101, df = 1, p-value = 0.8079

1981: LM test = 0.034358, df = 1, p-value = 0.8529

1990: LM test = 2.8172, df = 1, p-value = 0.09326

2001: LM test = 1.9363, df = 1, p-value = 0.1641

2007: LM test = 0.37806, df = 1, p-value = 0.5386

2020: NA

**Rainbow test, linearity (model fit)**

**raintest()**

MODEL FIT

>raintest()

Rainbow test

**DAILY**  
1969: Rain = 0.47843, df1 = 121, df2 = 118, p-value = 1

**1973: Rain = 1.3801, df1 = 174, df2 = 172, p-value = 0.01742**

1980: Rain = 1.2226, df1 = 66, df2 = 63, p-value = 0.2117

**1981: Rain = 2.5301, df1 = 175, df2 = 172, p-value = 1.037e-09**

**1990: Rain = 1.4798, df1 = 87, df2 = 85, p-value = 0.03563**

2001: Rain = 1.0026, df1 = 88, df2 = 85, p-value = 0.4958

2007: Rain = 0.53208, df1 = 195, df2 = 193, p-value = 1

2020: Rain = 0.32631, df1 = 22, df2 = 19, p-value = 0.993  
  
**WEEKLY**  
1969: Rain = 1.138, df1 = 22, df2 = 20, p-value = 0.3878

1973: Rain = 1.573, df1 = 32, df2 = 30, p-value = 0.1078

1980: Rain = 0.87543, df1 = 12, df2 = 10, p-value = 0.5924

**1981: Rain = 2.3516, df1 = 32, df2 = 30, p-value = 0.01039**

1990: Rain = 2.1175, df1 = 16, df2 = 14, p-value = 0.08273

2001: Rain = 1.1396, df1 = 16, df2 = 14, p-value = 0.4066

2007: Rain = 0.57563, df1 = 36, df2 = 34, p-value = 0.9473

2020: Rain = 0.021209, df1 = 4, df2 = 2, p-value = 0.9983  
  
**MONTHLY**

1969: Rain = 1.3624, df1 = 6, df2 = 3, p-value = 0.4313

1973: Rain = 2.9527, df1 = 8, df2 = 6, p-value = 0.102

1980: Rain = 4.2748, df1 = 3, df2 = 1, p-value = 0.3383

1981: Rain = 1.9553, df1 = 8, df2 = 6, p-value = 0.2147

1990: Rain = 0.44866, df1 = 4, df2 = 2, p-value = 0.7763

2001: Rain = 9.2756, df1 = 4, df2 = 2, p-value = 0.09968

2007: Rain = 0.49069, df1 = 9, df2 = 7, p-value = 0.842

2020: NA